ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

Cultural Translation in the Age of Artificial Intelligence: Limitations and Biases of Large Language Models

Chenxi Ban, Shu Wu*

University of Shanghai for Science and Technology, Shanghai, China
**Corresponding author*

Abstract: *Large language models (LLMs) have dramatically reshaped the landscape of translation practice, offering unprecedented speed and scalability across hundreds of language pairs. Yet the enthusiasm surrounding these systems has frequently outpaced a sober accounting of what they cannot do. This paper examines the structural limitations and embedded biases of LLMs when they are applied to cultural translation—understood here not merely as the conversion of words between languages but as the negotiation of meaning across culturally situated contexts. Drawing on recent empirical and theoretical scholarship, the paper argues that LLMs reproduce and, in certain conditions, amplify the cultural asymmetries already present in their training data; that they systematically underperform on figurative language, politeness conventions, and minority or low-resource languages; and that the growing integration of AI into professional translation workflows raises questions about epistemic authority that the field has not yet adequately addressed. The paper concludes by suggesting that the most productive path forward is not the wholesale rejection of AI-assisted translation but a theoretically informed recalibration of its role—one that keeps human cultural judgment at the center of the process.*

Keywords: *large language models, cultural translation, bias, figurative language, low-resource languages, translation education*

1. Introduction

Translation has never been a neutral act. Every choice a translator makes—which word to reach for, which connotation to preserve, which cultural reference to gloss or to leave intact—is saturated with judgments that are simultaneously linguistic, cultural, and ideological. For most of the twentieth century, debates about those judgments were conducted among human practitioners, and the field of translation studies developed a rich theoretical vocabulary for understanding them: equivalence, domestication, foreignization, skopos, thick translation. The arrival of neural machine translation, and more recently of large language models capable of generating fluent prose across dozens of languages, has not dissolved those debates. It has, if anything, made them more urgent.

The commercial success of systems such as Google Translate, DeepL, and the translation capabilities embedded in generative models like GPT-4 and Gemini has created a widespread impression that the translation problem is, for most practical purposes, solved. Fluency is often mistaken for accuracy, and accuracy for cultural adequacy. These are not the same things. A sentence can be grammatically impeccable and semantically faithful to its source while still failing to carry the social weight, the emotional register, or the cultural resonance that the original text possessed for its intended audience. This gap—between linguistic correctness and cultural meaning—is precisely where LLMs are most vulnerable, and it is the gap that this paper sets out to examine.

2. Training Data, Cultural Representation, and the Illusion of Neutrality

To understand why LLMs struggle with cultural translation, it is necessary to understand something about how they are built. These models are trained on vast corpora of text drawn predominantly from the internet, from digitized books, and from existing parallel translation databases. The scale is genuinely staggering—hundreds of billions of tokens, spanning thousands of domains and dozens of languages. But scale is not the same as representativeness, and the cultural distribution of that data is deeply uneven.

So do the other major European languages, along with Mandarin, Japanese, and a handful of other

ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

high-resource languages. The texts that constitute the training corpus are not a random sample of human linguistic production; they are a sample of what has been digitized, published, and made accessible online, which means they skew heavily toward educated, urban, Western, and economically privileged voices. The cultural assumptions embedded in those texts—about gender roles, social hierarchies, the relationship between the individual and the community, the appropriate register for discussing illness or death or sexuality—become, through the training process, the default assumptions of the model. The model does not know it has assumptions. It has no metacognitive access to its own biases. It simply generates text that is statistically consistent with the patterns it has absorbed.

This matters enormously for translation. When an LLM is asked to translate a text from one language into another, it is not consulting a neutral dictionary of equivalences. It is drawing on a probabilistic model of how words and phrases co-occur in its training data, and that model is culturally loaded in ways that are often invisible to the user. The appearance of fluency can mask the substitution of one cultural framework for another. A translation that reads smoothly in the target language may have quietly domesticated the source text—flattening its cultural specificity, normalizing its foreignness, and presenting the reader with something that feels familiar precisely because it has been stripped of what made it strange [1].

Zheng's (2025) analysis of AI-assisted translation of Chinese emergency management texts into English offers a useful illustration of this dynamic [2]. Emergency management discourse in China is embedded in a specific institutional and political culture, with characteristic rhetorical patterns, modes of authority, and assumptions about the relationship between the state and the individual. When LLMs translate these texts, they tend to produce English prose that conforms to the conventions of Western administrative writing—which is to say, they perform a kind of invisible cultural substitution. The resulting translations are readable, but they have lost something that the original possessed: a particular way of organizing authority and responsibility that is not merely a stylistic preference but a substantive feature of the text's meaning.

