STUDIA UBB PHILOLOGIA, LXX, 4, 2025, p. 197 - 218
(RECOMMENDED CITATION)
DOI:10.24193/subbphilo.2025.4.11

CULTURAL SENSITIVITY IN AI TRANSLATION OF CHINESE LANGUAGE: EVALUATING AI'S CAPABILITY TO GRASP CULTURALLY RICH TEXTS

Iulia Elena CÎNDEA[1], **Jiong WANG**[2]

Article history: Received 13 June 2025; Revised 12 November 2025; Accepted 20 November 2025; Available online 12 December 2025; Available print 30 December 2025

ABSTRACT. ***Cultural Sensitivity in AI Translation of Chinese Language: Evaluating AI's Capability to Grasp Culturally Rich Texts***. The widespread adoption of Artificial Intelligence (AI) in translation has fundamentally reshaped global communication by enhancing efficiency, cost-effectiveness, and accessibility. Despite these advancements, critical questions persist about AI's capacity to navigate the inherent complexities of language, particularly regarding cultural sensitivity and contextual accuracy. These challenges are especially pronounced in Chinese, a language abundant in idiomatic phrases, literary allusions, and culturally embedded references, where mistranslations risk significant misunderstanding.

[1] **Iulia Elena CÎNDEA** is a lecturer of Chinese Language, Translation, and Cultural Studies at Lucian Blaga University of Sibiu, Faculty of Letters and Arts, Chinese Chair. She obtained her Ph.D. from the University of Bucharest, Doctoral School of Literary and Cultural Studies. Her research interests span cultural studies, translation studies, cultural sociology, and cross-cultural communication, with a particular focus on Sino-Romanian intellectual and literary exchanges. She has published on topics including the sociology of translation, China's soft power strategies, and the role of translation in knowledge production. Her current work critically examines the intersections of translation, sinology, and geopolitical influence in post-1989 Romania, investigating how knowledge production on China is shaped by historical legacies, political alliances, and economic interests. She also explores decolonial perspectives in China studies, analyzing how Romanian intellectual traditions engage with and challenge Western-centric narratives while negotiating China's expanding cultural diplomacy. E-mail: iulia.cindea@ulbsibiu.ro

[2] **Jiong WANG** is a lecturer at Beijing Language and Culture University (BLCU), where she teaches university students majoring in English and train in-service and pre-service teachers of English. Wang also teaches short-term international students learning Chinese as a second language and served as Director of the Training Department and the E-learning Department at BLCU. E-mail: wangjiong@blcu.edu.cn

This study examines the capabilities of leading AI translation tools in rendering culturally rich Chinese texts into Romanian and English. Drawing on a curated corpus of classical literature, contemporary literature and news articles, the research evaluates each tool's performance in terms of accuracy, cultural sensitivity, contextual understanding, emotional tone, and overall fluency. Both qualitative and quantitative analyses reveal considerable variation among popular platforms. While high fluency and coherence are attainable, some persistent gaps remain in the handling of idiomatic expressions and nuanced cultural elements. These findings reaffirm the need for expert post-editing and a judicious, human-centered integration of AI into translation workflows.

Keywords: *AI translation tool, cultural sensitivity, hermeneutics, Chinese culturemes, post-editing, hybrid translation*

REZUMAT. ***Sensibilitate culturală în traducerea asistată de IA a limbii chineze: evaluarea capacității de redare a textelor cu densitate culturală*.** Adoptarea pe scară largă a inteligenței artificiale (IA) în traducere a remodelat fundamental comunicarea globală, sporind eficiența, costurile și accesibilitatea. În pofida acestor progrese, persistă întrebări critice privind capacitatea IA de a gestiona complexitățile inerente ale limbajului, în special sensibilitatea culturală și acuratețea contextuală. Provocările sunt deosebit de pronunțate în chineză, limbă bogată în expresii idiomatice, aluzii literare și referințe puternic ancorate cultural, unde erorile de traducere pot genera neînțelegeri majore. Studiul de față examinează capacitățile principalelor instrumente de traducere bazate pe IA în redarea în română și engleză a textelor chineze bogate cultural. Pe baza unui corpus atent selectat, cuprinzând literatură clasică, literatură contemporană și articole de presă, cercetarea evaluează performanța fiecărui instrument în termeni de acurateţe, sensibilitate culturală, înțelegere contextuală, ton emoțional și fluență generală. Analizele calitative și cantitative evidențiază variații considerabile între platformele populare. Deși fluența și coerența ridicate sunt posibile, persistă anumite lacune în tratarea expresiilor idiomatice și a elementelor culturale nuanțate. Aceste constatări reafirmă necesitatea post-editării de către traducători profesioniști și a unei integrări cumpătate, centrate pe om, a IA în fluxurile de traducere.

