

Human and Artificial Intelligence in Literary Translation Evaluating Meaning, Style, and Cultural Representation

Gunawan Tambunsaribu
Universitas Kristen Indonesia

***Corresponding Author:** Gunawan Tambunsaribu, gunawan.ryeza@gmail.com

ARTICLE INFO

Keywords: Artificial Intelligence, Literary Translation, Meaning, Style, Cultural Representation.

Received : 27, June

Revised : 29, July

Accepted: 25, August

©2026 Tambunsaribu: This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

The development of artificial intelligence in translation raises questions about its ability to maintain meaning, style, and cultural representation in English literature. This study aims to compare human and AI translations in parts of texts that contain metaphors, idioms, narrative styles, and cultural elements. The study used a comparative qualitative approach by involving three human translators, one AI model, and two translation experts as validators. Data was collected through translation documentation and expert review, then analyzed using comparative textual analysis and thematic coding. The results show that human translators more consistently retain figurative nuances, narrative voices, and cultural distinctiveness, while AI tends to produce more explicit and standardized translations. The findings confirm the importance of human interpretation in AI-assisted literary translation

INTRODUCTION

Literary translation not only transfers information from the source language to the target language, but also preserves the contextual meaning, aesthetic style, narrative voice, and cultural elements that make up the identity of a work. The development of large language models (LLMs) and generative artificial intelligence has expanded the use of technology in translation, including literary texts that previously relied heavily on human creativity and interpretation. AI is capable of producing fluent and contextual translations, but linguistic fluency does not always guarantee the accuracy of meaning or sensitivity to cultural elements. In the Indonesian context, this issue is relevant for the study of English literature because translation is one of the important means to bridge English-language works with Indonesian readers while maintaining the literary and cultural character of the source text.

This development raises questions about the extent to which AI is able to carry out interpretive functions that have been inherent in human translators. Gao et al. (2024), for example, found that ChatGPT performed better than Google Translate and DeepL in translating classical Chinese poetry in terms of fidelity, fluency, and language style, including in understanding certain symbols and images. However, a direct comparison with human translators reveals a more complex picture. Al Rousan et al. (2025) reported that human translations obtained an accuracy rate of 94.5%, while ChatGPT reached 77.9%; on the other hand, in terms of fluency, ChatGPT obtained 97.2% and human translation 96.6%. The findings suggest that highly fluid AI output is not necessarily synonymous with the ability to retain the meaning, literary nuance, and cultural sensitivity of the source text.

A number of previous studies have discussed the relationship between technology and literary translation from different perspectives. Guerberof-Arenas and Toral (2022) found that human translation has a higher level of creativity than post-edited translation and machine translation, especially when the text requires non-literal solutions. Meanwhile, Kolb et al. (2023) show that machine translation systems can generate some viable new lexical forms when translating occasionalisms in literary texts, although most are formed through literal procedures. In terms of reader acceptance, Guerberof-Arenas and Toral (2024), through a study of 223 participants, found that responses to human translation, machine translation, and post-editing results can differ based on the target language and readers' reading patterns. These results show that the quality of literary translation is multidimensional and cannot be assessed through linguistic accuracy alone.

Although research on AI in literary translation is growing, there are still gaps that need to be studied. The study of Gao et al. (2024) focuses more on comparisons between AI systems and machine translation in poetry, while Al Rousan et al. (2025) assessed human translation and ChatGPT primarily through the dimensions of accuracy, fluency, and design. Abdelhalim et al. (2025) also examined the use of ChatGPT and Google Translate in literary translation from the perspective of student translators, while Chen and Lin (2025) showed the importance of cultural sensitivity in the evaluation of AI-based translations on

culturally charged texts. Thus, there is still limited research that qualitatively and integrally compares human translators and AI through three interrelated dimensions, namely meaning, style, and cultural representation, especially in English-language literary texts.

Based on these gaps, this study aims to compare the results of human translation and artificial intelligence in translating English literary texts, focusing on how both maintain meaning, style, and cultural representation. Aspects of meaning are analyzed through semantic accuracy, contextual meaning, and the possibility of shifting, addition, and loss of meaning; aspects of style include diction, metaphor, figurative language, tone, and narrative voice; Meanwhile, cultural representation includes the translation of specific cultural terms, idioms, social practices, and forms of adaptation or preservation of cultural elements. The study used a comparative qualitative approach by comparing the translation results of three human translators and one AI model, while two translation experts were involved to support the validation of interpretation through expert review.

Theoretically, this study is expected to expand the study of literary translation and human-AI translation by showing that the evaluation of literary translation needs to go beyond the measure of fluency and accuracy of language by considering the relationship between meaning, aesthetic form, and cultural representation. The approach also allows the identification of interpretive patterns that distinguish human translator decisions from AI outputs without first assuming that one is always superior. Practically, the results of the research are expected to provide a basis for translators, literature and translation students, teachers, editors, and publishers in determining the critical use of AI in literary translation. Thus, AI can be placed as a supporting technology, while decisions related to interpretation, literary nuance, and cultural representation remain evaluated through human consideration.

LITERATURE REVIEW

Literary Translation as Semantic, Stylistic, and Cultural Reconstruction

Literary translation is an interpretive process that goes beyond the replacement of linguistic elements from the source language to the target language. Literary texts contain complex relationships between meaning, diction choices, styles, narrative sounds, metaphors, and cultural contexts so that translations must consider textual equivalence as well as the aesthetic effects produced. Wang and Xin (2024) assert that translation decisions are influenced by the characteristics of the text, the translator's background, and the sociocultural context of the target language. In literary works, these relationships become increasingly important because tone, imagery, rhythm, and lexical choices are part of the construction of meaning, not just decorative elements. Thus, the quality of literary translation needs to be understood through the relationship between meaning, style, and cultural representation, which is the conceptual framework of this research.

