

ARTICLE

ChatGPT in English-Arabic Translation: Learner Perceptions and the Treatment of Humour and Empathy

Muneer Hezam Alqahtani ^{1*} , Arif Ahmed Mohammed Hassan Al-Ahdal ² 

¹ Department of Curriculum and Instruction, College of Education, King Faisal University, Al-Ahsa 31982, Saudi Arabia

² Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah 52571, Saudi Arabia

ABSTRACT

Although English is truly the global lingua franca today, machine translation (MT) of paralinguistic elements like humour and empathy is not totally seamless in the English-Arabic language pair. This study examined the effectiveness of ChatGPT, a generative artificial intelligence (AI) model, in translating these elements, and how Saudi translation students who use AI tools for translation perceive its performance. The study applied a mixed-methods approach by gathering quantitative data through a Likert-scale questionnaire from 46 undergraduate translation students at Qassim University, Saudi Arabia, and qualitative data was gathered through an analysis of texts translated by ChatGPT. The questionnaire was validated and tested, demonstrating high reliability with a Cronbach's alpha of 0.670. The findings showed that, overall, translation output in ChatGPT was strong, and users have high confidence in the tool for formal and academic translation, with a high average score of 4.00. However, its performance in conveying humour (mean = 3.33) and empathy (mean = 3.37) is moderate, which can lead to neutral or culturally mismatched results. The study concludes that while ChatGPT is a useful educational tool for improving accessibility and motivating translation practice, it needs significant improvements in cultural and emotional intelligence. Additionally, the language dataset can be enhanced with rich Arabic content. Despite

*CORRESPONDING AUTHOR:

Muneer Hezam Alqahtani, Department of Curriculum and Instruction, College of Education, King Faisal University, Al-Ahsa 31982, Saudi Arabia;
Email: mhalqahtani@kfu.edu.sa

ARTICLE INFO

Received: 17 March 2025 | Revised: 19 September 2025 | Accepted: 23 September 2025 | Published Online: 10 November 2025
DOI: <https://doi.org/10.30564/fls.v7i12.11682>

CITATION

Alqahtani, M.H., Al-Ahdal, A.A.M.H., 2025. ChatGPT in English-Arabic Translation: Learner Perceptions and the Treatment of Humour and Empathy. *Forum for Linguistic Studies*. 7(12): 637–650. DOI: <https://doi.org/10.30564/fls.v7i12.11682>

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

the mixed results, it can be recommended that educators critically incorporate ChatGPT into curricula, while simultaneously focusing on post-editing and cross-cultural comparison exercises for best practices.

Keywords: Artificial Intelligence; ChatGPT; Humour; Empathy; Machine Translation; English-Arabic Translation; Research Work

1. Introduction

ChatGPT can handle the lexical and structural aspects of English-Arabic translation, but it fails to capture essential cultural and emotional nuances, such as humour and empathy, that define effective human communication. The value proposition of this study is to bridge computational linguistics and translation pedagogy by addressing the critical gap in understanding AI's handling of pragmatically complex translation through the novel lens of intersemiotic transfer and by integrating learners' perceptions. The results show that although ChatGPT is an effective tool for formal and structural translation, it has significant problems with pragmatic and cultural translation^[1,2].

Paralinguistic and prosodic features of language, including tone, emotion, and humour, play an essential role in effective communication^[3]. They are cues that contribute to the authenticity of communication and assist in achieving clarity in the reception of the message, especially in bilingual or multilingual settings. In translation, paralinguistic elements are fundamental to ensure that messages maintain their full meaning and emotional depth when translated into another language^[4].

Yet, translating such elements, particularly humour and empathy, presents considerable hurdles for large language models (LLMs) and AI systems like ChatGPT^[1,5]. These features often rely on context and implicature, which may not be expressed in the text. For example, humour often depends on wordplay, cultural references, or tone^[2], while empathy is shown through nuanced emotional expressions that may not translate directly into another language or culture^[6]. This is a pronounced challenge in the English-Arabic language pair due to notable linguistic and cultural distinctions^[7,8]. While previous studies have assessed the lexical and grammatical accuracy of neural machine translation (NMT) systems^[9], fewer have examined their ability to handle pragmatic and paralinguistic features, particularly humour and empathy. This gap is critical in educational contexts, like Saudi Arabia,

where English as a Foreign Language (EFL) students increasingly depend on AI tools for learning and translation^[10].

The main goal of translation is to attain semantic equivalence between a source and a target language. Such a goal follows from any sensible definition of translation, differentiating it from other language processing. There are, however, still many open questions regarding how one should determine an adequate level of equivalence—questions that take on an especially heightened significance in the age of large language models and machine translation tools. Complete, perfect equivalence is impossible to achieve; this is no new paradox. Whether translation occurs through the use of human professionals or through machines, elements like rhythm, symbolism, and pragmatic inference need to be carefully translated and often then post-edited to come close to an optimal result. Machine translation technologies more often than not generate many kinds of imperfect equivalence. These imprecisions may still be perfectly serviceable for specific practical uses: information transfer, approximate comprehension, or quick content localisation. However, only the target language and culture specialist can properly assess the fidelity and functional adequacy of a particular translation. That is to say, there is no one “best” translation in the abstract, but rather translations best adapted to specific communicative intentions. When the expressive goals of a text involve humour and empathy, the requirements for what constitutes an adequate translation change markedly. Evaluation of translation output is, therefore, fundamentally pragmatic. The success of a rendition rests on its purpose and its intended audience. For factual, utilitarian texts, fidelity to propositional content may suffice. In the case of texts aimed at provoking laughter, warmth, or emotional connection, the translator needs to re-create not just propositional content but also affective impact and social intention. Human communication—and its translation—serves as a realisation of individual cognition and cultural schemata. Whereas verbal material can be translated directly from one language to another, issues arise when non-verbal or culturally coded

material is at the heart of the message. Humour and empathy are reliant on a shared cultural toolkit: social conventions, idioms, intertextual allusions, timing, and performative markers. The punchline of a joke may rely on phonetic ambiguity, cultural stereotype, or historical reference that simply will not make it through a literal shift.

Empathy, too, is conveyed through subtle lexical options, deixis, honorifics, and the sensitivity of politeness strategies—factors that vary cross-linguistically and cross-culturally. Any translation agent, human or computerised, therefore stands to gain from a culturally aware strategy. For machine outputs in particular, post-editing by culturally aware translators is usually required to restore pragmatic intention and emotional richness. Practically, this means a layered process: preliminary machine translation can offer a semantic equivalence scaffold, but functional equivalence proper—especially for humour and empathy—needs human hands. The translator needs to diagnose what pragmatic functions are involved and choose strategies (compensation, adaptation, explanation, or substitution) that maintain the original's illocutionary force. In summary, successful translation of humour and empathy is not so much about word-for-word correspondence as about recapturing communicative impacts in the target culture. Only through placing primary emphasis on cultural competence and purposeful assessment can translation serve as a bridge between worlds of language.