Cuvinte-cheie: *instrumente de traducere bazate pe IA; sensibilitate culturală; hermeneutică; cultureme chineze; post-editare; traducere hibridă*

Introduction

Translation studies have long rejected a purely linguistic definition of its object, positioning translation as a socially situated, historically contingent practice (Gentzler 2014; Marais 2022). Recent work argues that the arrival of artificial-intelligence systems intensifies this interdisciplinarity, demanding theoretical input from anthropology, ethics, history, and epistemology (Nergaard &

Arduini 2022). Accordingly, research on machine translation (MT) must be anchored in a framework that acknowledges both its algorithmic foundations and its cultural embeddedness (Rundle 2014; Zheng, Tyulenev & Marais 2023).

Empirical studies confirm that state-of-the-art NMT narrows the gap with human translators on surface metrics, yet still underperforms on culturally embedded texts such as classical Chinese literature or dialectal media (Venuti 1995; Castilho et al. 2018).

Among languages that pose such challenges, Chinese stands out for its dense idiomatic legacy, wide array of literary allusions, and culturemes. Converting Confucian moral injunctions, Daoist paradoxes, or even everyday colloquialisms into equally nuanced target languages demands more than grammatical fidelity; it entails preserving the cultural and historical resonance that underpins meaning. Subtle oversights, such as mistranslating a classical phrase or ignoring the emotional register of a literary passage, may distort or occlude the source text's core significance. This problem becomes even more acute for specialized genres like philosophical treatises, classical poetry, or the colloquial dialogues embedded in modern Chinese media.

Responding to these concerns, the present study focuses on five key dimensions of translation quality—idiomatic accuracy, cultural references, contextual understanding, emotional tone, and overall fluency—in order to capture both linguistic precision and deeper interpretive demands (Castilho et al.2018). By systematically evaluating AI outputs for classic Chinese literature (e.g., Confucian texts, Daoist scriptures), contemporary colloquial conversations, mass-media news, and technical manuals, the research aims to ascertain how well current tools address the layered nature of Chinese language across diverse registers and domains.

Central to this investigation is an appreciation of hermeneutical considerations. Recent philosophical debates, particularly those revolving around John Searle's (1980) critique of "strong AI" and Schleiermacher's emphasis on the translator's empathetic engagement, highlight the gap between advanced pattern recognition and genuine semantic or cultural "understanding." Although recent NMT systems show strong gains in surface fluency, they still struggle with historically anchored, affective, and symbolic dimensions, especially outside training domains. This gap underscores the continuing relevance of human expertise—both as post-editors who correct errors and as co-creators who guide AI toward culturally faithful renderings.

Hence, this study explores how effectively current leading AI platforms handle cultural nuance and interpretive subtlety in Chinese-to-English or Chinese-to-Romanian translation. The findings elucidate each system's capabilities and limitations, illustrating why a hybrid approach—in which human translators validate or refine machine output—remains indispensable for culturally saturated

or context-sensitive material. The ultimate objective is to illuminate both the practical potential and the philological constraints of AI translation tools, thereby contributing to ongoing discussions in computational linguistics, translation studies, and cross-cultural communication.

1. Theoretical Background

Recent developments in machine translation have spurred renewed reflection on what precisely constitutes "translation" and how scholarly inquiry might deal with its increasingly transdisciplinary character. While discussions of AI translation often focus on computational or linguistic aspects, emerging perspectives suggest that translation cannot be circumscribed by such discrete disciplinary boundaries (Gentzler 2014; Marais 2022). Instead, it entails broader social, cultural, and historical processes that intersect with ethics, technology, and epistemology (Nergaard & Arduini 2022). Consequently, the study of AI-driven translation must engage with a wider theoretical foundation that includes not only computational linguistics but also translation history, anthropology, and hermeneutics.

This recognition of translation's scope also demands a rethinking of AI-based workflows in terms of historical contextualization and cultural immersion (Rundle 2014). For instance, classical Chinese texts, with their deep intertextual links and symbolic references, cannot be approached solely as computational challenges but must be viewed as historically and culturally embedded phenomena. The conceptual frameworks offered by Zheng, Tyulenev, and Marais (2023) propose "(re-)conceptualizing translation" to underscore the fluid boundaries between translation studies and technology, noting that the interplay of algorithmic processes and translator agency calls for an integrated theoretical lens.