This is not a marginal problem. It is structural. The cultural biases of LLMs are not bugs that can be patched in the next software update; they are consequences of the way these systems are trained, and they will persist as long as training data remains culturally skewed and as long as the evaluation metrics used to assess translation quality prioritize fluency over cultural fidelity.

3. Figurative Language: Where Statistical Patterns Break Down

Figurative language is perhaps the most demanding test of any translation system, human or machine. Idioms, proverbs, and metaphors are not merely decorative; they are the sites where a language's cultural logic is most densely concentrated. An idiom is not a phrase that happens to have a non-literal meaning; it is a crystallized piece of cultural knowledge, a shorthand that only works for someone who already knows what it refers to. Proverbs encode moral frameworks and social norms. Metaphors organize entire conceptual domains—the way a culture thinks about time, about the body, about social relationships—and these conceptual structures are not universal.

LLMs handle figurative language through pattern matching. When a model encounters an idiom it has seen frequently in its training data, it can often produce a reasonable equivalent in the target language, because the training data includes examples of how human translators have handled that idiom before. But this apparent competence is fragile. It depends on the idiom being common enough to appear repeatedly in the training corpus, on the target language having a reasonably close equivalent that also appears in the corpus, and on the model correctly identifying the figurative rather than the literal reading of the phrase. When any of these conditions fails, the results can range from awkward to actively misleading.

Tannous and Haider's (2025) study of Google Translate and Gemini's performance on Arabic idiomatic expressions and proverbs provides detailed empirical evidence of these limitations [3]. Arabic is a language with an exceptionally rich tradition of proverbial expression, and many Arabic proverbs carry cultural and religious resonances that have no direct equivalent in English. The study found that both AI tools performed reasonably well on idioms that have close English parallels but struggled significantly with proverbs whose meaning is deeply embedded in Arab cultural and religious contexts. The models tended either to translate literally—producing nonsense—or to substitute a semantically adjacent English expression that lost the cultural specificity of the original. Neither strategy is adequate for a translator whose goal is to convey not just what was said but what was meant, and what kind of person would say it in what kind of situation.

ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

Alazzam, Alzghoul, and Alzghoul (2025) reach broadly similar conclusions in their examination of AI's capacity to translate English metaphors into Arabic [4]. Metaphor translation is particularly challenging because metaphors are not simply figures of speech; they are, as cognitive linguists have argued since Lakoff and Johnson, the primary means by which abstract concepts are understood. The conceptual metaphors that structure English discourse—time as a resource, argument as combat, the mind as a container—are not universal, and their Arabic counterparts, where they exist, are often structured differently. The study found that LLMs frequently imposed English conceptual structures on Arabic translations, producing texts that were grammatically acceptable but culturally dissonant. The models were, in effect, translating English metaphors into Arabic words while leaving the English conceptual framework intact.

Matan and Velvizhy's (2025) work on a neuro-symbolic AI approach to translating children's stories from English to Tamil adds another dimension to this picture [5]. Tamil is a Dravidian language with a literary tradition stretching back more than two millennia, and it carries cultural and emotional associations that are quite distinct from those of English. The study found that standard neural translation models struggled with the emotional register of children's literature—the specific ways in which warmth, humor, and moral instruction are encoded in Tamil narrative conventions—and proposed a neuro-symbolic approach that attempts to incorporate explicit cultural and emotional knowledge into the translation process. The results were promising but also revealing: the need for such an approach in the first place is a testament to how poorly standard LLMs handle the cultural texture of even relatively simple texts.

4. Politeness, Pragmatics, and the Social Life of Language

Figurative language is not the only domain where cultural meaning resists algorithmic capture. Politeness is another. Every language has systems for encoding social relationships—deference, solidarity, distance, intimacy—and these systems are not simply mappings of social facts onto linguistic forms. They are constitutive of social relationships; they do not merely describe how people relate to each other but actively create and maintain those relationships. A translation that gets the propositional content right but mishandles the politeness system has not merely made a stylistic error; it has produced a text that does something different in the world.

The complexity of politeness translation is well illustrated by Sofian, Nababan, Santosa, and Djatmika's (2025) study of the Indonesian translation of Kazuo Ishiguro's novel Klara and the Sun [6]. The novel features an AI narrator—Klara—whose speech is characterized by a distinctive politeness register that encodes her status as a non-human entity navigating a human social world. The Indonesian translation faces a particular challenge because Indonesian has a complex system of honorifics and politeness levels that does not map neatly onto the English original. The study found that the translation made systematic choices about how to render Klara's politeness that had significant consequences for how her character is understood—choices that a human translator made deliberately but that an LLM, lacking any understanding of the social and narrative stakes involved, would be unlikely to make consistently or appropriately.