This study evaluates ChatGPT's ability to handle complex cultural and pragmatic elements, particularly humour and empathy, in English–Arabic translation. While it performs well in formal and structural tasks, it struggles with cultural nuances and pragmatic depth, highlighting limitations in current AI models. Utilising an intersemiotic perspective, the research emphasises the necessity of cultural and contextual awareness for effective translation—areas where AI currently lacks proficiency. The integration of learner perspectives provides actionable insights for educators and developers, advocating for enriched training data and critical AI literacy in curricula. The conclusions denote the importance of human cultural intelligence in translation tasks while recommending a collaborative human–AI approach in future studies for a deeper understanding of AI's role in translation education and practice.

Understanding ChatGPT's strengths and weaknesses in translating emotional content is crucial for developing

effective teaching strategies and improving AI systems. This study aims to fill a perceptible gap by evaluating ChatGPT's performance in translating humour and empathy between English and Arabic and examining Saudi translation students' views on its utility. The study aims to answer the following research questions:

1. How effective is ChatGPT in translating humour between English and Arabic?
2. How accurately does ChatGPT convey empathy in English-Arabic translation?
3. What do Saudi translation learners think about ChatGPT's performance in translating humour and empathy?

2. Literature Review

2.1. AI, Machine Translation, and the Rise of ChatGPT

The development of NMT marked a significant improvement over earlier rule-based and statistical systems, offering better fluency and contextual accuracy^[9]. ChatGPT, a large language model (LLM) developed by OpenAI, represents a shift from dedicated NMT systems to more generalised models that can perform translation as one of many tasks^[1]. Its strength lies in using vast training data to produce contextually appropriate translations. However, its performance varies across language pairs: It performs well in high-resource languages like English but struggles with languages with less online presence, such as Arabic, often prioritising fluency over accuracy^[11].

2.2. The Challenge of Paralinguistic Features in Translation

Translation goes beyond mere word-for-word transfer to include pragmatics and cultural nuances^[3]. Paralinguistic elements like humour and empathy are deeply rooted in cultural contexts, making them particularly challenging for AI to translate^[2,4,12]. Humour often relies on wordplay, sarcasm, and cultural references, while empathy is expressed through nuanced emotional language and shared understandings. The available AI systems, including ChatGPT, lack deep pragmatic awareness, leading to translations that may be semantically correct but emotionally flat or culturally inappropriate^[13].

2.3. The English-Arabic Context and Pedagogical Implications

The linguistic and cultural gap between English and Arabic is significant by virtue of their belonging to two distinct language families. Arabic communication is high-context and is interspersed with a liberal presence of religious references, proverbs, and indirectness to maintain its unique socio-cultural fabric^[7,14–18]. Translating English humour into Arabic requires an understanding of these cultural norms to prevent misinterpretation^[2]. Likewise, expressions of empathy in Arabic often have religious connotations or implications (e.g., “إنّا لله وإنّا إليه راجعون”) (To Allah we belong and to Him we shall return), many of which do not have direct counterparts in English. Consequently, AI translation tools are seen as one option that is also supported by the teacher community, which recognises their efficacy in vocabulary and grammar support. However, lack of caution and over-reliance on these is fraught with other problems^[10]. Therefore, it is essential to explore how well these tools perform in the context of cultural and emotional nuances. This understanding is crucial for creating effective learning strategies that use AI while addressing its shortcomings.

2.4. The Culture and Nature of Arabic Dialogue

Arabic dialogue, which incorporates both artistic and intellectual exchange, reflects the rich cultural, social, and linguistic heritage of the Arab world^[19]. Typical Arabic conversation is characterised by nuanced, emotional expression underpinning the need for cross-cultural understanding^[4]. Its foundations are in indirectness, hospitality, and respect. Culturally, politeness and circumvention are favoured in place of directness and confrontation^[7,20,21]. Phrases like “Perhaps you might consider AlaItibar Anna”, or “If it’s convenient” (Ala Rahtak) are used to soften requests, and disagreements may be delayed or hidden behind ambiguity to maintain harmony.

Proverbs, religious references, and poetic accompaniments are commonly employed in speech, making the language eloquent. The rhythmic, almost ceremonial feel of formal discourse is a result of classical Arabic’s influence on Modern Standard Arabic, even in informal dialects. People can introduce faith into everyday conversations by using phrases like Insha’Allah (God willing) or Masha’Allah (God

has willed it).

Arabic communication is also very high-context, relying on tone, pauses, and gestures that convey meaning. Silence can express disapproval or thoughtfulness, and a raised voice can express passion rather than anger. Social hierarchy also affects conversation; titles (like Sheikh, Sayyed, or Dr.) and avoiding too-familiar language highlight deference to elders and authority figures, though interruptions are acceptable in lively discussions in formal settings. Furthermore, emotional gestures, repetition, and dramatic statements (like “By my life, I promise you!” (which in some cases are considered more sincere than swearing by God’s name) are used to emphasise sincerity. Instead of being perceived as disturbing, this expressiveness is seen as an indication of genuine engagement. Arabic conversation is essentially a dance of tradition and connection, where words tell history, relationships set the tone, and the unsaid is just as significant as the spoken. Engaging in it entails negotiating a culture where language is not merely a tool for communication, but rather, a celebration of identity, religion, and social bonds.

3. Method

3.1. Research Design and Participants

This study used an explanatory sequential mixed-methods design. In the quantitative phase, perceptual data was gathered, followed by a qualitative phase involving an analysis of translation output by ChatGPT. A purposive sample of 46 undergraduate students (21 males, 25 females) majoring in English at Qassim University, Saudi Arabia, was engaged. Participants were aged 18–25 (mean = 21.7, SD = 1.8) and were proficient users of ChatGPT. **Table 1** summarises the demographic information of the participants.

3.2. Procedure and Instruments

In the qualitative phase, data were collected in April 2024 using the freely available ChatGPT -3.5. Participants provided five English sentences rich in humour, empathy, and cultural references. These sentences were translated into Arabic using ChatGPT. Examples included: “He really jabbed Al-Eid there!” (implying he made a silly mistake) and “I’m so sorry for your loss.” The translations produced by ChatGPT were evaluated independently by three professional English-Arabic

translators using a rubric assessing (a) linguistic accuracy, (b) preservation of humorous intent, (c) conveyance of empathetic tone, and (d) cultural appropriateness (1 = Poor, 5 = Excellent). The interrater reliability was high (Krippendorff's $\alpha = 0.87$). Qualitative translation data were analysed thematically to identify common error patterns.