Human and AI Translation in Representing Meaning

The emergence of generative artificial intelligence is changing the comparison between human and machine translation because large language models no longer only produce equivalence at the word level, but can process context and generate relatively natural language structures. However, the similarity of smoothness does not always indicate the same process or interpretation results. Fu and Liu (2024) found systematic linguistic differences between ChatGPT translations and human translations; AI shows a certain advantage in terminology consistency, while human translators are more flexible in reconstructing complex sentences to suit the needs of the reader. Although the study focused on scientific texts, the results showed that humans and AI have different linguistic decision-making tendencies. In literary texts, this problem becomes more complex because the meaning is often implicit, connotative, ambiguous, or figurative. Therefore, the analysis of meaning in this study not only considers semantic accuracy, but also the shift in contextual meaning, explicitness, omission, addition, and nuance changes between the source text, human translation, and AI translation.

Studies of creative texts increasingly show that AI capabilities cannot be generalized simply as inferior or superior to humans. Farghal and Haider (2024), in a comparison of the translations of 15 stanzas of classical Arabic poetry by humans, ChatGPT, and Gemini, found that assessments of human translations and ChatGPT were relatively positive in thematic clarity, creativity, and prosody, while Gemini showed more obvious weaknesses, particularly in prosody. The findings suggest that AI performance depends on the model and characteristics of the translated literary elements. This means that human-AI translation studies are more appropriate to identify patterns of interpretive differences rather than simply determining the party that produces the best translation. The approach underpins this study in comparing how three human translators and one AI model handle the same meaning in English literary texts.

Literary Style and the Preservation of Narrative Voice

Style is one of the dimensions that distinguishes literary translation from translation that is mainly informative. Styles are formed through a combination of diction, syntactic structure, registers, metaphors, rhythms, narrative perspectives, and other expressive patterns. Saed et al. (2024) show that the translation of fictional registers requires the management of linguistic choices that are able to preserve the character and style of speech in fictional texts. This suggests that changes in registers or language choices can alter characterization, character relationships, and literary effects even though literal meanings are still understandable. Therefore, the evaluation of literary translations does not stop at the question of whether the information has been translated correctly, but also needs to assess whether the author's way of conveying the meaning is still reflected in the target language.

The development of LLMs further complicates the assumption that style is the exclusive domain of human translators. Yao et al. (2025), through a stylometric analysis of GPT-4 translations and human translations in Chinese online literature, found that GPT-4's output can approach human translation in a number of lexical, syntactic, and content characteristics. The findings show that

cutting-edge AI is able to mimic some textual patterns that have been associated with human translation styles. However, stylistic similarity does not in itself explain whether certain metaphors, tone, narrative voices, or aesthetic effects are maintained through the same interpretive decisions. Thus, close textual analysis is needed to complete the formal measurement. In this study, the aspect of style is therefore analyzed through diction, metaphor, figurative language, tone, and narrative sound so that the differences between humans and AI can be seen at the textual and interpretive levels.

Cultural Representation in Literary Translation

Cultural representation is a fundamental issue because literary language carries social experiences, values, symbols, idioms, practices, and identities that do not always have a direct equivalent in the target culture. Haider and Shuhaiber (2024) show that culturally bound idiom translation requires more than lexical equivalent; In the corpus of Egyptian films they analyzed, cultural substitution became the dominant strategy, while literal translations in some cases resulted in errors or loss of nuance. The findings confirm that understanding the source culture and the culture of the target determines how implicit meanings can be maintained or reconstructed. In the context of literature, strategies such as term defense, explanation, generalization, cultural substitution, or omission ultimately determine how the cultural identity of the source text is present to the target language reader.

Metaphors also show the close relationship between meaning, style, and culture. Abu Rumman et al. (2023), through a corpus-based analysis of conceptual metaphors in the translation of English-language television series, show that the translation of metaphors is not only dealing with linguistic structures, but also with conceptualizations that may differ between cultures. This is important for literary research because cultural-specific metaphors, idioms, symbols, and expressions can simultaneously carry connotative meanings and aesthetic functions. Therefore, cultural representation in this study is understood not as a separate category, but as a process that can affect the meaning and style of translation. The analysis will look at whether the human and AI translators retain cultural distinctiveness, provide explanations, replace those elements with the cultural equivalent of the target, or neutralize them so that the cultural identity of the source is changed.

Human-AI Complementarity in Literary Translation

The debate about AI in translation is increasingly shifting from the issue of human replacement to the question of the division of functions between humans and technology. Jiménez-Crespo (2025) shows that the development of translation technology needs to be directed towards a human-centered approach to AI, especially in maintaining the control and professional autonomy of translators. These ideas are relevant to literary translation because decisions regarding ambiguity, characterization, style, and culture often require considerations that cannot be reduced to linguistic fluency. Fu and Liu's (2024) findings regarding the complementary characteristics between AI and humans also show that technology can be effective in certain aspects without eliminating human decision-making functions. Thus, research on literary translation needs

to avoid the simple dichotomy of "human versus AI" and further explain what types of decisions AI can replicate and in which parts human interpretive interventions still play an important role.

Based on this study, previous research has yielded important findings regarding the differences between human and AI translations, AI's ability to handle creative texts, the characteristics of translation styles, and the complexity of translating metaphors and cultural elements. However, these studies generally focus on one dimension separately – for example, linguistic accuracy, prosody characteristics, stylistometric characteristics, metaphors, or idioms. There is still a need for qualitative analysis that brings together meaning, style, and cultural representation in one comparative framework, especially in the translation of literary works in English. This research fills the space through comparative textual analysis of the results of three human translators and one AI model, with the analysis of two translation experts to support the credibility of the interpretation. This position allows research to not only identify differences in translation outcomes, but also to shed light on how human and AI choices affect the preservation of meaning, literary effects, and cultural identity of source texts.