This example is instructive because it involves not just the translation of politeness forms but the translation of a character whose identity is partly constituted by her relationship to politeness. Klara's way of speaking is not incidental to who she is; it is part of what makes her both alien and sympathetic. An LLM translating this text would have no access to that interpretive framework. It would process the politeness forms as linguistic data and produce target-language equivalents based on statistical patterns, without any understanding of what those forms are doing in the narrative.

The challenge is not confined to literary translation. Heathco's (2025) work on AI-translated task instructions for non-native English-speaking students demonstrates that even in relatively utilitarian contexts, the pragmatic dimensions of language—the implicit assumptions about the reader's background knowledge, the appropriate level of formality, the cultural expectations embedded in pedagogical discourse—can be seriously distorted by AI translation [7]. Students receiving AI-translated instructions were sometimes confused not because the translation was semantically inaccurate but because the pragmatic register was wrong: the translated text sounded too formal, or too informal, or made assumptions about the student's relationship to authority that did not match the cultural context of the classroom.

Andalib et al.'s (2025) validation study of AI-translated Spanish orthopedic medical texts raises similar concerns in a higher-stakes domain [8]. Medical communication is not just about conveying

ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

information accurately; it is about establishing trust, managing anxiety, and enabling patients to make informed decisions. The pragmatic dimensions of medical language—the way uncertainty is expressed, the way the patient's agency is acknowledged or elided, the way bad news is delivered—are culturally variable and clinically significant. The study found that while AI translation performed adequately on technical terminology, it was less reliable on the pragmatic and relational dimensions of medical communication, which are precisely the dimensions that matter most for patient understanding and compliance.

5. Low-Resource Languages and the Reproduction of Linguistic Inequality

The consequences of this asymmetry are not merely technical. They are political. When an LLM cannot translate adequately into or out of a low-resource language, it is not simply failing to solve a hard computational problem; it is reproducing and reinforcing the marginalization of the communities that speak that language. The implicit message is that some languages—and by extension, some cultures—are more translatable, more legible, more worthy of the resources required to build good translation systems. This is a form of what might be called algorithmic linguistic imperialism, and it deserves to be named as such.

Desrine's (2025) study of Trinidadian Bhojpuri translation in the AI era provides a particularly vivid illustration of this problem [9]. Trinidadian Bhojpuri is a creolized variety of Bhojpuri that developed among the Indo-Caribbean community in Trinidad and Tobago, and it carries a dense cultural history—of indenture, of diaspora, of cultural survival and transformation—that is encoded in its vocabulary, its phonology, and its pragmatic conventions. The study found that AI translation systems are essentially unable to handle this language: they either fail to recognize it as a distinct variety, assimilating it to standard Bhojpuri or to English, or they produce translations that strip away the cultural specificity that makes the language meaningful to its speakers. The loss is not merely linguistic; it is a loss of cultural memory and identity.

Peterson's (2025) reading of Kafka's "Josefine, the Singer or the Mouse People" as a metaphor for AI offers a different but complementary perspective on this problem [10]. Kafka's story is about a singer whose art is indistinguishable from ordinary mouse-squeaking—or perhaps it is not, and the difference is everything. Peterson argues that this ambiguity is a productive metaphor for AI translation: the model produces outputs that are formally indistinguishable from human translation in many cases, but the question of whether something has been lost in the process—some quality of understanding, of cultural embeddedness, of intentionality—remains open and perhaps unanswerable. The metaphor is apt precisely because it resists easy resolution. It does not tell us that AI translation is worthless; it asks us to think carefully about what we mean by value in translation, and whether fluency is a sufficient criterion.

The situation of low-resource languages also raises questions about the feedback loops built into LLM training. When a model is retrained on its own outputs—a practice that has become increasingly common as AI-generated text proliferates online—the cultural biases of the original training data are not corrected; they are amplified. Low-resource languages that were poorly represented in the first place become even more marginal as the training corpus fills up with AI-generated text in high-resource languages. This is a dynamic that the field needs to monitor carefully, and it is one that purely technical solutions are unlikely to address.

6. Translation Education and the Question of Epistemic Authority

The growing integration of AI into translation practice has significant implications for how translation is taught and how the profession understands its own expertise. These implications are not straightforward. AI-assisted translation is not simply a new tool that translators can add to their existing repertoire; it changes the nature of the task in ways that require a rethinking of what translation competence means and what translation education should aim to produce.