A Questionnaire was administered to collect the perceptual data. Participants filled out a 25-item online questionnaire adapted from El Shazly's study on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree)^[22]. This questionnaire measured four dimensions: (a) general per-

ception, (b) perception of handling paralinguistic elements, (c) cultural appropriateness, and (d) usefulness. This quantitative data was analysed using SPSS (v.28) for descriptive statistics (mean, standard deviation), ANOVA, and *t*-tests.

The reliability and the validity of the questionnaire were measured. For content validity, the questionnaire was sent to four professors to check the items' language and relevance to the dimensions. For the reliability, the tool was applied on a pilot study containing 23 students, and the tool was found to be valid and reliable for final application, as seen in **Table 2**.

Table 1. Participants' demographic data.

Variables		N	%
Gender	Male	21	45.7%
	Female	25	54.3%
Age	18–20	6	13.0%
	21–22	21	45.7%
	23–25	19	41.3%

Table 2. The reliability of the questionnaire.

Dimensions	Cronbach's Alpha	No. of Items
General perception of ChatGPT as a translation tool	0.658	8
Perception of Paralinguistic Features (Tone, Emotion, Formality)	0.650	8
Cultural and Contextual Appropriateness	0.624	9
Usefulness and Learning Values	0.525	5
Total	0.670	30

4. Results

4.1. General Perceptions and Usage

Students reported a highly positive view of ChatGPT as a translation tool (mean = 3.58, SD = 0.44), a finding consistent with other studies on students' perceptions of the use of AI tools in translation^[23]. They expressed strong reliance on it for translating academic content (mean = 4.00, SD = 0.73) and trusted its accuracy in English-Arabic translation (mean = 3.91, SD = 0.69). Most participants reported a preference for ChatGPT over other similar tools like Google Translate (mean = 3.83, SD = 0.83). However, perceptions of its efficiency in saving time were the lowest (mean = 3.11, SD = 0.92), indicating that post-editing is often perceived as necessary. Data reflecting this factor is summarised in **Table 3**.

According to the analysis of responses in **Table 3** and **Figure 1**, Saudi students generally have a positive opinion of

using ChatGPT for translation. On a five-point Likert scale, the items' mean scores range from 3.11 to 4.00, suggesting that the majority of students agree or strongly agree with the statements. With standard deviation values ranging from 0.62 to 1.10, responses appear to be moderately variable, with some items showing greater agreement amongst the participants than others.

4.2. Perceptions of Paralinguistic Feature Handling

ChatGPT demonstrated competence in handling structural and formal elements but struggled with nuanced paralinguistic features, corroborating the findings of Abu-Rayyash on humour translation^[5]. The responses are summarised in **Table 4** and depicted in **Figure 2**.

The analysis of ChatGPT's performance in paralinguistic feature management in English-to-Arabic translation

gives a comprehensive overview of its strengths and weaknesses. In tone and formality detection, the model shows a comparatively high performance, especially when differentiating between registers of formal and informal language, as indicated by a mean value of 3.50. However, the comparatively high standard deviation of 0.98 indicates a high degree of heterogeneity in user experience, indicating that although

the system performs comparatively well in register detection, its performance can vary depending on the type or complexity of the text. This heterogeneity is also indicated in the case of emotional tone conservation, where the mean value is 3.48, and user feedback shows an even split between agreement and disagreement, pointing towards potential variability in the detection of subtle emotional nuances.

Table 3. General perception of ChatGPT as a translation tool.

No.	Statement	SD	D	N	A	SA	Std.D	Mean
1	ChatGPT encourages me to learn more about translation.	0.00	0.09	0.30	0.37	0.24	0.92	3.76
2	ChatGPT is easy to use for translation purposes.	0.00	0.02	0.41	0.50	0.07	0.62	3.41
3	ChatGPT makes translation more accessible to me as a Saudi student.	0.00	0.02	0.59	0.35	0.04	1.10	3.33
4	ChatGPT saves me time when translating texts.	0.11	0.02	0.43	0.30	0.13	0.92	3.11
5	I believe ChatGPT is suitable for beginners in translation.	0.11	0.02	0.54	0.30	0.02	0.81	3.48
6	I prefer using ChatGPT over other translation tools like Google Translate or DeepL.	0.022	0.04	0.46	0.39	0.09	0.83	3.83
7	I rely on ChatGPT when translating academic content.	0.00	0.00	0.43	0.30	0.26	0.73	4.00
8	I trust ChatGPT to provide accurate translations between English and Arabic.	0.02	0.00	0.13	0.65	0.20	0.69	3.91

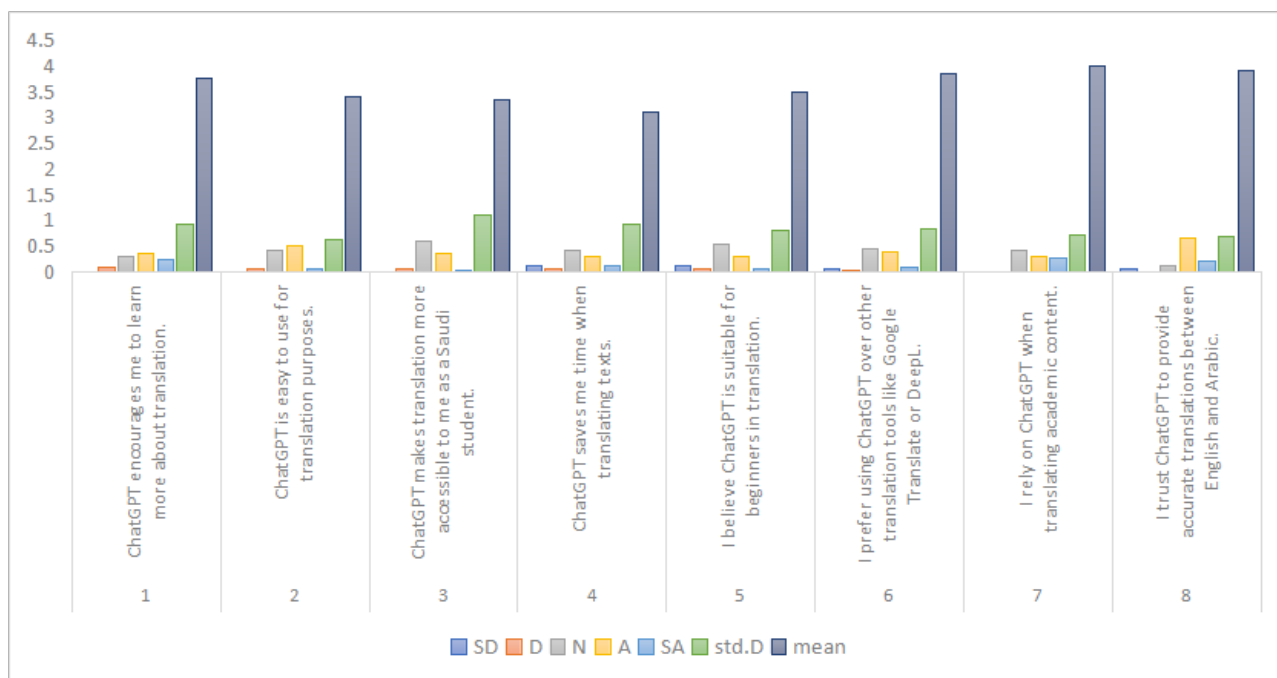


Figure 1. General perception of ChatGPT as a translation tool.

