Passerelle PP. 79-111
Vol. 14 No. 01 ISSN : 1112-6337 EISSN : 2716-8328

Clash of AIs: Which translates Saudi children's comics into English best?

Bahia, Zemni[1]
Translation Department, College of Languages, Princess Nourah bint Abdulrahman University / Saudi Arabia
baalzemni@pnu.edu.sa
Chaouki, Bounaas [2]
Department of Translation, Mohamed BOUDIAF University / Algeria
chaouki.bounaas@univ-msila.dz
Fadiah, Abdullah Alshehri [3]
Translation Department, College of Languages, Princess Nourah bint Abdulrahman University / Saudi Arabia
faaalshehri@pnu.edu.sa

Received: 13/6/2025, **Accepted:** 20/10/2025, **Published:** 12/12/2025

Abstract: This study investigates the performance of Large Language Models (LLMs)—ChatGPT, DeepSeek, and Claude—in translating literary texts, with a particular emphasis on children's literature. Recognizing the unique challenges posed by this genre, such as age-appropriate content, expressive language, and cultural nuances, the study employs Juliane House's Translation Quality Assessment (TQA) model to assess the quality of AI-generated outputs. Content and comparative analyses will be conducted on selected examples from Saudi children's comics translated by the three tools. The outputs will be evaluated based on linguistic accuracy, stylistic fidelity, and contextual appropriateness. The findings highlight strengths and limitations in each tool's performance in handling the intricate demands of literary texts, providing practical insights for translators and students regarding tool selection and effective usage.

Keywords: large language models (LLMs), machine translation (MT), artificial intelligence (AI), literary translation, children's literature.

Résumé : Cette étude examine la performance de trois grands modèles de langage — ChatGPT, DeepSeek et Claude — dans la traduction de textes littéraires destinés aux enfants. Reconnaissant les défis spécifiques posés par ce genre, tels que l'adéquation au public cible, le langage expressif et les nuances culturelles, l'étude met en œuvre le modèle d'évaluation de la qualité de traduction de Juliane House pour analyser la qualité des traductions générées par l'IA. Des analyses de contenu et des analyses comparatives sont menées sur des extraits choisis de bandes dessinées saoudiennes pour enfants traduites par les trois outils. Les traductions sont évaluées en fonction de leur précision linguistique, de leur fidélité stylistique et de leur pertinence contextuelle. Les résultats

mettent en évidence les forces et les limites de chaque outil face aux exigences complexes de la traduction littéraire, apportant des indications pratiques aux traducteurs et aux étudiants pour un choix éclairé et un usage efficace de ces technologies.

Mots-clés : grands modèles de langage, traduction automatique (TA), intelligence artificielle (IA), traduction littéraire, littérature pour enfants.

Introduction

In recent years, artificial intelligence (AI) has advanced significantly across various fields, performing intricate tasks with exceptional efficiency and minimal human involvement. In the domain of translation, developments such as deep learning, neural machine translation, and advancements in natural language processing (NLP) have positioned AI at the forefront of contemporary translation workflows. These technologies provide instantaneous bilingual translations across multiple languages and text formats, integrating functionalities such as speech recognition, grammatical and syntactical correction, and contextual adaptation. Large Language Models (LLMs), as advanced AI systems, utilize powerful machine learning algorithms to analyze, understand, and generate natural language, thereby enhancing users' ability to communicate intended meanings with greater precision.

Notwithstanding these advancements, identifying the most suitable AI translation technology remains a significant challenge. The burgeoning global market for AI translation, valued in the billions, is inundated with numerous competing programs that offer overlapping functionalities and claim superior performance, often resulting in user ambiguity. For professional translators and language practitioners, cultivating instrumental competence—defined as the effective use of resources and information and communication technology (Salamah, 2021)—is crucial for achieving accurate and efficient translation outcomes. Al-Batineh and Bilali (2017) found that instrumental competence was highly sought after in the workforce, with 50% of job requirements exceeding linguistic proficiency, which ranked second.

This study conducts a thorough quantitative and qualitative comparison of three leading LLM systems—ChatGPT, DeepSeek, and Claude—focusing on their efficacy in translating a culturally significant work of children's literature, specifically comics written by Fatima bint Yaâqoub Khoudja, a Saudi author. Literary works, especially those aimed at children, present significant challenges for AI due to their semantic complexity, stylistic nuances, and genre-specific conventions, all of which require meticulous preservation of meaning, tone, and educational significance. By assessing AI performance on this demanding text type, we seek to elucidate the strengths and limitations of each system within a sensitive and culturally rich framework.

We employ Juliane House's translation evaluation model to evaluate the translations. This model aligns well with our research aims, providing a comprehensive framework that addresses linguistic, stylistic, and contextual aspects of literary translation. Applying House's model to AI-generated translations enables the assessment of a human-centered evaluation tool on machine outputs, thereby revealing both the limitations of current AI systems and the adaptability of established theoretical frameworks. Unlike approaches that merely enumerate errors, House's methodology facilitates a nuanced analysis grounded in extensive translation expertise.

Accordingly, the objectives of this study are to:

• Assess and compare the translation accuracy and stylistic fidelity of ChatGPT, DeepSeek, and Claude in rendering a culturally embedded children's literary text.

• Identify the strengths and limitations of each AI translation system in handling the semantic and pragmatic complexities inherent in children's literature.

• Examine the extent to which these AI systems preserve genre-specific features, such as repetition, cultural references, and didactic elements, which are critical for young readers' engagement and comprehension.

• Provide practical recommendations for translators, educators, and students regarding the selection and effective use of AI translation tools in literary contexts.

Our research questions include:

• How effectively do ChatGPT, DeepSeek, and Claude translate literary texts, with particular focus on children's literature?

• What are the comparative strengths and limitations of these AI tools as evaluated through Juliane House's TQA model?

• To what extent do these systems maintain the stylistic fidelity and expressive qualities characteristic of children's literature?

• What practical recommendations can we offer translators and students for selecting and using AI translation tools effectively in literary contexts?

By addressing these questions through a mixed-methods approach combining quantitative error analysis and qualitative assessment, this study contributes to a deeper understanding of AI translation capabilities and their implications for cross-cultural literary transmission, especially in the sensitive domain of children's literature.

1. Literature review

The field of AI is booming and attracts many researchers, some aiming to highlight its shortcomings and improve its performance, while others seek to prove that it will never be able to replace humans. The field of translation is no exception to this trend. Machine translation, and more recently AI-based machine translation in the form of robots, is at the center of numerous discussions. Many studies are dedicated to this topic.

In 2024, Yu Yuxiu published a study titled "Application of Translation Technology Based on AI in Translation Teaching." The paper explored the use of AI-based translation technology in the classroom, employing a neural machine translation (NMT) algorithm to encode and decode the original text, generating the corresponding translation. Additionally, a statistical machine translation (SMT) algorithm was utilized to build the translation model, relying on statistical models to search for the optimal translation hypothesis through inference. This approach improved both the accuracy and readability of translations. Compared to traditional machine translation (MT), AI-based translation achieved a 97% accuracy rate, significantly higher than traditional MT, making it more suitable for teaching purposes. The improvement in students' translations was also evident, as reflected in their translation test scores. Furthermore, the teacher's satisfaction with the AI translation system was high, with an average score of 92. The findings highlighted that AI-powered translation teaching positively impacted students' translation skills and efficiency.

In the same year, Jiao et al. published a paper titled "Gradable ChatGPT Translation Evaluation," in which they proposed a generic taxonomy for gradable translation prompts. This taxonomy classified prompts based on expression type, translation style, part-of-speech (POS) information, and explicit statements. The paper emphasized the importance of a well-defined prompt taxonomy for translation tasks and identified essential design elements such as expression type, style, POS tagging, and few-shot examples. The researchers also explored how the explicit descriptions and contextual information within the gradable translation prompting taxonomy enhance the quality of prompts, improving translation accuracy, and reducing misunderstandings and ambiguity. The study concluded that these factors collectively improve translation quality and offer more precise guidance for ChatGPT-based translation tasks.

In the same year, Abkar Alkodimi et al. conducted a study titled "Human-AI Collaboration in Translation and Back Translation of Literary Texts." The study examined the impact of AI translation tools, such as ChatGPT, on the translation and back translation of literary texts. Using an experimental design within a qualitative framework, the researchers employed a translation test as the primary research tool. 80 English-major students from Imam Mohammed Ibn Saud Islamic University (IMSIU) were randomly assigned to four groups: two control groups and two experimental groups. These students were tasked with translating and back-translating an English short story, and qualitative data from their tests were analyzed through various comparisons. An independent samples t-test was used for statistical analysis. The results revealed that students who used AI tools produced better translations and back translations than those using traditional methods, with slightly superior performance in back translation.