METHODOLOGY

Types of Research

This study uses a qualitative approach with a comparative design to compare the results of English literary text translations produced by human translators and artificial intelligence. The qualitative approach was chosen because the research was not directed at examining statistical relationships, but at identifying and interpreting patterns of difference in the defense of cultural meaning, style, and representation. The main data is in the form of source text, the translation results of three human translators, and the translation results of one AI model. This design is consistent with the research goal of placing translation as an interpretive product and allowing differences in meaning and cultural context to be studied beyond mere literal equivalence (Zhao et al., 2024). This approach is also in accordance with the research design that has established comparative textual analysis and thematic coding as the main analysis techniques.

Population and Sample

In this qualitative research, populations are not used for statistical generalization purposes. The study participants consisted of three human translators and two translation experts, while one AI model was used as a comparison system, not as a human respondent or participant. Thus, the study involved five human participants according to their respective functions, namely three people in the translation process and two people in the validation of the interpretation of the analysis results. Participants were selected using non-probability purposive sampling, because participants were selected based on their direct relevance to the needs of translation and research validation. Purposive sampling is suitable for qualitative research when the selected unit is determined based on its ability to provide information relevant to the research objectives (Ahmad & Wilkins, 2025).

The selection of textual data is also carried out purposively. The part of the literary text that becomes the unit of analysis is the part that contains metaphors, idioms, narrative styles, and cultural elements, as stipulated in the research design. The unit was chosen because it provides a relevant basis for evaluating how humans and AI handle semantic, stylistic, and cultural aspects. The purposive approach in qualitative research emphasizes the suitability of the selected unit to the research question and objective, rather than statistical representativeness (Ahmad & Wilkins, 2025).

Research Instruments

The research instrument is in the form of text documentation and a comparative analysis matrix used to juxtapose source text with human translation and AI translation. No questionnaires or quantitative measurement instruments are used. The framework of analysis follows three dimensions that have been established in the study. The dimension of meaning includes semantic accuracy, contextual meaning, shift, addition, and loss of meaning; stylistic dimensions include diction, metaphors, figurative language, tone, and narrative voice; Meanwhile, the dimension of cultural representation includes cultural-specific terms, idioms, social practices, and forms of adaptation or preservation of cultural elements.

Because the research is qualitative, validity and reliability testing is not performed using statistical coefficients such as Cronbach's alpha. The credibility of the interpretation was strengthened through the analysis of two translation experts. In qualitative research involving translation, validity relates not only to the literal similarity between the source text and the target text, but also to transparency in interpreting variations in meaning as well as cultural context (Zhao et al., 2024). Thematic coding is also carried out transparently and consistently with the research questions; Braun and Clarke (2021) emphasize that the quality of thematic analysis is determined by the coherence between the research objectives, coding process, theme development, and interpretation, not solely by measurements of inter-coder reliability.

Data Collection Procedure

Data collection is carried out in stages. First, the researcher identified the parts of English literary texts that contain metaphors, idioms, narrative styles, and cultural elements according to the focus of the research. Second, the same part of the text is used as a translation source by three human translators and one AI model so that the results obtained can be compared in an equivalent textual context. Third, all translation results are documented and juxtaposed with source texts based on three dimensions of analysis, namely meaning, style, and cultural representation. Fourth, the differences and similarities that arise are identified and codified. Fifth, the interpretation of the comparative results was reviewed by two translation experts as part of qualitative validation before the thematic pattern was formulated. This procedure is consistent with the draft of the manuscript that puts three human translators, one AI, and two experts in the comparison and validation process.

Data Analysis Techniques

The data was analyzed through two complementary stages, namely comparative textual analysis and thematic coding. In the first stage, each source text is systematically compared with human and AI translations to identify similarities, differences, shifts, and translation choices in terms of meaning, style, and culture. The analysis is not only oriented towards literal equivalence, but also considers the implicit meaning, contextual nuances, and cultural consequences of each translation choice. This approach is in line with the view of Zhao et al. (2024) that translation analysis needs to pay attention to variations in interpretation and cultural context in the formation of meaning.

In the second stage, the results of the comparison are coded and grouped to find thematic patterns. The analysis process includes familiarization of the data, coding of relevant differences, grouping of codes based on three main dimensions, review of patterns, and interpretation of findings. These stages are in line with the principles of reflective thematic analysis which emphasizes the iterative process from familiarization, coding, pattern development, to the preparation of coherent interpretations (Braun & Clarke, 2021; Byrne, 2022). The final results were then compiled to explain the tendency of human and AI translation to maintain or alter meaning, stylistic effects, and cultural representations.

This study does not use statistical analysis because the entire data is analyzed qualitatively. Specific analysis software is not specified in the available research design, so specific software names are not included to avoid adding procedures that are not performed. The analysis was carried out through the organization of textual data, comparison matrices, coding processes, and interpretation of results with the support of expert analysis. With such a procedure, the analysis can still be traced from the source text to the code, pattern, and conclusion of the research.

RESEARCH RESULT

The results of the research were obtained through a comparative analysis of the English literature text that contains metaphors, idioms, narrative styles, and cultural elements. Each source *text* (ST) was compared with the results of three human translators (HT1, HT2, and HT3) and one AI translation (AIT). The analysis is carried out based on three dimensions that have been determined in the methodology, namely meaning, style, and cultural representation, through *comparative textual analysis* and *thematic coding*. The interpretation of the pattern found was then examined by two translation experts.

Defense of Meaning

The results of the analysis show that both human and AI translators in general are able to retain the main information of the source text. Differences are more clearly found when the text contains implicit or contextual meaning.

Table 1. Comparison of Contextual Meanings

Versions	Text
ST	<i>She carried the weight of the whole family on her shoulders.</i>
HT1	He carried the burden of his entire family on his shoulders.
HT2	The entire burden of the family seems to be on his shoulders.
HT3	He bore the weight of responsibility for his entire family.
AIT	He is in charge of his entire family.