Table 4. Perception of Saudi translation learners towards ChatGPT in translating humour and empathy.

No.	Statement	D	N	A	SA	Std.D	Mean
1	ChatGPT can recognise whether the original English text is humorous or not.	0.00	0.28	0.52	0.20	0.98	3.50
2	ChatGPT captures the humorous tone of the original English texts.	0.17	0.33	0.33	0.17	0.66	3.48
3	ChatGPT maintains the humour level required in formal Arabic writing.	0.04	0.48	0.43	0.04	0.82	3.33
4	ChatGPT struggles to express sarcasm between English and Arabic.	0.11	0.57	0.22	0.11	0.89	3.57
5	I feel that ChatGPT sometimes sounds too neutral in empathy.	0.07	0.50	0.24	0.20	0.77	3.37
6	The emphasis on certain words is well maintained in ChatGPT's translations.	0.09	0.54	0.28	0.09	0.78	3.46
7	The rhythm and tone of spoken English are preserved well in Arabic by ChatGPT.	0.09	0.46	0.37	0.09	0.86	3.46
8	The translated Arabic text reflects the same level of formality as the original English.	0.09	0.52	0.24	0.15	0.92	3.89

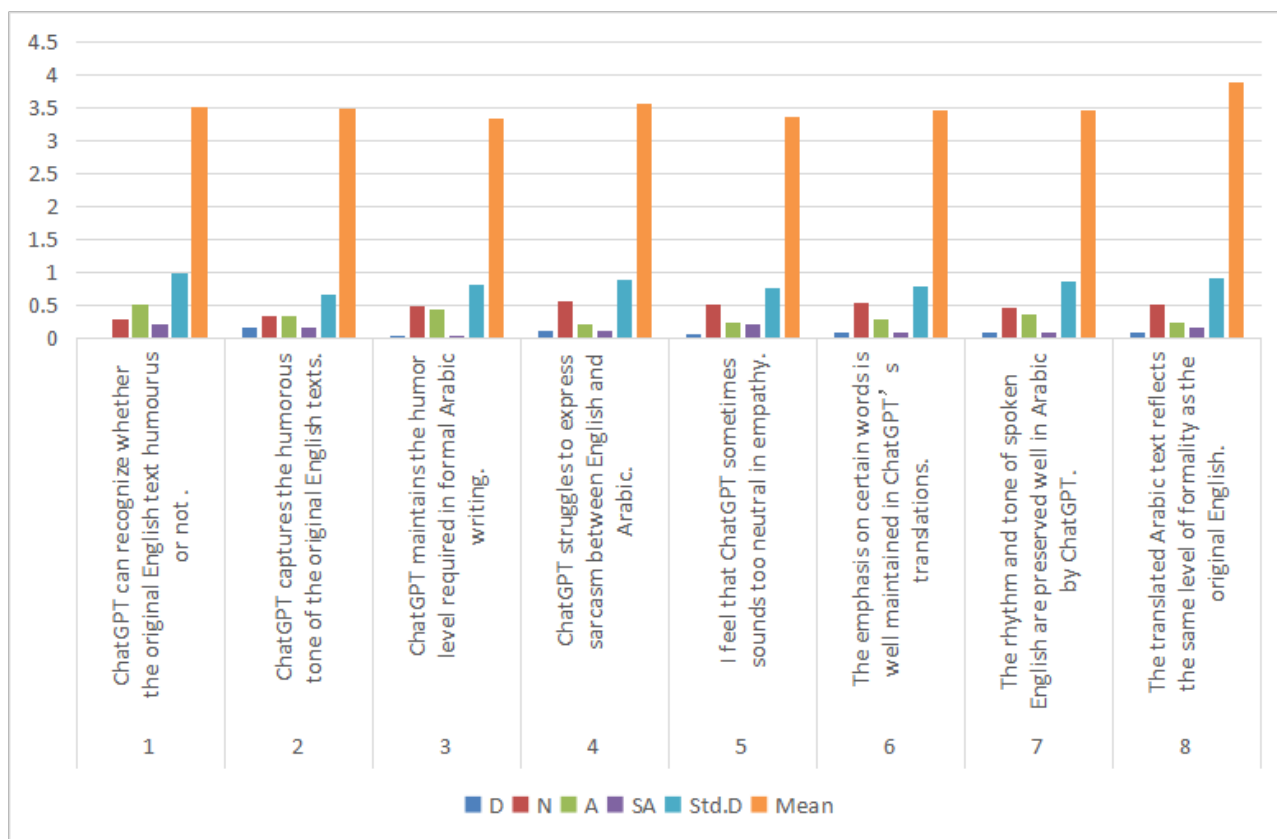


Figure 2. Perception of Saudi translation learners towards ChatGPT in rendering humour and empathy.

A particularly notable challenge emerged in the model's handling of culture-specific linguistic features. As per the data, the translation of sarcasm presents a significant difficulty with a mean of 3.57, reflecting the complex interplay between linguistic form and cultural context that current AI systems struggle to navigate in languages that have limited datasets in LLMs. Similarly, the tendency toward tonal neutralisation, with a mean of 3.37 (50% of the participants' responses were neutral), likely indicates that the model may prioritise semantic accuracy over expressive richness, potentially flattening rhetorical effects in the process. These limitations are especially relevant for literary texts where paralinguistic features carry substantial communicative value.

4.3. Cultural and Contextual Appropriateness

The most significant discrepancy was noted in its treatment of humour, where cultural incongruity frequently resulted in total communicative breakdown, and empathy, where translations, although occasionally neutral or generic, more often preserved the fundamental intent of the source material. ChatGPT showed reasonable yet inconsistent capa-

bility in managing cultural nuances, reflecting the broader challenges LLMs face with low-resource cultural content^[11]. It adapted translations to Saudi norms relatively well (mean = 3.78, SD = 0.99) and produced natural-sounding Arabic (mean = 3.83, SD = 0.80). However, its ability to avoid offensive language was weaker (mean = 3.22, SD = 0.99), and participants often detected cultural mismatches (mean = 3.43, SD = 0.86). The data pertaining to these elements is summarised in **Table 5** and **Figure 3**.