In 2023, Li, Nawi and Sook Kang presented a study titled "Human-Machine Translation Model Evaluation Based on Artificial Intelligence Translation." This study analyzed attention mechanisms and the technical challenges of conventional translation models. The researchers proposed an AI-based translation model that produced high-quality, accurate translations, serving as a reference to further refine AI-driven translation technologies. The study showed that the human-machine translation model improved mismatches between texts and contexts, enhancing the accuracy and efficiency of intelligent recognition and expression. The results indicated that the language fluency score increased from 4.97 for traditional Statistical Machine Translation to 6.63 for the AI-based model. Consequently, the human-machine translation model improved translation efficiency, speed, precision, and accuracy, strengthening the connection between semantic characteristics and intelligent recognition. This model advanced intelligent recognition, offering more precise and high-quality translations for users and facilitating the automatic processing of natural language input and output.

As demonstrated in the aforementioned studies, the majority focus on assessing the effectiveness of translation robots in terms of the quality of the translations produced. While some studies have attempted to develop models for translation evaluation, none have employed a model like that of Juliane House. Although this model has been applied and studied in the context of human translation of literary and technical texts, it has yet to be utilized specifically for machine translation or for texts in children's literature. Furthermore, to the best of our knowledge, no study has conducted an analytical and comparative evaluation of three widely-used translation tools, such as ChatGPT, Google Gemini, and Claude. The absence of such combined research, coupled with the lack of a rigorous methodology for applying evaluation models to machine translations, underscores the originality and significance of this study in the realm of AI-based translation.

2. Theoretical framework

2.1 Large Language Models (LLMs)

LLMs represent a major advancement in AI, particularly within NLP. Built upon deep learning techniques and transformer-based architectures, these models are designed to understand and generate human language by analyzing vast amounts of text data (Jiao et al., 2023). They can analyze and produce language that reflects intricate syntactic structures, semantic nuances, and pragmatic uses, making them versatile across various language tasks.

LLMs are trained on extensive, diverse datasets through a two-step process: pre-training and fine-tuning. During pre-training, the model ingests billions or trillions of tokens from unstructured textual data, learning to predict and generate coherent sequences of words based on context. This self-supervised learning enables the model to grasp complex linguistic patterns, semantics, grammar, and conceptual relationships without explicit labeling of the data. Tokenization plays a key role in this stage, breaking down input text into numerical representations that the model processes to identify meaningful patterns (Zhu et al., 2024).

Following pre-training, LLMs undergo fine-tuning using methods such as Supervised Fine-Tuning (SFT) and Reinforcement Learning with Human Feedback (RLHF). RLHF, in particular, incorporates human evaluators' guidance to reduce biases and hallucinations, aligning model outputs more closely with human preferences and expectations (Jiao et al., 2023).

Popular LLMs available to the public include OpenAI's ChatGPT, DeepSeek, Claude and many other systems. In the following sections, these three models will be presented, as they constitute the tools used in this study:

2.1.1 ChatGPT

ChatGPT is an artificial intelligence chatbot designed to mimic human conversation and linguistic abilities. Developed by OpenAI, it stands for "Generative Pre-trained Transformer," a name that reflects its core functionality: processing requests and generating responses using advanced transformer architecture. This deep neural network enables ChatGPT to produce coherent, human-like text, making it versatile for tasks like answering questions, writing articles, composing emails, and even coding (Ferrag and Bentounsi, 2024; Kalla et al., 2023).

One of ChatGPT's standout features is its ability to understand and generate natural language with remarkable fluency. Built on GPT-3.5 and refined through RLHF, it has surprised users with its capabilities in both comprehension and text generation (Gao et al., 2023). For translation tasks, ChatGPT excels in high-resource languages, often matching or surpassing professional systems. However, it struggles with low-resource languages, where training data is scarce. What sets it apart is its flexibility—users can tweak prompts to guide translations, adjusting outputs to fit specific needs (Gao et al., 2023).

Despite its strengths, ChatGPT isn't without flaws. Its knowledge is limited to its training data, which means it may falter on niche or highly specialized topics. There's also the risk of bias, as the model can inadvertently reflect biases present in its training datasets (Kalla et al., 2023). Yet, its impact is undeniable. From aiding students with personalized explanations to transforming Information Technology (IT) interactions, ChatGPT has carved a significant niche in multiple fields (Kalla et al., 2023).

2.1.2 DeepSeek

DeepSeek refers to a family of LLMs developed by DeepSeek-AI, optimized for specialized tasks such as code generation and mathematical reasoning. Its architecture is based on a fine-grained mixture of experts, illustrated by DeepSeek-V3 (671B parameters), which dynamically activates 37B parameters per token to balance performance and efficiency (DeepSeek-AI et al., 2024). The creation process includes pre-training on multilingual corpus enriched in code (60%) and mathematics (10%), followed by alignment through mechanisms such

as Group Relative Policy Optimization (GRPO), using rewards based on response accuracy and format (Zhu et al., 2024; DeepSeek-AI et al., 2025).

DeepSeek incorporates innovations to reduce memory usage by 93.3% and accelerate inference—an indirect benefit for translations of long documents (DeepSeek-AI et al., 2024). In addition, techniques such as Chain-of-Thought (CoT), embedded in DeepSeek-R1, improve sequential reasoning, applicable to the translation of technical texts requiring rigorous logic (DeepSeek-AI et al., 2025). These advances position DeepSeek among the leaders of open-source LLMs, combining performance, efficiency and versatility.

2.1.3 Claude

Claude, developed by Anthropic, is another powerful AI LLM, designed with a focus on speed and efficiency. Part of the Claude family, which includes Claude Instant, Claude 1, and Claude 2, this model shines in real-time applications like customer service and information retrieval. Its automated structure allows it to process text rapidly, delivering quick responses without sacrificing quality (Ferrag and Bentounsi, 2024).

Recent advancements, particularly with Claude 3 Opus, highlight its prowess in low-resource machine translation. Unlike ChatGPT, Claude excels in translating languages with limited training data, making it a valuable tool for multilingual accessibility. It also generates realistic synthetic data, which can enhance traditional neural machine translation systems, pushing the boundaries of accuracy and productivity (Ferrag and Bentounsi, 2024).

In practice, Claude delivers near-literary quality translations, especially in English. Minor imperfections may arise, but they rarely disrupt the overall coherence or meaning of the text. Its growing utility in academic and research settings underscores the rapid progress in AI and its practical applications (Ferrag and Bentounsi, 2024).

2.2 LLMs as translation tools

As mentioned before, AI has emerged as a revolutionary tool across various domains due to its remarkable efficacy and its ability to emulate aspects of human cognition. Within translation, LLMs such as ChatGPT have become increasingly valuable. Researchers like Ferrag and Bentounsi (2024) highlight the versatility of LLMs in performing a range of translation tasks, including translating, proofreading, and bridging cultural gaps between languages. These models

achieve notable speed and accuracy, especially when handling source texts (ST) with moderate complexity, familiar content, and when low-resource languages are not involved. Designed to optimize real-time response efficiency, LLMs are particularly suited for applications requiring rapid and reliable translation outputs (Ferrag and Bentounsi, 2024).

2.3 Children's literature and its translation

Literature, in its broadest sense, reflects human experiences, emotions, and cultures through storytelling. Within this vast landscape, children's literature stands out as a unique genre crafted specifically for young readers. As Diachuk (2024) notes, children's literature employs diverse forms—fairy tales, fables, poetry, comics and novels—all tailored to captivate young minds while nurturing their imagination and cognitive growth. Thus, children's books are more than just simplified stories; they are carefully designed to align with the emotional, linguistic, and moral development of their audience. For instance, repetition or cumulative patterns (Prodanović Stankić and Begonja, 2024), reinforces language acquisition and memory. Moreover, humor, wordplay, and vivid imagery help young reader widen their imagination and construct personalities.

Thus, children's books do more than entertain; they foster empathy and global awareness. Erten (2011) highlights how these stories help children "*understand and respect other cultures*" (cited in Rençberler, 8: 2021). Yet, the challenge lies in maintaining this cultural richness without overwhelming young readers. For instance, Dawson et al. (2021) found that children's books use lexically denser language than everyday speech.

Translating children's literature is then far to be an easy task. As (Prodanović Stankić and Begonja, 2024) assert translating puns, rhymes, or nonsense words demands creativity. A direct translation might lose the whimsy, so translators often recreate effects using target-language equivalents. Also proper nouns, folklore elements, or idioms rarely have direct equivalents. Strategies like domestication (adapting to the target culture) or foreignization (retaining source-culture elements) come into play (Venuti, 1995). Moreover, in the case of Picture books, text and visuals merge, requiring translators to adapt language without disrupting the visual narrative (Masi, 2021). For example, a translated joke must align with the illustration's context. Interestingly, translators of children's literature wear multiple hats: they're linguists, cultural mediators, and even educators. As Kučiš (2016) argues, they must navigate socio-cultural contexts to ensure stories resonate with young readers (cited in Prodanović Stankić and Begonja, 2024).