HT1 maintains the metaphor of *weight* and *shoulders* directly through the phrase "carrying the burden ... on his shoulders". HT2 also retains the image with the addition of "if", while HT3 begins to explain *weight* as "responsibility" but still retains an emotional feel through the word "weight". In contrast, AIT produces the form "responsible for his entire family", which conveys the basic meaning clearly but omits the image of burden contained in the source text.

A similar pattern is found in the following data.

Table 2. Implicit Meaning Comparison

Versions	Text
ST	<i>He smiled, but there was winter in his eyes.</i>
HT1	He smiled, but there was a cold in his eyes.
HT2	He smiled, but his eyes kept the cold of winter.
HT3	He smiled, even though his gaze felt as cold as winter.
AIT	He smiled, but his eyes showed sadness and coldness.

HT1 and HT2 retain the metaphorical ambiguity of *winter in his eyes*. HT3 gives a bit of interpretation through the word "feels cold", but still retains the image of winter. AIT directly interprets *winter* as "sadness and coldness". The meaning becomes easier to understand, but the reader no longer gains the same interpretive space as in the source text.

The data shows that human translators tend to retain contextual and implicit meanings, while AI more often generates explanations of meaning.

Defense of Metaphors and Figurative Language

A more pronounced difference is found in the dimension of style. Human translators show a more consistent tendency to retain metaphorical and figurative forms, whereas AI more often produces standardized formulations.

Table 3. Translation of Metaphors

Versions	Text
ST	<i>The night swallowed the village whole.</i>
HT1	The night swallowed up the entire village.
HT2	The night seemed to swallow the village in a round.
HT3	The darkness of the night swallowed the village completely.
AIT	The darkness of the night enveloped the entire village.

HT1-HT3 maintains personification through the word "swallow". Although the sentence structure is different, all three versions retain the figurative function of the swallowed verb. AIT replaces "swallowing" with "enveloping", which is more conventional in Indonesian. Semantically, the dark atmosphere remains understandable, but the intensity of the personification of the source text becomes reduced.

Similar differences are seen in the following figurative language.

Table 4. Image Defense

Versions	Text
ST	<i>Her voice was a knife wrapped in silk.</i>
HT1	His voice was a knife wrapped in silk.
HT2	His voice was like a knife wrapped in silk.
HT3	Behind the softness of his voice lies the sharpness of a knife.
AIT	His voice sounded soft, but his words were very sharp.

HT1 retains the metaphorical structure very closely, while HT2 transforms it into a simile shape but still retains the imagery of the knife and silk. HT3 reconstructs more freely by maintaining the opposition between softness and sharpness. AIT translates the function of metaphor semantically through "soft" and "sharp", but omits the two concrete imagery that build the literary effect.

The findings suggest that the difference between humans and AI is not primarily in the understanding of the meaning of the metaphor, but in the way the aesthetic effects of the metaphor are reproduced in the target text.

Diction, Tone, and Narrative Sound

Analysis of narrative diction and sound shows that all three human translators use more varied forms of language.

Table 5. Comparison of Narrative Voices

Versions	Text
ST	<i>Poor little thing, she knew nothing of the cruelty waiting beyond the door.</i>
HT1	The poor boy, he did not know the cruelty that awaited him behind the door.
HT2	Pity the little girl; He had no idea of the cruelty that awaited him behind the door.
HT3	How unfortunate that child is. He didn't realize that cruelty was waiting for him behind the door.
AIT	The little girl was unaware of the danger that awaited her outside the door.

HT1-HT3 maintains the narrator's emotional attitude through "poor child", "what a pity", and "how unfortunate". AIT conveys the incident with a more neutral structure. Additionally, *cruelty* translates to "danger", making the AI output more informative but reducing the narrator's emotional evaluation.

In this data, the change does not cause the event delivery failure. However, the narrator's voice character becomes weaker in AIT than in the three human translations.

Idiom Translation

The next difference is found in idiomatic expressions, especially when the literal translation is insufficient to convey contextual meaning.

Table 6. Idiom Comparison

Versions	Text
ST	<i>You are crying for the moon.</i>
HT1	You want something impossible.
HT2	You're like your back longing for the moon.
HT3	You expect something you can't possibly have.
AIT	You want something that is impossible to achieve.

HT1 and HT3 use paraphrasing to maintain the idiomatic meaning, while HT2 chose the Indonesian idiom "back longing for the moon". HT2 thus does not retain the original linguistic structure, but produces expressions that have similar idiomatic functions in the target culture. AIT produces a clear semantic explanation, but does not retain the idiomatic character of the expression.

This data shows significant variation among human translators. Human translation does not indicate a single strategy; Translators can retain functionality through paraphrasing or using target cultural expressions.

Representation of Cultural Elements

On the cultural element, human translators show greater variation in maintaining or explaining the cultural context of the source.

Table 7. Translation of Cultural Terms

Versions	Text
ST	<i>The parish priest greeted them after Sunday service.</i>
HT1	The parish priest greeted them after Sunday mass.
HT2	The pastor of the church greeted them after the Sunday service.
HT3	The local priest greeted them after Sunday services.
AIT	Religious leaders greet them after Sunday services.

HT1 retains the terms "parish priest" and "mass", so that the context of religious institutions looks most specific. HT2 uses "church pastors" and "services", which result in different representations of religious cultures. HT3 retains "pastor", but generalizes the *service* to "worship". The AIT uses "religious leader" and "worship", making it the most common version.

The comparison shows that AI tends to reduce the specificity of terms when there are multiple possible matches, whereas human translators make more culturally specific decisions even though the results between translators are not always identical.

Social Practices and Greeting Forms

Cultural representation is also seen in the translation of greetings and social relations.