The transition from SMT to NMT represents a fundamental shift in how translation systems process language. Whereas SMT approaches relied on phrase-based probabilistic models often prone to syntactic dislocations and constrained context windows, NMT leverages deep neural networks capable of learning context-aware representations across entire sentences (Bahdanau et al. 2014). A key inflection point in NMT design arrived with the Transformer (Vaswani et al. 2017), whose self-attention mechanism obviates the need for sequential recurrence. By modeling global dependencies, Transformers accommodate more coherent and fluent outputs, thereby addressing some of the limitations of recurrent neural networks (RNNs) and long short-term memory (LSTM) models. Building on the foundations of Transformer architectures, LLMs scale both training corpora and model parameters in ways that substantially enhance zero-shot and few-shot translation capabilities (Brown et al. 2020;

Chowdhery et al. 2022). Empirical evidence indicates that these models not only approximate human-level performance for specific high-resource language pairs but also exhibit emergent behaviors, managing nuanced tasks such as style transfer and context-based disambiguation more effectively than earlier NMT systems. Nonetheless, their escalating size raises computational and ethical challenges, including the significant carbon footprint associated with large-scale model training (Bapna et al. 2022, 266-280; Strubell et al. 2019, 3645–3650) and the propagation of biases embedded within uncurated data (Bender et al. 2021, 5185–5198). Further concerns center on "hallucination," where LLMs produce convincingly phrased yet inaccurate results. These issues underscore the importance of domain- or language-specific tuning, particularly in under-resourced settings, and highlight the ongoing need for human oversight to mitigate errors and ensure responsible deployment (Son & Kim 2023, 574; Guerriro et al. 2023, 1500–1517).

Cultural Sensitivity and Contextual Understanding

Despite the leaps in syntactic fluency, cultural adaptation persists as a formidable challenge. Scholars note that translation tools, even those fine-tuned with domain-specific corpora, can flatten linguistic variety by ignoring deeper cultural references—particularly in the context of literary Chinese or idiomatic expressions. Chinese rhetorical forms and historical allusions often require interpretive insight, not merely lexical equivalences (Venuti 1995). When dealing with Confucian classics or Daoist texts, mismatched conceptual metaphors can cause significant distortions in the target output. Similarly, the multiplicity of dialects in modern Chinese media complicates text standardization, indicating that purely corpus-driven AI systems risk producing literal but context-poor translations.

NMT excels at sentence-level coherence but often struggles to incorporate cultural cues that might be peripheral in a purely statistical sense (Castilho et al. 2018, 241–262). The model may capture grammatical well-formedness while overlooking connotations tied to unique historical events, region-specific humor, or social hierarchies. Thus, context must be understood not only as co-textual references within a particular passage but also as a broader cultural knowledge base. Models that fail to incorporate these deeper layers risk generating translations that appear superficially correct yet diminish the text's cultural and aesthetic depth.

Human-based evaluations remain indispensable in MT research, incorporating expert reviewers or professional translators who can detect omissions, distortions, or culturally insensitive phrasing (Zaretskaya 2018, 139-160). Studies

on post-editing processes reveal that advanced NMT outputs, while proficient, still require interventions to restore nuance or reintroduce connotations lost in the algorithmic leap from source to target. Furthermore, domain specialization (e.g., legal or medical) and culturally laden material exacerbate the mismatch between raw AI translations and the contextual adequacy demanded by stakeholders.

A central question in hermeneutical inquiry is whether neural architectures can truly "understand" a text's cultural substrata or whether they merely produce outputs that appear syntactically correct (Searle 1980, 417-424; Bender & Koller 2020, 5185-5198). Searle's Chinese Room Argument, in which Searle imagines himself in a room with a set of rules in English for manipulating Chinese characters and although he does not understand Chinese, he can follow the rules to produce appropriate Chinese responses to Chinese inputs. To an outside observer, it appears as if Searle understands Chinese, but in reality, he is merely following syntactic rules without any understanding of the semantics (Searle 1980). The experiment demonstrates how a system proficient in symbol manipulation may lack genuine intentionality, while Lonergan's (1958) account of intentional consciousness suggests that meaningful comprehension involves self-reflection, moral agency, and experiential depth. Despite major strides in pattern recognition, today's large language models have no intrinsic access to these human dimensions of understanding. Consequently, even advanced tools can overlook the deeper cultural or philosophical resonances central to classical Chinese works and other context-rich texts.

From a hermeneutic standpoint, scholars such as Schleiermacher (1838) and Gadamer (1960) emphasize that interpretation is inextricably tied to the translator's existential horizons and empathetic engagement with the author's voice. Language models like GPT-4 excel at mapping linguistic patterns (Bender & Koller 2020), yet they cannot internalize the lived, historical standpoint that undergoes human interpretation. This theoretical gap manifests practically when AI systems generate translations that, while grammatically sound, fail to capture emotional subtleties, ethical nuances, or symbolic layers. Researchers therefore continue to advocate a "human-in-the-loop" paradigm, ensuring that professional translators and domain experts refine AI-generated text in order to preserve moral, cultural, and existential dimensions of meaning (Zaretskaya 2018).