Accordingly, translating for children isn't just about swapping words—it's about preserving magic across cultures. Whether adapting a funny rhyme or a fantastical creature, translators must blend creativity with sensitivity, ensuring stories remain both authentic and accessible. As we've seen, this demands not only linguistic skill but also a deep understanding of how children read, learn, and imagine.

2.4 Julian House's TQA Model

Juliane House first introduced her TQA model in 1977, marking a foundational development in translation studies (Shakernia, 2014). Acknowledging the evolving nature of translation theory and practice, House revised her model significantly in 1997. This revision expanded the framework to incorporate elements of discourse analysis, drawing on Halliday's systemic functional linguistics as well as insights from the Prague School, speech act theory, and pragmatics (Vallès, 2014). Building on this interdisciplinary foundation, House updated the model again in 2015 to include recent research from corpus linguistics, psycholinguistics, neurolinguistics, intercultural communication, and globalization studies (Al-Aizari, 2023).

House's model is grounded primarily in Hallidayan systemic-functional theory, providing a robust framework for analyzing language use across multiple levels (Naidj and Motahari, 2019). It moves beyond superficial linguistic comparison by evaluating translations across three hierarchical levels: language/text, register, and genre (Faryad et al., 2021). At the language/text level, it examines grammatical, lexical, cohesion and coherence and syntactic elements. The register level analyzes situational variables through three components: field (the subject matter and lexical specificity), tenor (social relationships, formality, and attitudes between participants), and mode (the channel and style of communication) (Faryad et al., 2021). Genre represents the highest level, focusing on the overall communicative purpose and the cultural conventions that shape expectations for different text types (Faryad et al., 2021).

Central to House's model is the principle of functional equivalence, which holds that a translation should perform the same communicative function and evoke similar responses in its audience as the ST (Shakernia, 2014). She defines translation as essentially a process, rather than a product, emphasizing the dynamic nature of rendering meaning from one language to another (Cappelle, 2011). The model distinguishes between overt translation—where the source culture remains visibly present and the target audience is secondary—and covert

translation, which adapts the text to function naturally within the target culture, often employing cultural filters (Faryad et al., 2021; Beyranvand et al., 2024).

House's model also offers a detailed error classification system that differentiates two main error types: overt and covert errors (Hedayati and Yazdani, 2020). Overt errors are clear and measurable, including omissions, additions, distortions, and breaches of target language norms that affect meaning and readability (Hedayati and Yazdani, 2020). Covert errors, on the other hand, involve subtler mismatches at the level of register or genre, such as inappropriate tone, cultural misalignments, or failure to maintain the original's communicative intent, which may go unnoticed but undermine the translation's effectiveness (Hedayati and Yazdani, 2020; Vallès, 2014). Cultural filtering errors, a form of covert error where adaptations are either insufficient or excessive, represent a particular challenge in balancing fidelity and accessibility, with research showing they may constitute a significant portion of translation mistakes (Al-Aizari, 2023).

The evaluation process involves constructing detailed profiles of both source and target texts (TT), then systematically comparing these profiles to identify mismatches that signify errors (Vallès, 2014). This approach allows for a comprehensive understanding of translation quality beyond surface accuracy, taking into account semantic, pragmatic, and textual equivalences (Shakernia, 2014). House's model is especially valuable in literary translation, where the translator must balance faithfulness to the source with cultural resonance and readability for the target audience—a task complicated by the public's resistance to literature in translation (Landers, 2001; Al-Aizari, 2023).

By bridging theoretical insights and practical assessment tools, House's continually evolving model remains a cornerstone in translation quality assessment. Its interdisciplinary foundation and systematic methodology offer translators and scholars a nuanced framework for analyzing why certain translations succeed in preserving the ST's communicative functions while others falter. The model's ongoing revisions demonstrate its adaptability to contemporary challenges in translation studies, making it indispensable for evaluating both human and machine-generated translations today.

3. Methodology

To achieve the objectives outlined above, this study adopts a mixed-methods approach that integrates both qualitative and quantitative data collection and analysis. The methodology is structured as follows:

Three LLMs—ChatGPT, DeepSeek, and Claude—are selected for comparison. The selection aims to provide a diverse representation of available LLMs, each with distinct functionalities and performance capabilities.

The corpus consists of a culturally embedded children's comics entitled مَرْحى لِلْقَهْوَةِ السُّعُودِيَّةِ! authored by Fatima bint Yaâqoub Khoudja. Selected extracts from the corpus are input into each AI translation tool, with no human intervention during the translation process to preserve the integrity of AI-generated outputs. Each tool processes the same ST to establish a consistent basis for comparison.

The resulting translations undergo evaluation using Juliane House's TQA model, which assesses translation quality across multiple dimensions:

- Linguistic aspects: accuracy, grammar, and syntax
- Stylistic aspects: tone, register, and naturalness
- Contextual aspects: appropriateness for the target audience, cultural adaptation, and fidelity to the original text

To systematically organize error analysis, a comprehensive table will be conceived distinguishing two primary error types:

- Overt errors: those that are clearly identifiable and directly affect meaning or comprehension
- Covert errors: subtler inaccuracies that may retain grammaticality but alter nuance or stylistic fidelity

Each error type will be further divided into relevant subtypes, such as lexical choice, language, coherence and cohesion, field, tenor, mode and genre, enabling detailed categorization and comparison across the LLM outputs.

The qualitative analysis will complement the quantitative data by presenting a table that juxtaposes the ST with the TTs generated by each LLM. This table will be followed by a thorough explanation highlighting specific translation choices, errors, and culturally significant elements, allowing for an in-depth understanding of each system's performance. This will also include examining how each system handles children's literature-specific features such as repetition, didactic elements, and expressive language.

4. Analysis

4.1 Quantitative analysis

To evaluate the translation quality of the three AI models—Claude, ChatGPT, and DeepSeek—we adopted a quantitative approach aimed at identifying and counting errors. This method distinguishes between overt and covert errors. Overt errors, as previously defined, are clearly observable deviations from linguistic norms, including lexical inaccuracies, grammatical mistakes, orthographic faults, and issues related to cohesion and coherence. Covert errors, by contrast, are subtler and often more difficult to detect; they typically relate to the appropriateness of tone, mode, and genre. By categorizing and quantifying these two types of errors, we seek to provide a clearer comparison of the systems' performances and to highlight their respective strengths and weaknesses.

Table 01: Quantitative Comparison of the Three Systems' Translations

Errors		ChatGPT	DeepSeek	Claude
Overt	Lexical Choice	23	23	4
	Language	0	22	0
	Coherence And Cohesion	1	0	0
	Field	0	0	0
Covert	Tenor	2	2	1
	Mode			0
	Genre	40	30	2

As shown in the table above, all three models displayed a combination of strengths and weaknesses, although some performed better than others in translating the provided story. Claude clearly outperformed both ChatGPT and DeepSeek, producing fewer errors overall. We observe that DeepSeek frequently made errors related to language, meaning, and genre, while ChatGPT's errors were mainly associated with meaning and genre. In contrast, Claude's output contained only a small number of errors. The following section offers a detailed qualitative analysis to explain the nature of these errors.

4.2 Qualitative analysis

This section presents a qualitative analysis of the translations produced by ChatGPT, DeepSeek, and Claude. It is intended to complement the quantitative findings by offering explanations for specific errors observed in the translations.

- **Extract 1**

Table 02: Extract 01 – Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
مَرْحى لِلْقَهْوَةِ السُّعُودِيَّةِ!	Hurray for Saudi Coffee!	Hooray for Saudi Coffee!	Hooray for Saudi Coffee!

In this extract, ChatGPT translated the underlined word as *hurray*, while both DeepSeek and Claude rendered it as *hooray*. It is important to note that hooray and hurray are variant forms of the same interjection—less commonly also rendered as hurrah (Merriam-Webster, n.d.)—typically used in children's literature to convey excitement. Although both variants are correct, hurray is less frequently used. Therefore, we consider ChatGPT's choice a covert genre-related error, as it diverges from the more conventional and familiar form expected in children's narratives.