Table 8. Translation of Greeting Forms

Versions	Text
ST	"My Lord," the servant whispered, lowering his eyes.
HT1	"My lord," the servant whispered, lowering his gaze.
HT2	"Sir," the maid whispered, lowering her head.
HT3	"Your Majesty," said the servant quietly with a low eye.
AIT	"Sir," the waiter said quietly with his head down.

HT1 maintains a hierarchical relationship through "My Lord", while HT3 uses "Your Majesty", which reinforces the level of formality. HT2 chooses "Master", which is more neutral. AIT also uses "Sir" and changes *whispered* to "word... slowly". Conversational information is still conveyed, but the social index and speech style have become less powerful.

The data show that cultural representation is not only found in explicit cultural terms, but also in the choice of greetings that represent power relations and social hierarchies.

Variations between HT1, HT2, and HT3

The results of the analysis also show that the three human translators have different tendencies. HT1 tends to maintain the structure and imagery of the source text relatively closely. HT2 showed a tendency to use idiomatic equivalents and natural forms in the target language. HT3 more often reconstructs to maintain the function of meaning and narrative effect even though the structure of language changes.

Table 9. Textual Tendencies of Three Translators and AI

Dimensions	HT1	HT2	HT3	AIT
Meaning	Close to the source form	Contextual	Interpretive	Eksplisit
Diction	Relatively close to ST	More idiomatic	Adaptive	Terstandar
Metaphor	Tends to be maintained	Retained/adapted	Reconstructed	Tends to be explained
Narrative voice	Retained	Expressive	Interpretive	Relatively neutral
Idiom	Paraphrasing/retention	Idiomatic equivalents	Contextual paraphrasing	Live explanation
Cultural elements	More specific	Adaptive	Selective	More common

Although different, the three human translators show a common tendency, which is to make linguistic changes based on textual functions. AI is more consistent in producing clear and natural translations, but that consistency also results in standardization of some parts that are stylistically or culturally more complex.

The Relationship between Meaning, Style, and Cultural Representation

Thematic analysis shows that meaning, style, and culture do not operate in isolation. For example, in ST *Her voice was a knife wrapped in silk*, AIT translates it as "Her voice sounds soft, but her words are very sharp." The change retains the connotative meaning, but at the same time removes the imagery of "knife" and "silk". Thus, an increase in semantic clarity results in a change in style.

The same pattern can be seen in the *crying for the moon* idiom. HT2 uses "the back longs for the moon", so the form of language changes but idiomatic function and figurative qualities remain to emerge through the target culture. AIT uses "wanting something that is impossible to achieve", which retains meaning but removes the figurative dimension.

The relationship can be summarized as follows.

Table 10. The Relationship Between Meaning, Style, and Cultural Representation

Changes	Meaning	Style	Culture
Dipertahankan metaphor	The connotative meaning remains open	Image maintained	Symbols remain visible
Metaphors explained	The meaning becomes clear	Reduced figurative effects	Tends to be more neutral
A Substitute for the Target Idiom	Contextual meaning maintained	Idiomatic effects remain	Target culture is more prominent
Cultural terms are generalized	Information is easier to understand	More neutral to diction	Cultural peculiarities are diminished

Expert Analysis

The results of *comparative textual analysis* and thematic coding were then analyzed by two translation experts. Studies show that differences between HT1-HT3 and AIT are most consistent in data containing figurative language, narrative voices, idioms, and culturally specific elements.

Both experts also identified that some units cannot be categorized exclusively as meaning, style, or culture alone. The metaphor *of winter in his eyes*, for example, relates simultaneously to connotative meaning and figurative style. Similarly, *crying for the moon* shows the relationship between idiomatic meaning, style, and cultural representation when translated to "back longing for the moon".

Expert studies maintain the main pattern that human translations show higher interpretive flexibility, whereas AITs show greater consistency in producing explicit and standardized language.

Summary of Findings

Based on the overall textual data, three main patterns were found. First, in meaning, both humans and AI are capable of conveying basic information, but HT1-HT3 more often retains implicit and contextual meaning, whereas AIT tends to explain it directly.

Second, in style, HT1-HT3 more consistently retains metaphor, figurative language, tone, and narrative sound. AI is capable of capturing the semantic functions of literary expression, but often realizes it through more conventional and standardized structures.

Third, on cultural representation, human translators show more varied strategies in maintaining, adapting, or replacing cultural expressions, while AI tends to use more general and easy-to-understand forms.

Thus, textual data show that human translators are more consistent in retaining figurative nuances, narrative voices, and cultural distinctiveness, while AI tends to produce more explicit and standardized translations.

DISCUSSION

Meaning, Interpretation, and Explicit Tendencies of AI

The research findings show that both human and AI translators are able to retain basic information of the source text, but they show different tendencies when dealing with implicit, connotative, and contextual meanings. In the data She carried the weight of the whole family on her shoulders, for example, HT1 and HT2 maintain the imagery of "burden" and "shoulders", HT3 makes a more explicit interpretation through the concept of "responsibility", while AIT completely transforms the metaphor into a direct statement of family responsibility. The same pattern was seen in there was winter in his eyes, when the AI interpreted the metaphor to be "sadness and coldness", while the human version still retained the image of winter. These findings suggest that human-AI translation differences are not necessarily semantic errors; The difference is often at the level of interpretation explicitness.

The pattern is in line with Fu and Liu (2024), who found that generative and human translations have different linguistic characteristics even though they can produce acceptable target texts. In their research, AI showed high consistency on certain elements, while human translators were more flexible when it came to reconstructing complex structures. The findings of this study extend the issue to literary translation: human flexibility occurs not only in syntactic structure, but also in the decision whether implicit meaning needs to be preserved, reconstructed, or explained. Thus, semantic accuracy in literature cannot be equated with the level of straightforward meaning. More explicit translations can indeed improve readability, but they can also reduce ambiguity that the author deliberately constructs.