Jiménez Crespo's (2025) analysis of human-centered AI and intelligence augmentation in translation education addresses this challenge directly [1]. The paper argues that the dominant model of AI integration in translation pedagogy—in which students learn to use AI tools as productivity aids, with human post-editing as the primary skill—is inadequate because it frames the human translator's role as essentially corrective rather than creative. Students who learn to translate "like a robot," as the paper's title puts it, are not developing the cultural and interpretive competencies that distinguish human translation from machine output; they are learning to smooth over the rough edges of machine output,

ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

which is a very different skill. The paper calls for a reorientation of translation education around the concept of intelligence augmentation—the idea that AI should enhance rather than replace human judgment, and that the goal of education should be to develop the kind of deep cultural knowledge and interpretive sophistication that AI cannot replicate.

This argument has considerable force, but it also raises a difficult question: if the distinctive value of human translation lies in cultural knowledge and interpretive sophistication, how do we assess whether students are developing those qualities? The metrics that have traditionally been used to evaluate translation quality—accuracy, fluency, consistency—are precisely the metrics on which AI performs best. Developing assessment frameworks that can capture the cultural dimensions of translation competence is a significant pedagogical challenge, and one that the field has not yet fully met.

Buckley, O'Sullivan, McKean, and Frizelle's (2026) work on AI literacy for speech and language therapists, while not primarily about translation, raises a related point about the importance of domain-specific knowledge in evaluating AI outputs [11]. The paper argues that practitioners who lack a clear understanding of how AI systems work—what they are optimized for, what their failure modes are, what kinds of errors they are likely to make—are poorly positioned to use them responsibly. This argument applies with equal force to translation practitioners. A translator who uses an LLM without understanding its cultural biases is not simply using a tool; they are delegating cultural judgment to a system that has no cultural judgment to offer, and they may not even notice when the delegation has gone wrong.

Tao's (2025) discussion of Google Translate in speech-language pathology contexts makes a similar point from a clinical perspective [12]. The paper notes that practitioners who rely on AI translation for patient communication may be unaware of the cultural and pragmatic distortions that the translation introduces, and that these distortions can have real consequences for patient care. The problem is not that AI translation is useless in this context; it is that its limitations are not visible to practitioners who lack the linguistic and cultural knowledge to detect them. This is a general problem, not a domain-specific one: the opacity of AI translation errors is one of its most dangerous features.

7. Toward a More Honest Account of AI and Cultural Translation

None of the foregoing is an argument for abandoning AI-assisted translation. The practical benefits of these systems are real: they dramatically reduce the time and cost of translation, they make multilingual communication accessible in contexts where human translation would be prohibitively expensive, and they perform well enough on routine, low-stakes tasks to be genuinely useful. The question is not whether to use them but how to use them with appropriate awareness of their limitations.

What that awareness requires, at minimum, is a clearer account of what LLMs are actually doing when they translate. They are not understanding texts in any meaningful sense; they are generating statistically probable sequences of tokens in the target language, conditioned on the source text and on the patterns in their training data. This is a remarkable technical achievement, but it is not the same as translation in the full sense of the word—the sense that involves understanding what a text means, why it matters, and what would be lost if it were rendered differently. The conflation of these two things—fluent output with genuine understanding—is the source of much of the overconfidence that currently surrounds AI translation.

A more honest account would acknowledge that LLMs are powerful tools for certain kinds of translation tasks and poor tools for others. They are reasonably reliable for translating technical documentation, legal boilerplate, and other genres where cultural specificity is low and the premium is on accuracy and consistency. They are much less reliable for literary translation, for the translation of culturally embedded discourse, for figurative language, for politeness-sensitive communication, and for low-resource languages. The appropriate response to this situation is not to pretend that the limitations do not exist but to develop practices and institutional structures that keep human cultural judgment in the loop where it matters most.

The broader implication is that the AI era does not make translation studies less relevant; it makes it more so. The theoretical frameworks that the field has developed for thinking about equivalence, cultural mediation, power, and identity are precisely the frameworks needed to evaluate what AI translation does and does not do. The challenge for the field is to bring those frameworks to bear on the new technological landscape with the same rigor and critical acuity that it has applied to the work of human translators.

ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

8. Conclusion

The argument of this paper can be stated simply, even if its implications are not. Large language models are genuinely impressive translation tools, and their capabilities will continue to improve. But they are not culturally competent translators, and the gap between fluency and cultural adequacy is not a temporary technical limitation that will be closed by the next generation of models. It is a structural consequence of how these systems are built and what they are optimized for. Recognizing this gap is not a counsel of despair; it is a precondition for using AI translation responsibly.