- **Extract 2**

Table 03: Extract 02 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
اِسْتَيْقَظَ القِرْدُ (مَيْمُون) مُتَشَوِّقًا لِشَيْءٍ مَا! وَمَا إِنْ فَتَحَ الْبَابَ، وَرَأَى الْمَاعِزَ (فَايِز)، وَبَيْنَ يَدَيْهِ كَوْمَةٌ مِنْ حَبَّاتِ الْبُنِّ الْخَوْلَانِيِّ؛ حَتَّى تَهَلَّلَ، وَهَتَفَ بِفَرَحٍ: **مَرْحى لِلْقَهْوَة! مَرْحى لِلْقَهْوَة!**	The monkey, **Maymoon**, woke up bubbling with excitement! As soon as he opened the door and saw **Fayez** the goat holding a heap of **Khawlani coffee beans**, he beamed and shouted with joy: **"Hurray for coffee! Hurray for coffee!"**	Monkey (Maimoon) woke up excitedly, eager for something! As soon as he opened the door and saw Goat (Fayiz) standing there with a pile of Khulani coffee beans, his face lit up, and he cheered	The monkey (Maymoun) woke up excited about something! As soon as he opened the door and saw the goat (Fayez) with a pile of Khawlani coffee beans in front of him, his face lit up and he shouted joyfully: "Hooray for

		joyfully: "Hooray for coffee! Hooray for coffee!"	coffee! Hooray for coffee!"

In this extract, we observe that DeepSeek omitted the definite article at the beginning of the sentence. This can be considered a covert genre-related error, as children's literature—particularly when introducing characters—typically retains articles such as "the" to support language development by reinforcing noun forms and syntactic patterns. Additionally, DeepSeek opted for alternative transliterations of the monkey and goat's names, using *Maimoon* and *Fayiz*, while ChatGPT and Claude used *Maymoon* and *Fayez*. We consider this an overt orthographic error, as DeepSeek's versions deviate from the Arabic pronunciation and are less phonetically accurate. Moreover, ChatGPT omitted the parentheses around character names, whereas DeepSeek and Claude preserved them.

For the word (مُتَشَوِّقًا), meaning "excited," ChatGPT rendered it as *bubbling with excitement*—a figurative, idiomatic expression implying enthusiasm and joy (Reverso, n.d.). While expressive, this translation introduces a stylistic exaggeration and can be classified as an over-translation, thus constituting an overt error. DeepSeek used *excitedly, eager*, combining two terms where one (excited) would have sufficed. This redundancy also represents an overt error due to unnecessary addition. Claude, in contrast, used excited, which we find both accurate and contextually appropriate. The author's original phrasing was straightforward and did not rely on idiomatic language, making Claude's choice the most fitting.

DeepSeek also misspelled (الْخَوْلَانِيّ)—a term derived from the place name Khawlan—and failed to apply diacritical signs appropriately. This constitutes another overt orthographic error. In contrast, both ChatGPT and Claude correctly rendered the term.

Finally, the expression (تَهَلَّلَ، وَهَتَفَ بِفَرَحٍ) was translated differently by each system. We believe that DeepSeek and Claude outperformed ChatGPT here. ChatGPT used to beam, which typically means to smile, leading to an under-translation that diminishes the intensity of the original expression. Both DeepSeek and Claude used his face lit up, which more accurately reflects the meaning of (تَهَلَّلَ)—a phrase often used to describe a face suddenly glowing with joy, particularly in poetic or emotional contexts (Almaany, n.d.). Regarding (هَتَفَ بِفَرَحٍ), ChatGPT rendered it as *shouted with joy*, a construction that leans more formal or literary (verb + prepositional phrase). Claude's choice, *shouted joyfully* (verb +

adverb), is more concise and typical of narrative prose in children's literature. Therefore, ChatGPT's rendering can be seen as a covert tenor-related error. DeepSeek's use of *cheered joyfully* attempts to highlight the subject's action, but the verb to cheer already implies joy. The addition of joyfully is thus redundant and constitutes an overt error due to unnecessary repetition.

- **Extract 3**

Table 04: Extract 03 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
اِسْتَمْهَلَهُ الْمَاعِزُ (فَايِز) قَائِلًا: "مَهْلًا! مَهْلًا! لَنْ نَسْتَطِيعَ احْتِسَاءَ الْقَهْوَةِ الآنَ! لَا بُدَّ مِنْ تَحْمِيصِ الْبُنِّ"!	But **Fayez** gently said: "Wait! Wait! We can't drink the coffee yet. First, we have to roast the beans!"	Goat (Fayiz) stopped him, saying: "Wait! Wait! We can't drink the coffee yet! First, we must roast the beans!"	The goat (Fayez) stopped him, saying: "Wait! Wait! We can't drink the coffee now! We must roast the beans first!"

At the beginning of this passage, we observe that ChatGPT omitted the word *goat*, retaining only the name *Fayez*, whereas the other two systems consistently included both the noun and the name throughout the translation. This choice by ChatGPT can be classified as a covert genre-related error, as it overlooks one of the key features of children's literature: cumulative repetition. This stylistic device is often employed to reinforce linguistic structures and aid memory development. Additionally, both ChatGPT and DeepSeek omitted the definite article—a pattern repeated throughout the text—while Claude preserved it. This omission further contributes to a covert genre-related error, as explained earlier.

In the same passage, ChatGPT also failed to convey the meaning of *(استمهله)* and instead added the adverb *gently* to modify *said*. While the tone may be inferred from context, the original text does not explicitly describe the manner of speech. As a result, this constitutes an overt error through the unwarranted addition of meaning. In contrast, both DeepSeek and Claude translated the term more faithfully.

We also note a difference in how the temporal adverb *(الآن)* was rendered. ChatGPT and DeepSeek translated it as *yet*, while Claude chose *now*. We consider Claude's version more appropriate, as it is more direct, common, and aligned with

the natural tone of children's storytelling. This discrepancy represents an overt error in tenor for ChatGPT and DeepSeek.

Finally, when translating *(لابد)*, DeepSeek and Claude used *must*, while ChatGPT chose *have to*. Although both modals express obligation, *must* is more appropriate in this context, as it conveys logical necessity rather than imposed rules. Therefore, ChatGPT's choice reflects an overt lexical error.

- **Extract 4**

Table 05: Extract 04 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
تَعَاوَنَ الْقِرْدُ (مَيْمُون) وَالْمَاعِزُ (فَايِز) فِي تَحْمِيصِ الْبُنِّ الْخَوْلَانِيِّ فِي الْمِحْمَاسِ الْحَدِيدِيِّ عَلَى نَارٍ هَاااادِئَةٍ، حَتَّى صَارَ ذَهَبِيَّ اللَّوْنِ، وَفَاحَتْ رَائِحَتُهُ الْجَمِيلَةُ! هَتَفَ الْمَاعِزُ (فَايِز) بِفَرَحٍ: **مَرْحى لِلْقَهْوَة! مَرْحى لِلْقَهْوَة!**	So, Maymoon and Fayez worked together, roasting the **Khawlani beans** in an iron roaster over a **gentle, slow flame**. The beans turned a **golden color**, and a wonderful aroma filled the air! Fayez cheered: **"Hurray for coffee! Hurray for coffee!"**	Monkey (Maimoon) and Goat (Fayiz) worked together to roast the Khulani beans in an iron roaster over a gentle flame. The beans turned golden and released a delightful aroma! Goat (Fayiz) cheered: "Hooray for coffee! Hooray for coffee!"	The monkey (Maymoun) and the goat (Fayez) worked together to roast the Khawlani coffee beans in an iron roaster over a very gentle fire, until they turned golden in color and their beautiful aroma filled the air! The goat (Fayez) shouted with joy: "Hooray for coffee! Hooray for coffee!"

In this example, the three systems offered different translations for the phrase *(عَلَى نَارٍ هَاااادِئَةٍ)*. ChatGPT rendered it as *gentle, slow flame*, DeepSeek as *gentle flame*, and Claude as *very gentle flame*. Since *gentle flame* is a commonly accepted expression in cooking, storytelling, and soothing imagery, we find Claude's translation the most appropriate. ChatGPT's use of two adjectives (*gentle* and *slow*) instead of a single, more precise expression introduces redundancy, constituting an overt error due to unnecessary addition. Moreover, it failed to convey the intensification implied by the elongated spelling in the ST. Similarly, DeepSeek omitted an adverb indicating degree, thereby overlooking the

expressive intent of the original. Both ChatGPT and DeepSeek, therefore, committed overt lexical errors.

For the expression *(صَارَ ذَهَبِيَّ اللَّوْنِ)*, DeepSeek outperformed ChatGPT and Claude by employing a more idiomatic and concise formulation, one that is commonly used in culinary and narrative contexts. In contrast, the other two systems provided less natural renderings. As a result, we classify ChatGPT and Claude's translations as covert tenor-related errors, as they failed to reflect the stylistic conventions expected in this genre.