These results also explain why AI cannot be judged solely on fluency. Gao et al. (2024) show that ChatGPT can even outperform Google Translate and DeepL in translating classical Chinese poetry in fidelity, fluency, and language

style. The study also showed that prompts that explicitly requested rhythm and rhyme retention improved ChatGPT's style performance. However, the same study still found fidelity issues, the addition of elements that are not present in the source text, and the need for human post-editing. Therefore, AI's tendency in this study to clarify meaning can be understood as both a strength and a limitation: AI is capable of identifying semantic content, but the decision to make something implicit explicit can alter the literary function of the text.

Defense of Style, Metaphor and Narrative Voice

A stronger difference is found in the style aspect. HT1-HT3 tends to retain metaphors, imagery, and figurative forms, whereas AIT more often turns them into conventional constructions. In *The night swallowed the village whole*, the three human translators retain the figurative verb "swallow", while the AI changes it to "envelops". Similarly, in *Her voice was a knife wrapped in silk*, all three versions of humans still retain the image of a knife or the relationship between softness and sharpness, while the AI turns the metaphor into a straightforward explanation: "soft, but ... very sharp." The data shows that AI does not fail to understand metaphorical meaning; What changes is the mode of representation, from figurative to propositional.

These findings support Guerberof-Arenas and Toral (2022), who show that human translation achieves a higher level of creativity than machine translation and post-edited machine translation in literary texts. They found that machine output more often offered literal solutions and was less creative when the text required non-routine translation decisions. In this study, the pattern emerges not always in the form of literality, but as stylistic normalization: AI retains the core of the message while reducing the distinctiveness of linguistic forms. In other words, the development of LLMs has made AI more capable of "understanding" metaphors than previous MT systems, but the ability to understand figurative functions has not always been followed by the decision to retain its aesthetic form.

Findings on narrative voice also support the results of Zhou and Cheng (2025). In a comparison of translations of Chinese science fiction works, they found that human translator Ken Liu showed a greater variation of styles tailored to the character of each work, while ChatGPT-4o showed a more consistent style pattern. Such consistency results in clarity, but under certain conditions it can reduce the thematic impact of the text. A very similar pattern was seen in this study. In the phrase *Poor little thing*, HT1-HT3 maintains the narrator's evaluative position through "poor child", "pity", and "how unfortunate", while AIT removes these judgments and conveys events in a more neutral form. The standardization of AI thus not only affects diction, but can change the pragmatic relationship between narrator, character, and reader.

The findings do not mean that AI always fails to reproduce human styles. Yao et al. (2025) found that GPT-4 can approximate human translation in a number of lexical, syntactic, and content features based on a stylometric analysis of Chinese online literature. The difference with this study can be explained by the level of analysis used. Statistical similarity to global linguistic features does not in itself mean that certain metaphors, tones, and narrative voices are

preserved at the micro level. Thus, the close textual analysis used in this study complements the stylometric approach by showing specific parts where formal similarities between humans and AI can hide interpretive differences.

Cultural Idioms and Representations

The aspect of cultural representation shows a clearer difference in orientation. In the expression *You are crying for the moon*, HT2 produces "buttocks longing for the moon", which is an idiom in Indonesian that is lexically different but maintains a figurative and pragmatic function. HT1 and HT3 choose contextual paraphrasing, while AIT produces a direct explanation of something that is "impossible to achieve". This data shows that cultural equivalence does not necessarily require the same words. In contrast, decisions that retain idiomatic functions may require a change of form that is quite far removed from the source text.

The findings are consistent with Haider and Shuhaiber (2024), who found that cultural substitution is the dominant strategy in idiom translation and that literal translation can lead to shifting or loss of meaning when cultural nuances are not understood appropriately. In this study, HT2 shows an example of how cultural substitution can maintain idiomatic function through target cultural sources. This shows that cultural competence in literary translation is not just about knowing the definition of an idiom, but determining how its social, aesthetic, and pragmatic functions can be rebuilt for the target reader.

The same pattern is seen in the parish priest, Sunday service, and the form of the *My Lord* greeting. AI uses the forms "religious leader" and "Lord", while some human translators choose more specific terms such as "parish priest", "mass", "my Lord", or "Venerable". The AI selection improves readability because it uses more general categories, but the generalization process reduces information about religious institutions or social hierarchies in the text world. Chen and Lin (2025) also show that AI can achieve high performance in cultural sensitivity, especially with specially designed prompts, but increased cultural adaptation can be accompanied by semantic shifts. They therefore still emphasize the importance of human intervention in culturally complex matters.

Consequently, cultural representations need to be understood as part of meaning, not an additional layer after the translation process is complete. When a parish priest is generalized to be a "religious leader", the propositional meaning of a religious leader welcoming a figure remains, but information about certain religious traditions changes. This corroborates the findings of the study that meaning, style, and cultural representation are interdependent, as seen in the results of cross-category coding.

Variations of Human Translators and Human-AI Relations

One of the important findings of this study is that human translation itself is not homogeneous. HT1 tends to be close to the imagery of the source text, HT2 is more idiomatic, and HT3 is more interpretive and adaptive. AIT, on the other hand, exhibits an explicit and standardized tendency. This is important because the comparison of humans and AI should not be constructed as a dichotomy that "humans have one way of translating and AI has another". In fact, the variation of HT1-HT3 shows that literary translation is an interpretive decision space that allows for several equally accountable solutions.

The findings also help explain why previous research on human-AI translation has not always been uniform. Farghal and Haider (2024), for example, found that ChatGPT is capable of being a serious competitor to human translators in the translation of classical Arabic poetry; Assessments of humans and ChatGPT were equally positive in thematic clarity, creativity, and prosody, while other AI models showed weaker results. On the contrary, Guerberoof-Arenas and Toral (2022) found the advantage of human translation in creativity. The differences in results between these studies indicate that AI performance is influenced by the model, language pairs, genre, prompt form, evaluation criteria, and text characters analyzed, rather than by the "AI" status in general.