The assessment of ChatGPT's ability to adapt English-to-Arabic translation function culturally and contextually highlights some critical issues regarding its sociolinguistic and cultural corpus. The findings indicate that ChatGPT does relatively well in accommodating content to Saudi cultural conventions, scoring a mean of 3.78, the highest mean score amongst all items. Moreover, ChatGPT showed reasonable local cultural sensitivity because the model's training data included information specific to the region. Still, a standard deviation of 0.99 suggests notable inconsistency in the participants' responses. This means different types or categories of content may yield different comparisons of the context consistency.

Table 5. Cultural and contextual appropriateness.

No.	Statement	D	N	A	SA	Std.D	Mean
1	ChatGPT adapts its translations to suit Saudi cultural norms.	0.11	0.28	0.33	0.28	0.99	3.78
2	ChatGPT avoids offensive language in translation.	0.24	0.46	0.15	0.15	0.99	3.22
3	ChatGPT recognises gender-specific language correctly in Arabic.	0.04	0.50	0.39	0.07	0.69	3.48
4	ChatGPT respects the linguistic traditions of Arabic while translating from English.	0.07	0.54	0.30	0.09	0.75	3.41
5	ChatGPT understands when literal translation is not appropriate.	0.09	0.41	0.30	0.20	0.91	3.61
6	I can detect cultural mismatch in some ChatGPT translations.	0.11	0.48	0.28	0.13	0.86	3.43
7	I would use ChatGPT to translate content for a Saudi audience.	0.11	0.39	0.33	0.17	0.91	3.57
8	The Arabic translations produced by ChatGPT match the intended meaning of English proverbs.	0.07	0.52	0.26	0.15	0.84	3.50
9	The translated text by ChatGPT feels natural to a native Arabic speaker.	0.04	0.28	0.48	0.20	0.80	3.83

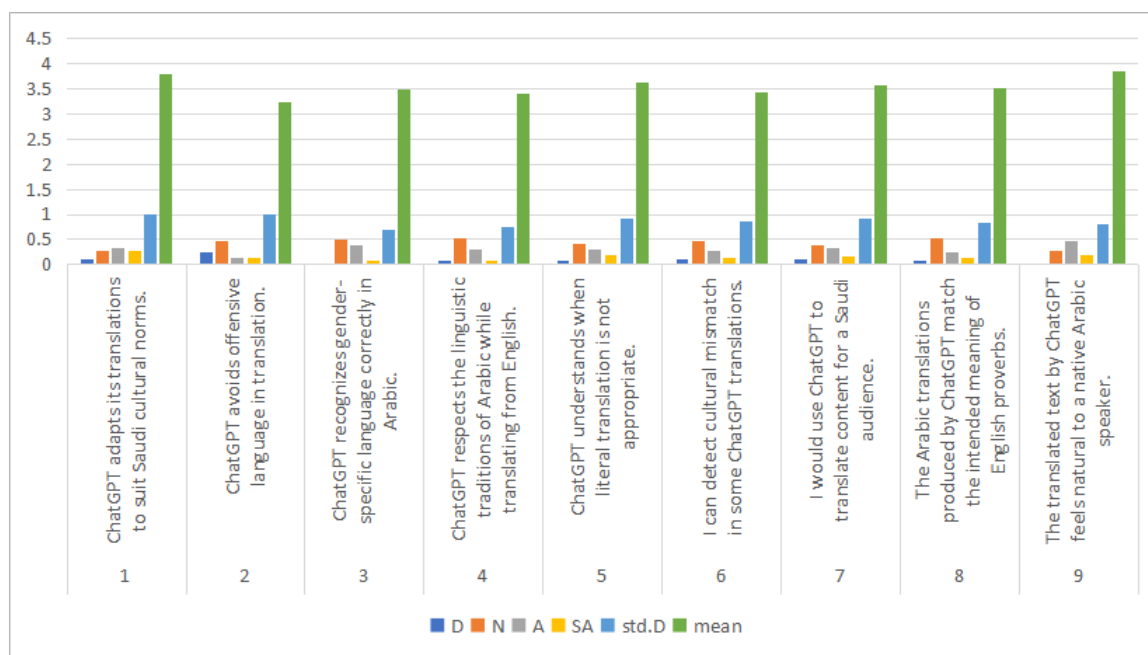


Figure 3. Cultural and Contextual Appropriateness.

For native Arabic speakers, authentic and fluent output is consistently provided by ChatGPT (mean = 3.83). This goes hand in hand with ChatGPT's ability to identify instances when literal translation is unsuitable, which some scholars may consider a form of cultural and linguistic judgment. However, specific aspects of cultural sensitivity yield more mixed results for ChatGPT. It does tend to avoid using offensive expressions on average, achieving 3.22, but this relatively lower score, especially in combination with a high standard deviation of 0.99, suggests that some translations known to be culturally inappropriate may still be produced.

ChatGPT's treatment of culturally loaded content reveals interesting tensions. On the one hand, it shows reasonable ability to render English proverbs in Arabic translation with a mean value of 3.50; on the other hand, users reported being able to detect cultural mismatches in some translations,

which indicates that ChatGPT doesn't always successfully navigate cultural subtleties. This dichotomy may reflect the challenge of balancing linguistic accuracy with cultural appropriateness, particularly for content that relies on cultural context. Despite these limitations, a majority of respondents indicated they would still use ChatGPT for translating content for Saudi audiences, with a mean value of 3.57, which reflects users' trust in its performance as being acceptable for practical purposes.

These findings showed that ChatGPT can produce Arabic translations that sound natural and, at the same time, ensure the exclusion of major cultural mistranslations. The generally positive but not outstanding scores across most dimensions indicate that while the technology has reached a level of practical usability, there is still room for improvement in achieving culturally-adept machine translation in the

English-Arabic language pair. **Table 6** below summarises the perceptual data collected, which highlights that age

can be an influential factor in user perception of ChatGPT in translation—one-way ANOVA test.

Table 6. One-way ANOVA of the perceptual factors and the age of the participants.