At the end of the passage, we also observe different renditions of *(وَفَاحَتْ رَائِحَتُهُ الْجَمِيلَةُ)*. Here, ChatGPT produced the most effective version, capturing both the intended nuance and the mood of the original by foregrounding the subject (*aroma*) and maintaining a vivid, sensory tone appropriate for children's storytelling. While DeepSeek's version was semantically accurate, it emphasized the action of release over the sensory impact, resulting in a less poetic effect. Claude's translation, although focused on the aroma, used the adjective *beautiful*, which is uncommon in English to describe scents. Alternatives such as *pleasant*, *delightful*, *wonderful*, or *aromatic* would have been more appropriate. Therefore, we consider both DeepSeek and Claude to have committed overt lexical errors due to imprecise word choice.

- **Extract 5**

Table 06: Extract 05 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
قَالَ القِرْدُ (مَيْمُون): "مَهْلًا! مَهْلًا! لَنْ نَسْتَطِيعَ احْتِسَاءَ الْقَهْوَةِ الآنَ! لَابُدَّ مِنْ طَحْنِ الْبُنِّ"! سَارَ القِرْدُ (مَيْمُون)، وَالْمَاعِزُ (فَايِز) إِلَى الْفَرَسِ (كَحِيلَة)؛ لِيَسْتَعِيرَا مِنْهَا الْهَاوُنَ النُّحَاسِيَّ. مَا إِنْ رَأَتْهُمَا الْفَرَسُ (كَحِيلَة)، وَشَمَّتْ رَائِحَةَ الْبُنِّ الْخَوْلَانِيِّ الْمُحَمَّص؛ حَتَّى عَرَفَتْ مَا يُرِيدَانِ، وَهَتَفَتْ بِفَرَح: **مَرْحى**	But Maymoon reminded him: "Wait! Wait! We still can't drink the coffee. We need to grind the beans!" They went to **Kaheela** the horse to borrow her **brass mortar**. As soon as Kaheela saw them and smelled the roasted	Monkey (Maimoon) interrupted: "Wait! Wait! We can't drink the coffee yet! We need to grind the beans!" Monkey (Maimoon) and Goat (Fayiz) hurried to Horse (Kheelah) to	The monkey (Maymoun) said: "Wait! Wait! We can't drink the coffee now! We must grind the beans first!" The monkey (Maymoun) and the goat (Fayez) went to the mare (Kaheela) to borrow her brass mortar.

لِلْقَهْوَة! مَرْحى لِلْقَهْوَة!	Khawlani coffee, she knew what they were up to and called out joyfully: **"Hurray for coffee! Hurray for coffee!"**	borrow her brass mortar. When Horse (Kheelah) saw them and smelled the roasted Khulani beans, she instantly knew what they wanted and cheered: "Hooray for coffee! Hooray for coffee!"	As soon as the mare (Kaheela) saw them and smelled the aroma of the roasted Khawlani coffee beans, she knew what they wanted and shouted joyfully: "Hooray for coffee! Hooray for coffee!"

For the verb *(سار)*, DeepSeek used *to hurry*, which typically means to move quickly toward a destination (Reverso, n.d.). This constitutes an overt lexical error, as the ST simply conveys the act of walking without specifying speed. A more accurate translation would be *to go*, which was the choice made by both ChatGPT and Claude.

Additionally, both ChatGPT and DeepSeek used the term *horse* to refer to the character *(كَحِيلَة)*, while the ST explicitly describes her as a female horse, or *mare*. Claude accurately captured this distinction. This oversight can be considered an overt error resulting from an imprecise transfer of meaning. It is particularly significant in the context of children's literature, where gender representation may be intentional. The author may have deliberately assigned specific genders to the animal characters to reflect an inclusive Saudi social setting, challenging stereotypical views of strict gender separation in the region. Alternatively, the choice of a mare may carry symbolic meaning in Arabic culture. The automatic use of the masculine form by the machine raises a broader ethical concern, highlighting the potential for embedded bias in AI-generated translations.

Finally, DeepSeek misspelled the name of the horse, which we classify as an overt orthographic error.

- **Extract 6**

Table 07: Extract 06 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
تَعَاوَنَ القِرْدُ (مَيْمُون) مَعَ الْمَاعِزِ (فَايِز) وَالْفَرَسِ (كَحِيلَة) فِي طَحْنِ الْبُنِّ الْخَوْلَانِيِّ الْمُحَمَّص، حَتَّى صَارَ نَاعِمَ الْمَلْمَسِ، شَهِيَّ الرَّائِحَةِ!	So, Maymoon, Fayez, and Kaheela worked together to grind the roasted coffee until it was **soft to the touch and smelled delicious!**	Together, Monkey (Maimoon), Goat (Fayiz), and Horse (Kheelah) ground the roasted beans until they became smooth and fragrant!	The monkey (Maymoun) worked together with the goat (Fayez) and the mare (Kaheela) to grind the roasted Khawlani coffee beans until they became smooth to the touch and delightfully fragrant!

ChatGPT and DeepSeek used all three animal characters as the subject of the sentence, emphasizing that they collaborated equally. In contrast, Claude selected the monkey as the grammatical subject and included the other animals in a prepositional complement, which slightly shifts the focus and places more emphasis on the monkey. We consider this an overt error due to an imprecise rendering of meaning, as it unintentionally alters the narrative balance among the characters. Furthermore, DeepSeek translated *(تَعَاوَنَ)* using only the word *together*, which is somewhat imprecise and fails to fully capture the sense of mutual cooperation inherent in the original. This constitutes an overt lexical error. On the other hand, both ChatGPT and Claude used *worked together*, a more accurate and contextually appropriate rendering, especially given that cooperation is a central theme in the story and arguably one of its moral lessons.

When translating *(نَاعِمَ الْمَلْمَس)*, an Arabic collocation, both ChatGPT and Claude opted for idiomatic expressions that preserved the stylistic tone and expressive impact typical of the genre. DeepSeek, however, used only *smooth*, which, although not incorrect, lacks the stylistic richness of the original and fails to align with the literary style of children's literature. We therefore classify it as a covert error linked to genre insensitivity.

Finally, Claude outperformed the other two systems in rendering *(شَهِيَّ الرَّائِحَةِ)* by using a natural and widely accepted English collocation. ChatGPT chose *delicious*, a term more commonly associated with taste rather than smell, making it an awkward fit in this context and thus an overt lexical error. DeepSeek used *fragrant*, which is appropriate but fails to convey the intensity and appeal suggested by the original. This makes DeepSeek's rendering an overt error due to imprecise expression.

- **Extract 7**

Table 08: Extract 07 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
قَالَتِ الْفَرَسُ (كَحِيلَة): "مَهْلًا! مَهْلًا! لَنْ نَسْتَطِيعَ احْتِسَاءَ الْقَهْوَةِ الآنَ! لَابُدَّ مِنْ غَلْيِ الْبُنِّ مَعَ الْمَاءِ"! سَارَ الْقِرْدُ (مَيْمُون)، وَالْمَاعِزُ (فَايِز)، وَالْفَرَسُ (كَحِيلَة) إِلَى الْغَزَالَةِ (أَصَالَة). مَا إِنْ رَأَتْهُمُ الْغَزَالَةُ (أَصَالَة)، وَشَمَّتْ رَائِحَةَ الْبُنِّ الْخَوْلَانِيِّ الْمُحَمَّصِ الْمَطْحُونِ؛ حَتَّى هَتَفَتْ بِفَرَحٍ: **مَرْحى لِلْقَهْوَة! مَرْحى لِلْقَهْوَة!**	Then Kaheela said: "Wait! Wait! We still can't drink the coffee. We have to **boil it with water** first!" Maymoon, Fayez, and Kaheela headed to visit **Asala the gazelle**. When Asala saw them and sniffed the ground, roasted coffee in the breeze, she joyfully exclaimed: **"Hurray for coffee! Hurray for coffee!"**	Horse (Kheelah) said: "Wait! Wait! We can't drink the coffee yet! We need to boil the beans with water!" Monkey (Maimoon), Goat (Fayiz), and Horse (Kheelah) rushed to Gazelle (Asalah). When Gazelle (Asalah) saw them and caught the scent of the roasted, ground beans, she cheered: "Hooray for coffee! Hooray for coffee!"	The mare (Kaheela) said: "Wait! Wait! We can't drink the coffee now! We must boil the coffee with water!" The monkey (Maymoun), the goat (Fayez), and the mare (Kaheela) went to the gazelle (Asala). As soon as the gazelle (Asala) saw them and smelled the aroma of the roasted, ground Khawlani coffee, she shouted joyfully: "Hooray for coffee! Hooray for coffee!"