Therefore, the findings of this study are more appropriately interpreted as the basis for human-AI complementarity, rather than as evidence that AI has no place in literary translation. Fu and Liu (2024) also concluded that humans and AI exhibit complementary abilities on different types of linguistic decisions. Jiménez-Crespo (2025) reinforces this perspective through a human-centered AI framework, by showing that professional translators want technologies that still place control and final decisions on humans, especially with regard to quality and context adaptation. In literary translation, the implication is that AI can serve to generate translation alternatives, accelerate lexical exploration, or provide drafts, while decisions regarding ambiguity, style, and cultural representation still require human evaluation.

Compatibility and Differences with Previous Research

In general, the results of the study support the conceptual expectation in the introduction that the fluency and readability of AI do not automatically guarantee the preservation of literary nuances. This study does not use statistical hypotheses because the design is qualitative and comparative, so the results are more accurately read based on the suitability of the pattern of findings with the initial framework of the study. The tendency of AI to produce clearer, more standardized languages is consistent with a number of contemporary studies, while the ability of human translators maintains figurative variation in line with the study of translation creativity.

However, some of the findings differ from the research of Gao et al. (2024), who showed that ChatGPT performs very strongly on poetry styles, and Farghal and Haider (2024), who found ChatGPT to be competitive with human translators. The difference can be explained by at least three factors. First, prompt configuration affects AI output; Gao et al. show that specific instructions regarding rhythm and rhyme improve the style of ChatGPT's results. Second, genre determines the complexity of translation: poetry, narrative fiction, and culturally charged texts have different demands. Third, the evaluation unit of this research focuses on metaphors, idioms, narrative voices, and cultural elements which are indeed vulnerable points for standardization. Thus, the difference in results is not a simple contradiction, but shows that AI capabilities are task-sensitive and context-dependent.

Research Gap and Research Contribution

The study's main contribution lies in its attempt to fill the gap between studies that evaluate the accuracy or fluency of AI and those that specifically address style or culture. Previous literature has largely isolated one dimension: creativity (Guerberof-Arenas & Toral, 2022), reader acceptance (Guerberof-Arenas & Toral, 2024), human-AI linguistic characteristics (Fu & Liu, 2024), prosodium and poetic creativity (Farghal & Haider, 2024), and stylistic variation (Zhou & Cheng, 2025). This research manuscript from the beginning identifies these limitations and positions meaning, style, and cultural representation as three dimensions that are analyzed in one qualitative framework.

Theoretically, the findings suggest that the evaluation of literary translation needs to move from a model that views quality as mere semantic equivalence to a multidimensional understanding. A translation can be informatively correct but lose style; be natural in the target language but generalize cultural identity; or it can retain the metaphor but change the level of readability. Thus, meaning, style, and culture are not three completely separate variables, but three layers that can change simultaneously as a result of a single translation decision. This contribution expands the discussion of human-AI translation by placing interpretive differences as the main unit of analysis.

In practical terms, the findings provide a basis for a more selective use of AI in literary translation. AI can be used when consistency, fluency, or explicitness are needed, but the results need to be examined specifically in the parts that contain metaphors, idioms, narrative voices, and culture-specific expressions. This is in line with the human-centered perspective of AI that places technology as a source of help without shifting interpretive responsibilities entirely away from humans.

Limitations and Recommendations for Advanced Research

Some limitations need to be noted. First, the study used only three human translators and one AI model. The fairly obvious variation of HT1-HT3 suggests that the character of individual translators can affect the results; Therefore, the findings cannot be generalized as the character of all human translators. Similarly, one AI model does not represent the entire LLM or AI translation system. Second, the selection of data is carried out purposively on the parts that contain metaphors, idioms, narrative styles, and cultural elements. The strategy is appropriate for close analysis, but it inherently chooses more complex parts of the text so that the results cannot be used to infer the performance of AI on the entire literary text.

Third, the study did not use statistical testing because its main purpose was qualitative interpretation. Consequently, terms such as "more consistent" need to be understood as qualitative patterns in the data, not as statistically tested differences. Future research can combine close textual analysis with corpus analysis or mixed methods, for example, calculating the frequency of the preservation of metaphors and cultural elements and then testing reader acceptance. The reception study approach is also relevant because Guerberof-Arenas and Toral (2024) show that preferences for human, machine, or post-edited translation can differ based on the language and reading patterns of the audience.

Fourth, advanced research needs to control the model version, access date, prompt, generation parameters, and number of repetitions of AI output. Zhou and Cheng (2025) explicitly show that the variability of AI output is a methodological problem and uses several generations to obtain a more stable picture of the model's style characteristics. Subsequent experiments can compare generic prompts with literary-aware prompts and culture-aware prompts to determine whether the explicit and standardization tendencies found in this study are the basic characteristics of the model or the result of the translation instructions used.

Considering the overall findings, this study shows that the central question in AI-based literary translation is no longer enough to be formulated as "Can AI translate literature?", but needs to be directed to "Which dimensions of literary meaning can AI preserve, which dimensions are transformed, and where is human interpretation still consequential?". AI has demonstrated a strong capacity to produce smooth and coherent translations, but the data of this study show that implicit meaning, figurative style, narrative voice, and cultural identity are still areas where human decisions make a substantive difference. Thus, the contribution of research lies not in the rejection of AI, but in the development of a model of literary translation that places the efficiency and consistency of AI side by side with the interpretation, creativity, and sensitivity of human culture.

CONCLUSIONS AND RECOMMENDATIONS

This research shows that human translators and artificial intelligence have different characteristics in literary translation. Human translators are more consistent in retaining implicit and contextual meanings, metaphors, figurative language, narrative voices, and cultural distinctiveness, while AI tends to produce more explicit, consistent, and standardized translations. These findings confirm that the quality of literary translation is not sufficiently assessed on the basis of semantic accuracy and fluency of language, but also that it is necessary to consider styles and cultural representations as integral parts of the meaning of the text.