The evidence reviewed in this paper suggests that LLMs systematically underperform on figurative language, on politeness and pragmatic meaning, and on low-resource and minority languages—and that their failures in these domains are not random but patterned, reflecting the cultural biases embedded in their training data. These failures matter because they are not always visible to users, because they can have real consequences in high-stakes contexts, and because they tend to reproduce and amplify existing cultural and linguistic inequalities.

The path forward requires a combination of technical improvement, institutional reform, and theoretical clarity. Better training data, more culturally diverse evaluation metrics, and more transparent documentation of model limitations would all help. So would educational frameworks that develop genuine cultural competence in translators rather than training them to be efficient post-editors of machine output. And so would a broader cultural conversation about what we want from translation—not just speed and accessibility, but fidelity to the full complexity of human meaning-making across cultural difference. That conversation is one that AI cannot have for us.

Acknowledgements

This research was funded by the China Youth & Children Research Association "Climb Plan", (G2025-0901-005), the Fund of Shanghai Education Science Research Project (C2024196); The Ministry of Education's Industry University Cooperation Collaborative Education Project (231101913174352); The Ministry of Education's Supply-Demand Docking Employment and Education Project (2023123031821); Funds of University of Shanghai for Science and Technology (YLKC202424443, CFTD2025ZD16, SH2025268, SH2025265, XJ2025564, 2025 Special Project for Women's Civilization Construction, 2025 Practical Innovation Project for Academic Atmosphere Construction); 2025 "Social Innovation and Education Integration" Community Education Innovation Project(SCRJ20250501005); the Humanities and Social Sciences Cultivation Fund Project of USST (Grant No. 24SKPY04); The Talent Development Model for High-Skilled Professionals in China Vocational Education Society 2025 Annual Planning Research Project (ZJS2025ZD36).

References

[1] Jiménez Crespo, M. A. (2025). "If students translate like a robot…" or how research on human-centered AI and intelligence augmentation can help realign translation education. The Interpreter and Translator Trainer, 19(3–4), 277–295.
[2] Zheng, Y. (2025). On the underlying logic of translating Chinese texts on emergency management into English in the AI context. Scientific and Social Research, 7(12), 251–260.
[3] Tannous, B., & Haider, A. S. (2025). Assessing the accuracy of AI tools (Google Translate and Gemini) in translating Arabic idiomatic expressions and proverbs into English. International Journal on Artificial Intelligence Tools, 34(04–05).
[4] Alazzam, T. S., Alzghoul, M. A., & Alzghoul, R. M. (2025). Exploring AI's capability in translating English metaphors into Arabic. Theory and Practice in Language Studies, 15(7), 2319–2325.
[5] Matan, P., & Velvizhy, P. (2025). A neuro-symbolic AI approach for translating children's stories from English to Tamil with emotional paraphrasing. Scientific Reports, 15(1), 20348.
[6] Sofian, E., Nababan, M., Santosa, R., & Djatmika. (2025). Translating politeness: Mapping the characterisation of Klara the AI from Klara and the Sun and its Indonesian translation. Forum for Linguistic Studies, 7(11).
[7] Heathco, G. J. (2025). AI translated task setup for non-native English speaking students. Communication Teacher, 39(1), 26–32.
[8] Andalib, S., Spina, A., Picton, B., Solomon, S. S., Scolaro, J. A., & Nelson, A. M. (2025). Using AI to translate and simplify Spanish orthopedic medical text: Instrument validation study. JMIR AI, 4, e70222.
[9] Desrine, B. (2025). Translating the Indo-Caribbean in the AI era: The case of Trinidadian Bhojpuri.

ISSN 2616-5783 Vol.9, Issue 3: 32-38, DOI: 10.25236/AJHSS.2026.090305

Archipélies, (19), 114–146.
[10] Peterson, D. J. (2025). Translating Franz Kafka's "Josefine, the Singer or the Mouse People" as a metaphor for AI. Humanities, 14(2), 21.
[11] Buckley, A. O., O'Sullivan, B., McKean, C., & Frizelle, P. (2026). An AI tutorial for speech and language therapists: Translating concepts from the AI literature into accessible knowledge and clinically relevant applications. International Journal of Language & Communication Disorders, 61(2), e70201.
[12] Tao, H. (2025). Google Translate: AI in interpreting and translating for speech-language pathology. Journal of Clinical Practice in Speech-Language Pathology, 27(2), 215–216.