Dimensions	Age	Mean	Std. Deviation	Sig.
General perception of ChatGPT as a translation tool	18–20	3.5417	0.18819	0.026
	21–22	3.4167	0.38188	
	23–25	3.7829	0.48204	
	Total	3.5842	0.43784	
Perception of Paralinguistic Features (Tone, Emotion, Formality)	18–20	3.8542	0.57237	0.034
	21–22	3.5595	0.45003	
	23–25	3.3355	0.34874	
	Total	3.5054	0.45258	
Cultural and Contextual Appropriateness	18–20	3.5741	0.49400	0.273
	21–22	3.6349	0.42621	
	23–25	3.4152	0.41050	
	Total	3.5362	0.43148	
Usefulness and Learning Values	18–20	3.6667	0.24221	0.509
	21–22	3.4095	0.50389	
	23–25	3.5263	0.55462	
	Total	3.4913	0.50014	

4.4. Impact of Demographics

A one-way ANOVA showed a significant effect of age on the perception of paralinguistic features, with younger participants (18–20 years old) rating it higher than older peers (23–25 years old). No significant differences were found based on gender across any dimensions, as summarised in **Table 7** below.

These results showed no significant differences in perceptions across the four dimensions as far as the factor of gender is concerned. Both genders had similar overall perceptions of ChatGPT, with men scoring a mean of 3.61 and women scoring a mean of 3.57. This pattern was consistent across paralinguistic features, with females rating ChatGPT's emotional and tonal abilities slightly higher than males. In terms of cultural and contextual appropriateness, both genders again gave similar ratings. However, there was a slight difference in usefulness and learning value perceptions, with female participants rating it slightly higher than males. This could indicate that men and women value ChatGPT's educational usefulness in different ways. The study's findings have implications for AI developers and teachers who use AI translation tools. While there aren't many significant differences between genders, the usefulness dimension could suggest that men and women use and value AI translation

tools for learning in different ways.

4.5. The Size Effect of ChatGPT on English-Arabic Translation of Humour and Empathy

Cohen's *d* is an effect size statistic that measures the standardised difference between two means and eta squared for multigroups, see **Table 8**. They quantify the size of a difference between groups by expressing it in terms of standard deviation units, allowing for comparisons across different studies. The scale of Cohen's *d* is that for small: 0.20–0.49, medium: 0.50–0.79, large: 0.80+, and for eta squared, a value near 0 means an insignificant effect, around 0.01 is small, 0.06 is medium, and 0.14 and above is considered a large effect.

Analysis of independent samples' effect sizes for gender differences across four variables indicates no statistically significant distinctions. All Cohen's *d* values revealed that the observed effects on all dimensions are not sufficiently strong. The wide confidence intervals imply low statistical power, leading to the conclusion that gender is not a reliable or significant predictor of performance on these metrics.

Table 9 presents ANOVA effect sizes (eta-squared) for age across four dependent variables. Age contributes non-trivial variance to some measures, with large effect sizes ($\eta^2 = 0.157$ for General perception of ChatGPT as a translation

tool and $\eta^2 = 0.145$ for Perception of Paralinguistic Features (Tone, Emotion, Formality).

Effect sizes for the other two variables are smaller: 5.9% variance explained for Cultural and Contextual Appropriateness ($\eta^2 = 0.059$, medium effect) and 3.1% for Usefulness and Learning Values ($\eta^2 = 0.031$, small effect). Despite

the apparent large effects for the first two variables, none reach statistical significance. Age may be a factor for the first and second dimensions, but findings require cautious interpretation and further research with larger samples for validation. This can be ascribed to the more young people, the more use of new technologies.

Table 7. Independent sample *t*-test for mean comparison based on the gender variable.

Perception: Gender		Mean	Std. Deviation	Std. Error Mean	Sig.
General perception of ChatGPT as a translation tool	male	3.6071	0.41886	0.09140	0.749042
	female	3.5650	0.46086	0.09217	
Perception of Paralinguistic Features (Tone, Emotion, Formality)	male	3.4405	0.50097	0.10932	0.378238
	female	3.5600	0.41003	0.08201	
Cultural and Contextual Appropriateness	male	3.5661	0.47389	0.10341	0.671532
	female	3.5111	0.40062	0.08012	
Usefulness and Learning Values	male	3.3714	0.50313	0.10979	0.137943
	female	3.5920	0.48470	0.09694	

Table 8. Independent samples effect sizes for gender using Cohen's *d*.

Dimensions	Standardizera	Point Estimate
General perception of ChatGPT as a translation tool	0.44227	0.095
Perception of Paralinguistic Features (Tone, Emotion, Formality)	0.45363	-0.263
Cultural and Contextual Appropriateness	0.43546	0.126
Usefulness and Learning Values	0.49316	-0.447

Table 9. ANOVA effect sizes for age.

Dimensions	Point Estimate
General perception of ChatGPT as a translation tool	0.157
Perception of Paralinguistic Features (Tone, Emotion, Formality)	0.145
Cultural and Contextual Appropriateness	0.059
Usefulness and Learning Values	0.031

4.6. Qualitative Translation Analysis

The evaluation of specific translations identified key limitations. Here are some examples:

- Example 1 (Humour): The English sentence, "He really jabbed Al-Eid there!" was literally translated as, "لقد طعن العيد حقاً هناك." This misses the cultural idiom, humorously meaning "he made a silly mistake." A better translation would be, "لقد ارتكب خطأ سخيفاً," using a culturally recognised phrase for foolishness.
- Example 2 (Empathy): The phrase, "I'm so sorry for your loss", was translated as, "أنا أسف جداً لـ خسارتك," which is grammatically correct but emotionally cold

compared to the more resonant Islamic expression, "إن الله وإن إلهي راح عون."

These patterns illustrate how the divergence in ChatGPT's ability to handle formal versus pragmatic and emotional content—such as humour and empathy—distinctly shaped learner perceptions and critical engagement with machine translation output, as discussed in the following section.

5. Discussion

5.1. Key Findings and Results

This study provides a nuanced evaluation of ChatGPT's performance in English-Arabic translation, particularly in

terms of humour and empathy. The findings confirm that while ChatGPT is a powerful tool for formal and structural translation, it struggles notably with pragmatic and cultural aspects^[1,2].

Translation is seldom a simple act; more often, it is a sequence of processes that deal with complex and singularly unique socio-cultural and linguistic entities. Historically, translation studies have focused on verbal texts—written or oral—that are construed in linguistic terms. But when considering subjects like humour and empathy, translation cannot be limited to a purely linguistic approach. It is here that intersemiotic translation comes into play, as it goes beyond words and includes non-linguistic media and expressive forms. Here, translation can be found in the form of converting verbal signs to non-verbal or vice versa, and needs interpretive schemes that span language systems.