In this example, the author depicts a scene in which Asala encounters her friends and perceives the aroma of coffee, without specifying whether she notices the scent intentionally. Therefore, the verb *to smell* serves as a more accurate translation, since *to sniff* implies taking a brief, deliberate breath through the nose, typically to detect or examine a scent. Consequently, ChatGPT's translation fails to convey the intended meaning, as it introduces a detail absent from the ST, resulting in an overt error due to imprecise interpretation. In contrast, DeepSeek employs the phrase *caught the scent*, which, while more verbose and using the more general term *scent* rather than *smell*, can be regarded as acceptable. Furthermore, ChatGPT adds the phrase *in the breeze*—a reference to a gentle wind—not present in the original text. This addition misrepresents the desert climate, which is characterized by heat and dryness, and thus constitutes an overt error arising from an inaccurate rendering of meaning.

- **Extract 8**

Table 09: Extract 08 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
تَعَاوَنَ القِرْدُ (مَيْمُون) مع الْمَاعِز (فَايِز)، وَالْفَرَس (كَحِيلَة)، وَالْغَزَالَةُ (أَصَالَة) فِي وَضْعِ الْبُنِّ الْخَوْلَانِيِّ الْمُحَمَّصِ الْمَطْحُونِ مَعَ الْمَاءِ فِي إِبْرِيقٍ فَوْقَ نَارٍ هَاااادِئَةٍ، حَتَّى بَدَأَ الْغَلَيَانَ، وَوَصَلَتْ رَائِحَتُهُ الطَّيِّبَةُ أَنْفَ الصَّقْرِ (جَسُور)! اِسْتَيْقَظَ الصَّقْرُ (جَسُور) مِنْ قَيْلُولَتِهِ، وَنَفَضَ جَنَاحَيْهِ فِي سُرُورٍ، وَهَتَفَ بِفَرَحٍ: **مَرْحى لِلْقَهْوَة! مَرْحى لِلْقَهْوَة!**	Together, the four friends placed the coffee and water in a pot over a **gentle flame**. As the water started to boil, the coffee's rich scent reached the nose of **Jasoor the falcon**. Jasoor awoke from his nap, stretched his wings with delight, and cried out: **"Hurray for coffee! Hurray for coffee!"**	The friends worked together to simmer the ground Khulani beans with water in a pot over a gentle flame. The rich aroma soon reached Eagle (Jasoor), who was napping nearby! Eagle (Jasoor) woke up, flapped his wings happily, and cheered: "Hooray for coffee! Hooray for coffee!"	The monkey (Maymoun) worked together with the goat (Fayez), the mare (Kaheela), and the gazelle (Asala) to put the roasted, ground Khawlani coffee with water in a pot over a very gentle fire, until it began to boil, and its wonderful aroma reached the nose of the falcon (Jasour)! The falcon (Jasour) woke up from his nap, flapped his wings

			in delight, and shouted joyfully: "Hooray for coffee! Hooray for coffee!"

At the beginning of this passage, unlike Claude, both ChatGPT and DeepSeek omit the references to the animals, substituting them with "the four friends" and "the friends," respectively. These translations constitute covert errors stemming from a disregard for genre conventions, as they eliminate a key feature of children's literature—namely, cumulative repetition—which the original author appears to employ deliberately.

We also observe an overt error resulting from an omission by DeepSeek: the phrase حَتَّى بَدَأَ الْغَلَيَانَ is left out, although this does not significantly affect the meaning since water naturally boils over flame. Notably, ChatGPT and Claude retain this phrase. Similarly, DeepSeek omits the word "nose," which is present in the ST, whereas the other two systems preserve it. While DeepSeek may assume that the scent's arrival at a nose is implicit in the olfactory context, the original author includes it intentionally to educate young readers about the senses and the corresponding organs. This omission thus constitutes both an overt error—due to inaccurate meaning rendering—and a covert error—resulting from a failure to respect the didactic style typical of children's literature.

Regarding the phrase اِسْتَيْقَظَ الصَّقْرُ (جَسُور) مِنْ قَيْلُولَتِهِ, we contend that DeepSeek performs better than the other two systems by employing the continuous tense, which implies that the scent reached the falcon's nose while it was napping and awakened it. In contrast, ChatGPT and Claude commit overt grammatical errors by failing to convey this aspect. However, DeepSeek mistranslates الصَّقْرُ as "eagle," which represents a different species; this constitutes an overt lexical error. Similarly, ChatGPT errs by misrepresenting the falcon's movement, depicting it as extending its wings outward rather than moving them up and down, as the ST specifies.

- **Extract 9**

Table 10: Extract 09 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
طَارَ الصَّقْرُ (جَسُور) إِلَى أَصْدِقَائِهِ، وَلَكِنَّهُ حِينَمَا اقْتَرَبَ، سَمِعَ	He flew toward his friends. But as he approached, he heard	He flew off to gather his friends but heard	The falcon (Jasour) flew to his friends, but

الْغَزَالَةَ (أَصَالَة) تَقُولُ: "مَهْلًا! مَهْلًا! لَنْ نَسْتَطِيعَ احْتِسَاءَ الْقَهْوَةِ الآنَ! لَابُدَّ مِنْ بَعْضِ حَبَّاتِ الْهَيْلِ"! حَلَّقَ الصَّقْرُ (جَسُور) عَالِيًا، وَفِي ثَوَانٍ قَلِيلَةٍ كَانَ رَاجِعًا وَبَيْنَ مَخَالِبِهِ حَفْنَةٌ مِنَ الْهَيْلِ.	Asala saying: "Wait! Wait! We still can't drink the coffee. We need some **cardamom pods**!" With powerful wings, Jasoor soared high into the sky. In just a few seconds, he returned with a **handful of cardamom in his talons**.	Gazelle (Asalah) say: "Wait! Wait! We can't drink the coffee yet! We need some cardamom pods!" Eagle (Jasoor) soared high and returned moments later with a handful of cardamom clutched in his talons.	when he approached, he heard the gazelle (Asala) saying: "Wait! Wait! We can't drink the coffee now! We need some cardamom pods!" The falcon (Jasour) soared high, and in just a few seconds he was back with a handful of cardamom in his talons.

In this passage, DeepSeek omits حِينَمَا اقْتَرَبَ, resulting in an overt error likely due to the assumption that it is self-evident the falcon cannot hear its friend unless nearby. ChatGPT introduces the phrase *with powerful wings*, which does not appear in the ST; this constitutes a covert error arising from an unjustified addition of meaning. DeepSeek translates وَفِي ثَوَانٍ قَلِيلَةٍ as *moments later*, but this choice is slightly inaccurate and imprecise, amounting to an overt error. While *moments later* conveys a short duration, it remains vague—potentially spanning several seconds or even minutes—whereas the original author specifically indicates a matter of seconds.

- **Extract 10**

Table 11: Extract 10 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
أَضَافَ الْأَصْدِقَاءُ الْهَيْلَ إِلَى الْبُنِّ الْخَوْلَانِيِّ الْمُحَمَّصِ الْمَطْحُونِ فِي الْإِبْرِيقِ مَعَ الْمَاءِ؛ فَاخْتَلَطَا مَعًا، وَأَكْمَلَا الْغَلَيَانَ. أَخِيرًا، جَهَزَتِ الْقَهْوَةُ؛ فَهَتَفَ الْجَمِيعُ فِي فَرْحَةٍ	The friends added the cardamom to the coffee pot and let it **simmer gently together**. Finally, the coffee was ready! Everyone cheered in	The friends added the cardamom to the pot and let it boil until the flavors blended perfectly. Finally, the	The friends added the cardamom to the roasted, ground Khawlani coffee in the pot with the water; they mixed together

كَبِيرَةٍ: **"مَرْحى لِلْقَهْوَةِ السُّعُودِيَّةِ"!**	delight: **"Hurray for Saudi coffee!"**	coffee was ready! Everyone cheered: "Hooray for Saudi Coffee!"	and completed the boiling process. Finally, the coffee was ready, and everyone shouted in great joy: "Hooray for Saudi coffee!"

Both ChatGPT and DeepSeek omit the phrase مَعَ الْمَاءِ؛ فَاخْتَلَطَا مَعًا، وَأَكْمَلَا الْغَلَيَانَ, possibly assuming that it was already mentioned in earlier passages and therefore need not be repeated. In contrast, Claude faithfully renders all the meanings as presented in the ST. The omissions by ChatGPT and DeepSeek represent overt errors due to the loss of essential meaning, as well as covert errors reflecting a failure to uphold genre conventions—specifically, the characteristic repetition found in children's literature, which serves to reinforce young readers' memory and establish a rhythmic reading experience.