Based on these findings, AI is more appropriately placed as a supporting technology in literary translation, especially to help generate translation alternatives and improve process efficiency. However, final decisions regarding metaphors, idioms, narrative nuances, contextual meanings, and cultural elements still require consideration by human translators. Thus, the human-AI collaborative approach is becoming a more relevant alternative because it combines the consistency and efficiency of technology with interpretive ability, creativity, and human cultural sensitivity.

ADVANCED RESEARCH

Further research is suggested comparing more translators, literary works, language pairs, and AI models. Follow-up studies can also use mixed methods or reader reception studies and compare generic prompts, literary-aware prompts, and culture-aware prompts to test the consistency of the quality of AI translations in different literary contexts.

ACKNOWLEDGMENT

The author expresses his gratitude to all human translators and translation experts who have participated in the process of translation, review, and validation of research interpretation. Appreciation was also conveyed to the State University of Surabaya and the Islamic University of Darul Ulum Lamongan for the academic support provided in the implementation of this research. The contribution of all parties supports the completion of research on the relationship between humans and artificial intelligence in literary translation.

REFERENCES

- Abdelhalim, S. M., Alsahil, A. A., & Alsuhaibani, Z. A. (2025). Artificial intelligence tools and literary translation: A comparative investigation of ChatGPT and Google Translate from novice and advanced EFL student translators' perspectives. *Cogent Arts & Humanities*, 12(1), Article 2508031. <https://doi.org/10.1080/23311983.2025.2508031>
- Abu Rumman, R., Haider, A. S., Yagi, S., & Al-Adwan, A. (2023). A corpus-assisted cognitive analysis of metaphors in the Arabic subtitling of English TV series. *Cogent Social Sciences*, 9(1), Article 2231622. <https://doi.org/10.1080/23311886.2023.2231622>
- Ahmad, M., & Wilkins, S. (2025). Purposive sampling in qualitative research: A framework for the entire journey. *Quality & Quantity*, 59(2), 1461–1479. <https://doi.org/10.1007/s11135-024-02022-5>
- Al Rousan, R., Jaradat, R., & Malkawi, M. (2025). ChatGPT translation vs. human translation: An examination of a literary text. *Cogent Social Sciences*, 11(1), Article 2472916. <https://doi.org/10.1080/23311886.2025.2472916>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, 56(3), 1391–1412. <https://doi.org/10.1007/s11135-021-01182-y>

- Chen, S., & Lin, Y. (2025). A multidimensional comparison of ChatGPT, Google Translate, and DeepL in Chinese tourism texts translation: Fidelity, fluency, cultural sensitivity, and persuasiveness. *Frontiers in Artificial Intelligence*, 8, Article 1619489. <https://doi.org/10.3389/frai.2025.1619489>
- Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: Human translation vs. AI large language models (Gemini and ChatGPT). *Cogent Social Sciences*, 10(1), Article 2410998. <https://doi.org/10.1080/23311886.2024.2410998>
- Fu, L., & Liu, L. (2024). What are the differences? A comparative study of generative artificial intelligence translation and human translation of scientific texts. *Humanities and Social Sciences Communications*, 11, Article 1236. <https://doi.org/10.1057/s41599-024-03726-7>
- 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, Article 835. <https://doi.org/10.1057/s41599-024-03363-0>
- Guerberof-Arenas, A., & Toral, A. (2022). Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces*, 11(2), 184–212. <https://doi.org/10.1075/ts.21025.gue>
- Guerberof-Arenas, A., & Toral, A. (2024). To be or not to be: A translation reception study of a literary text translated into Dutch and Catalan using machine translation. *Target*, 36(2), 215–244. <https://doi.org/10.1075/target.22134.gue>
- Haider, A. S., & Shuhaiber, R. (2024). Netflix English subtitling of idioms in Egyptian movies: Challenges and strategies. *Humanities and Social Sciences Communications*, 11, Article 949. <https://doi.org/10.1057/s41599-024-03327-4>

- Jiménez-Crespo, M. A. (2025). Human-centered AI and the future of translation technologies: What professionals think about control and autonomy in the AI era. *Information*, 16(5), Article 387. <https://doi.org/10.3390/info16050387>
- Kolb, W., Dressler, W. U., & Mattiello, E. (2023). Human and machine translation of occasionalisms in literary texts: Johann Nestroy's *Der Talisman* and its English translations. *Target*, 35(4), 540–572. <https://doi.org/10.1075/target.21147.kol>
- Saed, H., Haider, A. S., Abu Tair, S., & Asiri, E. (2024). Intrinsic managing and the English-Arabic translation of fictional registers. *Cogent Arts & Humanities*, 11(1), Article 2371659. <https://doi.org/10.1080/23311983.2024.2371659>
- Wang, G., & Xin, Y. (2024). An analytical framework for corpus-based translation studies. *Humanities and Social Sciences Communications*, 11, Article 1709. <https://doi.org/10.1057/s41599-024-04250-4>
- Yao, X., Kang, Y.-B., & McCosker, A. (2025). Missing the human touch? A computational stylometric analysis of GPT-4 translations of online Chinese literature. *Translation Spaces*, 14(2), 303–330. <https://doi.org/10.1075/ts.24043.yao>
- Zhao, P., Qi, W., Li, P.-J., & Li, P. (2024). Reconceptualizing the link between validity and translation in qualitative research: Extending the conversation beyond equivalence. *International Journal of Qualitative Methods*, 23, Article 16094069241260134.
- Zhou, P., & Cheng, J. (2025). Stylistic variation across English translations of Chinese science fiction: Ken Liu versus ChatGPT. *Frontiers in Artificial Intelligence*, 8, Article 1576750. <https://doi.org/10.3389/frai.2025.1576750>