Humour and empathy are especially difficult to translate due to the intersemiotic aspect of translation output. Unlike technical or informative texts, humour and empathy are built upon cultural codes, symbolic meanings, and affective connections that run very deep in certain contexts. Translating them is not a matter of meaning transfer but a creative process of reinterpretation. One expression can have several levels of meaning, and when transferred between languages, particularly between English and Arabic, which are dissimilar in structure, semantics, and culture, the possibility of several interpretations grows. Therefore, the human or machine translator is not only responsible for communicating the literal meaning but also for reconstructing the emotional and cultural effect of the original text.

From this point of view, the translation of humour and empathy demands a high level of sensitivity, creativity, and cultural knowledge. The translator (including machines) needs to understand not only the surface of the language but also the depth of the semiotics in the message. Humour tends to depend upon wordplay, irony, or incongruity, while empathy relies upon tacit hints of tone, vantage point, and cultural connotation. For a machine translation engine like ChatGPT, this is particularly challenging. While the model can mirror linguistic equivalents at decent fluency, the transference of affect, irony, and cultural subtlety is more uncertain. ChatGPT might get away with maintaining surface forms but not with mimicking the layered sense or the desired emotional impact, especially when English humour finds its base in id-

iomatic wordplay or when Arabic cultural backdrops require a different kind of expression.

Such translation difficulties cannot be well accounted for by communication theories that view language as mechanical information transmission. Shannon's theory, for example, envisions communication as a process in which a message is completely encoded, transmitted, and decoded in determinate ways^[24]. Such a model has been used for effect in the research on information systems, but is incomplete in describing the translation of humour and empathy, where meaning is co-constructed rather than pre-construed in the interface of text, translator, and reader. Human beings communicate not only to exchange something practically, but also for aesthetic, cultural, and emotional experience, which defies formulaic transmission.

5.2. Comparison with Prior Studies

The high level of student trust and reliance on ChatGPT, especially for academic texts, aligns with findings from previous studies^[10,23]. This suggests it is inevitable to integrate it into translation education. However, the moderate scores in translating humour and empathy highlight a significant limitation, a recurring issue in AI translation studies^[3,5]. The observation that ChatGPT often produces "neutral" empathetic translations emphasises that AI lacks genuine emotional understanding^[6]. Additionally, its difficulty with sarcasm and cultural humour supports the argument that translating humour needs a deep grasp of cultural nuances that ChatGPT currently lacks^[2].

The stronger performance in matching formality is due to more predictable patterns in formal language. Conversely, the variability in cultural adaptation suggests that, while vast, ChatGPT's training data lacks enough nuance in its corpus for the complexities of Arabic social and linguistic norms^[7,11]. The difference in perception based on age indicates that younger users may be more tuned to or forgiving of AI's limitations^[25,26].

6. Recommendations

For developers, this study highlights the urgent need to create training datasets that are richer in dialectical Arabic, cultural idioms, and pragmatically annotated examples^[11]. Enhancing emotion recognition algorithms should be a pri-

ority for generative model developers^[6].

For educators, rather than banning ChatGPT, harnessing it as a teaching tool, as suggested by Lee, can be effective for comparative exercises—contrasting AI and human translation outputs—and for teaching essential post-editing skills, thus improving students' language and cultural awareness^[10].

For learners, the message is to engage critically. While ChatGPT can be a helpful assistant for creating and checking formal translations, it is not a reliable partner for translating cultural nuances and emotions^[4].

7. Limitations of the Study

This study is limited by its sample size, its focus on a single AI tool (ChatGPT-3.5), and its focus on text translation only. Future research should include comparisons with other NMT systems, involve larger and more diverse samples, and use longitudinal designs to track learning outcomes. Investigating the translation of multimodal paralinguistic cues, such as tone of voice, also represents an important future direction.

8. Conclusions

The achievement of ChatGPT in translation poses important questions: how far can a machine, pre-trained with copious linguistic data, simulate the cultural imagination and interpretive range necessary for translating humour and empathy from English to Arabic? How does it deal with intrasemiotic elements like irony, hyperbole, or metaphor that are fundamental to humour? Does it convey the relational and affective aspects of empathy, or does it fall back to literal translations that flatten emotional complexity? In answering these questions, the current research places ChatGPT's translation within the wider context of translation studies and semiotics theory and emphasises the importance of assessing not just accuracy but also emotional equivalence and cultural appeal.

This study shows that ChatGPT has become a go-to translation tool for students, offering significant advantages in accessibility and formal discourse^[23]. However, its ability to translate the essence of human communication—humour and empathy—remains lacking^[3,5]. These elements are not just linguistic challenges but are deeply woven into our cul-

tural fabric and human experience and hence, command a central position in any communicative event^[4]. This study examined ChatGPT's ability to convey empathy and humour in English-to-Arabic translation.

The results show a complex AI-assisted translation scenario with linguistic and cultural constraints and technological potential. ChatGPT excels in formal and academic contexts but struggles with nuanced paralinguistic features. These findings suggest that ChatGPT can improve vocabulary comprehension and grammatical precision in language acquisition and translation education^[10]. Its inability to handle emotional and humorous content stresses the importance of culturally sensitive human translation expertise^[2].

The technological implications of these findings highlight several AI language model possibilities. Current systems are still in dire need of improving their cultural and emotional intelligence to reflect the complexity of human communication^[1,13].

Future research should focus on machine translation's emotional resonance, cultural sensitivity, and contextual awareness. These improvements may demand improved training datasets, pragmatic information integration, and social cue and rhetorical strategy recognition and emulation systems^[11]. Despite its ability to help with translation tasks quickly, ChatGPT's limitations apply to emotionally charged or culturally nuanced content.

The best method combines human intervention with AI assistance, fostering a symbiotic relationship between human insight and technical efficacy. Therefore, the future of AI-assisted translation should be about creating a partnership between humans and AI. Developers need to build more culturally aware systems, and educators need to guide learners on how to use these powerful tools alongside critical thinking, ensuring that the art of translation retains its human touch. Furthermore, the scope of the study should be widened to cover as many institutions as possible in Saudi Arabia.

Author Contributions

Conceptualization, M.H.A.; methodology, M.H.A.; software, A.A.M.H.A.-A.; validation, A.A.M.H.A.-A. formal analysis, M.H.A.; investigation, A.A.M.H.A.-A.; resources, A.A.M.H.A.-A.; data curation, A.A.M.H.A.-A.; writing—original draft preparation, A.A.M.H.A.-A.;

writing—review and editing, A.A.M.H.A.-A.; visualization, M.H.A.; supervision, M.H.A.; project administration, M.H.A.; funding acquisition, M.H.A. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. KFU252599].

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

The researchers have strictly adhered to ethical guidelines by ensuring that the participants gave full consent for sharing data. They were duly assured before commencement of the study that any information they shared would be fully confidential and used only for the purpose of the study. They were also informed that they could choose to exit the study at any point of time they wished.