- **Extract 11**

Table 12: Extract 11 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
اِسْتَمْهَلَهُمُ الصَّقْرُ (جَسُور) قَائِلًا: "مَهْلًا! مَهْلًا! لَنْ نَسْتَطِيعَ احْتِسَاءَ الْقَهْوَةِ الآنَ! لَابُدَّ مِنَ الدَّلَّةِ وَالْفِنْجَانِ!" وَإِذَا بِالْجَمَلِ (سَنُوم) قَادِمٌ. وَمَا إِنْ رَآهُمُ الْجَمَلُ (سَنُوم)؛ حَتَّى تَهَلَّلَ وَرَحَّبَ بِهِمْ، "مَرْحَبًا أَلف. مَرْحَبًا أَلف"، وَأَخْرَجَ مِنْ خُرْجِهِ دَلَّةً نُحَاسِيَّةً، وَفَنَاجِينَ خَزَفِيَّةً مُزَخْرَفَةً، وَأَيْضًا سَلَّةَ تَمْرٍ كَبِيرَةً.	But Jasoor flapped his wings and said: "Wait! Wait! We still need a **dallah (traditional coffee pot)** and **finjaan (cups)**!" Just then, **Sanoom the camel** arrived. When he saw them, his face lit up with joy. "Welcome, a thousand welcomes!" he said warmly. He opened his	But Eagle (Jasoor) interrupted: "Wait! Wait! We can't drink the coffee yet! We need the dallah (coffee pot) and finjan (cups)!" Just then, Camel (Sanoom) arrived. Camel (Sanoom) greeted them warmly: "A thousand	The falcon (Jasour) stopped them, saying: "Wait! Wait! We can't drink the coffee now! We need the coffee pot (dallah) and cups!" And there came the camel (Sanoum). As soon as the camel (Sanoum) saw them, he beamed and welcomed them warmly,

	saddlebag and brought out a shiny **copper dallah**, beautifully decorated **ceramic cups**, and a big **basket of dates**.	welcomes!" He pulled out a gleaming copper dallah, ornate ceramic finjan cups, and a large basket of dates from his saddlebag.	"Welcome a thousand times. Welcome a thousand times," and he took out from his saddlebag a brass coffee pot (dallah), decorated ceramic cups, and also a large basket of dates.

ChatGPT introduces an additional detail by describing the falcon as flapping its wings, despite the ST offering no information about the falcon's actions. This constitutes an overt error caused by an unjustified addition. Regarding the term الدَّلَّةِ, a traditional coffee pot, all three systems borrow the original word and provide explanatory notes. This approach can be viewed as a covert error stemming from a disregard for the literary genre and style. A similar pattern and error appear in the translations of الْفِنْجَانِ by ChatGPT and DeepSeek.

Moreover, ChatGPT adds the adjective *shiny* and the adverb *beautifully*, while DeepSeek uses the adjective *beaming*—none of which are present in the ST. These additions represent overt errors due to unwarranted elaboration. In this passage, Claude commits an overt lexical error by translating the material as *brass* instead of *copper*, which is more appropriate given that traditional functional coffee pots typically exhibit a reddish-brown hue characteristic of copper.

Finally, both DeepSeek and Claude choose *large* to describe the basket, whereas ChatGPT opts for *big*. We consider the choice of *large* a tenor error leading to a covert error, as *big* conveys a simpler, more approachable, and familiar tone better suited to young readers.

- **Extract 12**

Table 13: Extract 12 Translations by the Three Systems

ST	ChatGPT	DeepSeek	Claude
تَحَلَّقَ الْجَمِيعُ حَوْلَ الدَّلَّةِ، مُسْتَمْتِعِينَ بِالطَّعْمِ اللَّذِيذِ،	Everyone gathered around the pot,	The friends gathered around	Everyone gathered around

وَالرَّائِحَةِ الزَّكِيَّةِ لِلْقَهْوَةِ السُّعُودِيَّةِ.	enjoying the **delicious taste** and **aromatic smell** of **Saudi coffee**.	the dallah, savoring the delicious taste and enchanting aroma of Saudi coffee.	the coffee pot, enjoying the delicious taste and the pleasant aroma of Saudi coffee.

Although ChatGPT previously used the term *dalla*, in this passage it opts for *pot*, resulting in a cohesion and coherence error—an overt error due to inconsistency in terminology. Additionally, both ChatGPT and Claude translate the action of appreciating the coffee as *enjoy*, whereas DeepSeek uses *savour*. We consider DeepSeek's choice superior, as *to savour* conveys a deeper, more nuanced appreciation, which aligns logically with the effort invested and the joy expressed by the friend throughout the story.

The analysis revealed that all three systems committed errors, including overt lexical and orthographic mistakes (e.g., DeepSeek's "eagle" for "falcon"), covert genre violations (e.g., CHATGPT's omission of cumulative repetition), and stylistic mismatches (e.g., ChatGPT's over-translation of "متشوقا" as "bubbling with excitement"). Claude consistently outperformed the others in semantic precision and genre awareness, while ChatGPT and DeepSeek exhibited recurring issues with additions, omissions, and cultural misalignment. These findings underscore the need for tailored fine-tuning to address context-specific demands.

5. Discussion

Our findings reveal a marked disparity in the performance of the three evaluated LLMs—Claude, ChatGPT, and DeepSeek—when translating a culturally embedded children's comics. Claude's output was markedly superior, with only seven total errors, while ChatGPT and DeepSeek produced 66 and 75 errors respectively. These outcomes align with Ferrag and Bentounsi's (2024) conclusion that Claude generates near-literary quality translations, characterized by a coherent structure and a strong grasp of semantic nuance. Claude's automated structure and its enhanced processing capacity likely contribute to its ability to maintain linguistic integrity and stylistic fidelity (Ferrag and Bentounsi, 2024).

In contrast, the high number of grammatical and syntactic errors found in DeepSeek (n=22), compared to none in the other systems, suggests fundamental structural limitations. These results mirror the difficulties noted in prior research where systems like ChatGPT and Google Translate struggled with idiomatic expressions and abbreviations, undermining overall acceptability (Cahyaningrum,

2024). DeepSeek's recurrent misspelling of proper names further supports Diachuk's (2024) assertion that translating proper nouns and culturally loaded terms often presents significant challenges due to the absence of direct equivalents or the risk of alienating young readers.

The prominence of covert errors across all systems is particularly noteworthy. While overt errors are more easily identifiable, covert ones—such as mistranslations that retain grammaticality—are subtler and potentially more disruptive in literary contexts. These findings highlight the relevance of Al Rousan, Sami Jaradat and Malkawi's (2023) estimation that even high-performing models like ChatGPT achieve only 77.9% accuracy in literary translation, leaving room for nuance loss and cultural misrepresentation.

Despite their shortcomings, all three models demonstrated some ability to capture and convey cultural nuances. This is encouraging, given the cultural and textual-linguistic constraints that often limit the accurate rendering of elements in children's literature (Bulut, 2006, as cited in Rençberler, 2021). Our analysis confirms that, while Claude better navigated these constraints, both ChatGPT and DeepSeek attempted to address cultural subtleties, albeit inconsistently.

Lexical errors—particularly omissions and additions—were common in both ChatGPT and DeepSeek. This pattern resonates with Cahyaningrum's (2024) observations regarding ChatGPT's tendency toward natural but sometimes over-adaptive outputs. These inaccuracies are especially critical in children's literature, where word choice directly impacts clarity, engagement, and educational value (Yalçın and Aytaş, 2002, as cited in Rençberler, 2021).

Ultimately, the results point to Claude's stronger ability to render texts suitable for a child audience in both form and function. Its apparent competence in low-resource machine translation (Ferrag and Bentounsi, 2024) further underscores its potential in cross-cultural contexts, particularly when translating literature for young readers who rely heavily on repetition, clarity, and culturally resonant content for comprehension and engagement (Dawson et al., 2021; Prodanović Stankić and Begonja, 2024).

However, these technological gains must be critically examined. As Ferrag and Bentounsi (2024) caution, the increasing reliance on AI for translation raises ethical concerns related to authorship, neutrality, and data use. The translation of children's literature, in particular, involves high stakes: texts not only entertain but also shape values, language development, and cultural perceptions. Therefore,

while Claude's performance is promising, its integration into educational or publishing contexts requires careful consideration.

6. Solutions and implications

Our findings underscore the urgent need to refine AI translation systems for culturally rich children's literature, where linguistic precision and genre-specific conventions are paramount. To address the disparities observed, developers should prioritize enhancing systems' capacity to preserve stylistic features such as cumulative repetition, lexical consistency, and didactic elements critical to young readers' comprehension and engagement. Specifically, integrating genre-sensitive training data and fine-tuning models on children's literature corpora could improve their ability to maintain narrative rhythm and educational intent.

Given Claude's relative success, its architecture and processing capabilities offer a promising foundation. We recommend leveraging Claude's strengths by developing hybrid workflows that combine its outputs with targeted human post-editing, particularly focused on cultural nuances and subtle semantic distinctions. Such an approach can mitigate covert errors—often overlooked in automated evaluation—while preserving the efficiencies gained through automation.