Data Availability Statement

Information about data and materials used in the study is available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Hadi, M.U., Al-Tashi, Q., Qureshi, R., et al., 2023. Large Language Models: A Comprehensive Survey of Its Applications, Challenges, Limitations, and Future Prospects. *TechRxiv*. 4, 1–43. DOI: <https://doi.org/10.36227/techrxiv.23589741.v4>
- [2] Mohebbi, A., 2023. The Use of Cultural Conceptualisations as a Translation Strategy for Culture-Specific Jokes and Humorous Discourse: A Remedy for a Malady? *Ampersand*. 11, 100150. DOI: <https://doi.org/10.1016/j.amper.2023.100150>
- [3] Vandaele, J., 2016. *Translating Humour*. Routledge: London, UK.
- [4] Bekouche, M., Elsaid, K., 2024. Exploring Emotional and Cultural Dimensions in Translation: A Sociolinguistic Approach to Arabic Literary Forms. *Revue Algérienne Des Sciences Du Langage*. 9(1), 111–134.
- [5] Abu-Rayyash, H., 2024. AI Meets Comedy: Viewers' Reactions to GPT-4 Generated Humour Translation. *Ampersand*. 12, 100162. DOI: <https://doi.org/10.1016/j.amper.2023.100162>
- [6] Parish, L.A., 2022. *Empathy in Conversational AI: Rhetorical Analyses and Recommendations* [Master's thesis]. Texas State University: San Marcos, TX, USA.
- [7] Almadani, W., 2021. *How Hijazi Men and Women Say "NO": A Pragmatic and Discourse Analysis Study of the Speech Act of Refusal, Gender and Culture in Saudi Arabia* [PhD thesis]. University of Sunderland: Sunderland, UK.
- [8] Manasiq, G.Z.A.S., 2024. Interdisciplinary Approaches in Arabic Language Teaching: Integration with Islamic Studies to Deepen Cultural Understanding. *Halimi: Journal of Education*. 5(2), 36–55.
- [9] Guzmán, F., Joty, S., Márquez, L., et al., 2017. Machine Translation Evaluation with Neural Networks. *Computer Speech & Language*. 45, 180–200.
- [10] Lee, S.M., 2023. The Effectiveness of Machine Translation in Foreign Language Education: A Systematic Review and Meta-Analysis. *Computer Assisted Language Learning*. 36(1–2), 103–125. DOI: <https://doi.org/10.1080/09588221.2021.1901745>
- [11] Zhong, T., Yang, Z., Liu, Z., et al., 2024. Opportunities and Challenges of Large Language Models for Low-Resource Languages in Humanities Research. *arXiv preprint. arXiv:2412.04497*. DOI: <https://doi.org/10.48550/arXiv.2412.04497>
- [12] Al-Wasy, B.Q., Mohammed, O.S.M., 2024. Strategies of Translating Euphemistic Expressions from Arabic into English: A Comparative Study of Artificial Intelligence Models with Human Translation. *Humanities and Educational Sciences Journal*. 40, 826–855. DOI: <https://doi.org/10.55074/hesj.vi40.1121>
- [13] Sterken, R.K., Kirkpatrick, J.R., 2025. Conversational Alignment with Artificial Intelligence in Context. *Philosophical Perspectives*. 38(1), 89–102. Available from: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/phpe.12205>
- [14] Alghamdi, N., 2025. Utilising Artificial Intelligence (AI) for Vocabulary Learning by Saudi EFL Students: Perspectives, Practices, and Challenges. *Forum for Linguistic Studies*. 7(9), 15–28. DOI: <https://doi.org/10.30564/fls.v7i9.10870>
- [15] Alowed, N.A., Al Ahdal, A.A.M.H., 2023. Artificial Intelligence Based Arabic-to-English Machine versus Human Translation of Poetry: An Analytical Study of

- Outcomes. *Journal of Namibian Studies*. 33.
- [16] Alqarni, M.D., 2025. Investigating the Translation Dynamics of Arabic-English Code-Switching by AI and EFL University Students in Saudi Arabia. *International Journal of Linguistics, Literature and Translation*. 8(6), 72–84. DOI: <https://doi.org/10.32996/ijllt.2025.8.6.7>
- [17] Alsaied, M.A., 2024. Teaching the Arabic Language Between the Challenges of Globalization and the Enhancement of National Identity. *Journal of Ecohumanism*. 3(8), 3968–3977.
- [18] Alzabidi, A.S., Al-Ahdal, A.A.M.H., 2022. Translanguaging in Saudi Classrooms: A Study of Upper Secondary Learner Perceptions. *Journal of Language and Linguistic Studies*. 17(3).
- [19] Qian, S., 2025. Evaluating Machine Translation of Emotion-Loaded Chinese User-Generated Content [PhD thesis]. University of Surrey: Surrey, UK.
- [20] Orazova, B., Zharylgapov, Z., Kelmaganbetova, A., et al., 2025. The Art of Adaptation: Translation in the Context of Intercultural Communication. *Forum for Linguistic Studies*. 7(5), 561–572. DOI: <https://doi.org/10.30564/fls.v7i5.8980>
- [21] Naeem, A., ur Rehman, A.S., Rasheed, A., 2025. Evaluating Cultural Adaptation in AI Translations: A Framework and Implications for Literary Works. In: Ali, M.M., Anwar, M.N., Fazal, S., et al. (Eds.). *AI Applications for English Language Learning*. IGI Global Scientific Publishing: Hershey, PA, USA. pp. 223–252.
- [22] El Shazly, R., 2021. Effects of Artificial Intelligence on English Speaking Anxiety and Speaking Performance: A Case Study. *Expert Systems*. 38(3), e12667. DOI: <https://doi.org/10.1111/exsy.12667>
- [23] Almhasees, Z., Abu-Rayyash, H.K., Olimat, S.N., et al., 2024. An Analytical Study on Student Perceptions of Using ChatGPT in Language Translation. *Language Value*. 17(2), 1–23. Available from: <https://dialnet.unirioja.es/servlet/articulo?codigo=9937277>
- [24] Shannon, C.E., 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal*. 27(3), 379–423. DOI: <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- [25] Abdelgawad, N.A., 2022. The Netflix Translation of Humour from the Egyptian Vernacular Dialect into MSA: The Case of “el-Limby” and “Illy Baly Balak” Films. *Journal of Linguistics and Education Research*. 5(1), 1–13. DOI: <https://doi.org/10.30564/jler.v5i1.4814>
- [26] Thompson, P., 2013. The Digital Natives as Learners: Technology Use Patterns and Approaches to Learning. *Computers & Education*. 65, 12–33.