For systems like ChatGPT and DeepSeek, addressing structural limitations and lexical inaccuracies should be a priority. This entails refining syntactic parsing modules and enhancing context-aware semantic interpretation, especially for idiomatic and culturally specific expressions. Moreover, implementing mechanisms to handle proper nouns and culturally embedded terms with greater sensitivity will reduce alienation risks for young readers and maintain narrative authenticity.

Finally, these results suggest broader implications for cross-cultural literary exchange. The demonstrated capacity of AI systems to navigate complex linguistic and cultural landscapes, albeit unevenly, opens avenues for more inclusive and diverse children's literature worldwide. However, achieving this potential demands continuous collaboration between AI developers, literary scholars, educators, and cultural experts to ensure translations honor both the ST's integrity and the target audience's needs.

Conclusion

This study has highlighted significant disparities in the performance of three leading LLMs—ChatGPT, DeepSeek, and Claude—when tasked with translating

a culturally embedded piece of children's literature. Our combined quantitative and qualitative analyses reveal that Claude demonstrates a markedly superior ability to preserve linguistic accuracy, stylistic fidelity, and cultural nuances, thereby producing translations that better align with the demands of young readers and the conventions of children's literature. In contrast, ChatGPT and DeepSeek, while capable of capturing some cultural subtleties, exhibit frequent overt and covert errors, particularly regarding lexical choices, syntactic structure, and genre-specific features such as repetition and didactic elements.

The implications of our research emphasize the importance of enhancing LLMs with genre-sensitive training and hybrid human-LLM workflows, particularly when translating culturally and pedagogically sensitive texts such as children's literature. Moreover, cultivating instrumental competence among translators and language practitioners remains essential for navigating the complex landscape of LLM-based translation technologies effectively.

Ultimately, while Claude's performance signals promising advances toward literary-quality LLM translation, ethical considerations surrounding authorship, cultural integrity, and the educational impact of translated children's literature must guide the integration of such technologies into professional and educational settings. Future research should focus on exploring these ethical issues in greater depth and conducting studies with larger and more diverse corpora to better understand the capabilities and limitations of LLMs in literary translation.

Acknowledgements

This research received grant no. (625/2024) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

Bibliography

ABKAR ALKODIMI Khaled, ABDULRHMAN ALQAHTANI Osama and AL-WASY Baleigh Qassim, (2024), "Human-AI collaboration in translation and back translation of literary texts", in, Journal of Social Studies 30 (n°2), pp. 173–192. https://doi.org/10.20428/jss.v30i2.2404

ALMAANY, (n.d.), Almaany.com dictionary [en ligne]. Available on: https://www.almaany.com. Accessed on 28 May 2025.

AL ROUSAN Rafat, SAMI JARADAT Raghad and MALKAWI Mona, (2025), "ChatGPT translation vs. human translation: an examination of a literary text", in,

Cogent Social Sciences 11(n°1), pp. 1-21. https://doi.org/10.1080/23311886.2025.2472916
BEYRANVAND Rezvan, NAZARI Ali and GHASEMI MOUSAVI Seyed Esmaeil, (2024), "Quality assessment of the Persian translation of the novel Fardghan by Youssef Zaydan based on the Julian House's Model", in, Translation Researches in the Arabic Language and Literature 13(n°29), pp. 255-284.
BULUT Alev, 2006, "Joyce Carol Oates'un "Düş Ağı" öyküsünün çevirisinde fantastik özelliklerin getirdiği kısıtlamalar", *in*, Akşit GÖKÜÜRK (eds.), Akşit Göküürk'ü Anma Toplantısı: Yazında ve Çeviride Fantastik, İstanbul, İstanbul Üniversitesi Basim Yay, pp. 153-164.
CAHYANINGRUM Ika Oktaria, (2024), "Chat GPT vs Google Translate for translation», in, AICONICS Proceedings, n°1, pp. 450-456.
DAWSON Nicola et al., (2021), "Features of lexical richness in children's books: comparisons with child-directed speech", in, Language Development Research 1(n°1), pp. 9-53. https://doi.org/10.34842/Swel-yk94
DEESEEK-AI, (2024), "DeepSeek-V2: a strong, economical, and efficient mixture-of-experts language model", in, arXiv [en ligne]. Available on: https://arxiv.org/abs/2405.04434v5.
DEESEEK-AI, (2025), "DeepSeek-R1: incentivizing reasoning capability in LLMs via reinforcement learning", *in*, arXiv [en ligne]. Available on: https://arxiv.org/abs/2501.12948v1.
DIACHUK Nataliia, (2024), "Challenges and strategies in children's literature translation", in, Current Issues of Foreign Philology 21, pp. 38-43. https://doi.org/10.32782/2410-0927-2024-21-6
ERTEN Asalet, (2011), Çocuk yazını çevirisine yaklaşımlar, Ankara, Hacettepe Yayıncılık.
FARYAD Waqas, AZIZ Asif and BIBI Tahseen, (2021), "The Application of House' Model of Translation Quality Assessment on Pakistani Fictionist Saadat Hasan Manto's short story", in, İlköğretim Online 20(n°5), pp. 1308-1318. doi:10.17051/ilkonline.2021.05.146
FERRAG Faris and BENTOUNSI Ikram Aya, (2024), "The use of artificial intelligence in academic translation tasks: case study of Chat GPT, Claude and Gemini", in, RAZLE 11(n°2), pp. 173-192.
HEDAYATI Elham and YAZDANI Mohammad, (2020), "Translation Quality Assessment Based on House's Model: English translations of Iran's Supreme Leader letters to European youth", in, Journal of Foreign Language Teaching and Translation Studies 5(n°2), pp. 99-118. doi: 10.22034/efl.2020.230069.1039
JIAO Wenxiang et al., (2023), "ParroT: translating during chat using large language models tuned with human translation and feedback", in, arXiv (Cornell University) [en ligne]. https://doi.org/10.48550/arxiv.2304.02426.

KALLA, Dinesh et al., (2023), “Study and Analysis of Chat GPT and its Impact on Different Fields of Study”, in, International Journal of Innovative Science and Research Technology 8 (n°3), pp. 827-833.
KUČIŠ Vlasta, (2016), Translatologija u teoriji i praksi, Zagreb, Hrvatsko komunikološko društvo, Nonacom.
LI Ruichao, Nawi Abdullah Mohd and KANG Myoung Sook, (2023), “Human-machine Translation Model based on Artificial Intelligence Translation”, in, Journal for ReAttach Therapy and Developmental Diversities, 6(n°10s), pp. 1122–1129. https://jrtdd.com/index.php/journal/article/view/1900
MASI Silvia, (2021), “Translating non-fiction picturebooks for children across age groups and languages: the case of informative books on geography in English and Italian”, in, Translation Matters 3(n°2), pp. 60-74. https://doi.org/10.21747/21844585/tm3_2a4
MERRIAM-WEBSTER, (n.d.), Merriam-Webster.com dictionary [en ligne]. Available on: https://www.merriam-webster.com. Accessed on 27 May 2025.
NAIDJ Sara and MOTAHAI Masoud Seyed, (2019), “Translation Quality Assessment in the Literary Text based on House Model”, in, Journal of Humanities and Social Sciences Studies 1(n°2), pp. 19-27. https://al-kindipublisher.com/index.php/jhsss/article/view/158
PRODANOVIĆ STANKIĆ Diana and BEGONJA Helga, (2024), “Expressive language in translation of books for children: a corpus-based study”, in, Círculo de Lingüística Aplicada a la Comunicación 100, pp. 235-243. https://dx.doi.org/10.5209/ciac.80532
RENÇBERLER Alize Can, (2021), “Unveiling the translation procedures of fantastic items and names in children’s literature”, in, Turkish Studies - Language and Literature 16(n°1), pp. 423-437. https://doi.org/10.47454/TurkishStudies.49218
REVERSO, (n.d.), Reverso dictionary [en ligne]. Available on: https://www.reverso.net. Accessed on 25 May 2025.
SALAMAH Dania, (2021), “Translation competence and translator training: a review”, in, International Journal of Linguistics, Literature and Translation 4(n°3), pp. 276-291. https://doi.org/10.32996/ijllt.2021.4.3.29
SHAKERNIA Shabnam, (2014), “Study of House’s Model of Translation Quality Assessment on the short story and its translated text”, in, Global Journal of Human-Social Science XIV(n°III), pp. 9-14.
VENUTI Lawrence, (1995), The translator’s invisibility: a history of translation, London, Routledge.
YALÇIN Alemdar and AYTAŞ Aytas, (2002), Çocuk edebiyatı, Ankara, Akçağ Yayınları.

ZHU Qihao et al., (2024), "DeepSeek-Coder-V2: breaking the barrier of Closed-Source Models in Code Intelligence", in, arXiv (Cornell University) [en ligne]. Available on: https://doi.org/10.48550/arxiv.2406.11931.