The longstanding philosophical question of whether AI truly "understands" language (Searle 1980) gains renewed significance in light of large language models that convincingly emulate interpretive depth. Bender and Koller (2020) present a sharp critique of conflating form with meaning, emphasizing that pattern recognition alone does not equate to genuine semantic awareness. Such arguments resonate with hermeneutic theories put forth by Gadamer (1960) and Schleiermacher (1838), who prioritize the translator's existential horizon and empathetic immersion—dimensions that purely data-driven models cannot replicate.

From an ethical perspective, Floridi (2013) accentuates how accountability and transparency become central concerns when machines process culturally or historically sensitive material. If the technology mechanically flattens nuanced references or perpetuates biases, the cost may be eroded cultural integrity and ethical oversight. The balancing act between leveraging AI's efficiency and preserving the moral and existential dimensions of translation emerges as a major concern for practitioners and ethicists alike.

Cognizant of these limitations, Green et al. (2014) demonstrate that post-editing by professional translators improves both the efficiency and accuracy of machine outputs, especially in content where figurative language or domain-specific knowledge is non-negotiable. This collaborative approach, explored in multiple studies, underscores the practical synthesis of high-speed automation and interpretive fidelity, the latter being a hallmark of human expertise in reading cultural cues and rhetorical nuance.

2. Methodology

In order to examine the nuanced challenges that culturally rich Chinese texts pose to AI translation tools; this study adopted a multifaceted methodological approach. The ensuing sections describe the selection of the corpus, the choice of translation systems, and the criteria employed for both quantitative and qualitative evaluation. Although these procedures are structured around established best practices in machine translation research, they are equally informed by insights drawn from philological scholarship, which underscores the importance of historical context, idiomatic richness, and interpretive depth in textual analysis.

2.1. Corpus Selection

To evaluate whether AI systems preserve *cultural meaning* we require texts that systematically elicit (a) idioms and culturemes, (b) intertextual and historical allusions, (c) register and politeness management, (d) elliptical/zero-pronoun structures, and (e) emotionally marked narrative voice. Chinese provides all of these in concentrated form, but no single genre covers them evenly. We therefore constructed a two-part corpus that intentionally spans high-context literary discourse and lower-context expository prose, enabling a contrast between culture-dense and culture-light conditions while holding Chinese as the source language constant.

1. Literary texts:
    - Classical literature subcorpus (sourced from the *Chinese Text Project*): *The Analects* (论语, 2.1; 2.4), *Dao Dejing* (道德经, ch. 1), *The Art of War* (孙子兵法, opening aphorisms), *Dream of the Red Chamber* (红楼梦, Baoyu passages). These provide aphoristic density, parataxis, archaic lexicon, and canonical allusions that are well known to challenge literal systems.
    - Contemporary literature subcorpus (five excerpts): Mo Yan (L*ife and Death Are Wearing Me Out*), Yu Hua (*To Live*), A Cheng (*The Chess King*). These authors contribute colloquial idiom, rural/dialectal flavor, euphemism, irony, and culture-bound lexis, i.e., phenomena where tone and social meaning are easy to flatten.
2. Non-literary articles (sourced from *The Chairman's Bao*): five short texts on education policy, public health, social issues, and technology. These balance the corpus with formal register, domain terminology, and explicit propositional content—useful as a *control* condition in which we expect systems to perform comparatively well.

Selection followed three practical principles. First, phenomenon coverage: each passage was chosen because it exemplifies at least one of the target stressors (idioms/allusions, politeness and particles, ellipsis/zero pronouns, emotionally salient voice). Second, comparability: passages were kept short and self-contained to minimize length effects and facilitate consistent rating across systems and target languages. Evaluation required stable anchors. Each criterion was scored on a five-point scale (1–5) from poor to exceptional. For the literary set, published translations by experienced sinologists in English and Romanian served as primary references. For the news items, editor-style references ensured consistent terminology. These references ground judgments of idiomatic accuracy, cultural faithfulness, contextual understanding, emotional tone, and overall fluency, so that scoring reflects interpretive adequacy rather than surface overlap.

Finally, the corpus is deliberately bilingual on the target side—English as a high-resource language and Romanian as a comparatively lower-resource one—to expose how resource asymmetries affect idiom mapping, register calibration, and error profiles. In sum, the literary–news pairing offers a diagnostic "challenge-plus-control" design. If modern LLMs truly surpass earlier baselines, the gains should be visible not only in the expository set but, crucially, in the culture-dense classical and contemporary fiction where meaning is inseparable from context, intertext, and social pragmatics. This corpus therefore fits the study's aim: it makes cultural sensitivity measurable, comparable across systems, and attributable to more than general improvements in fluency.

2.2. Translation Tools

We evaluated a mixed set of neural machine translation (NMT) and large-language-model (LLM) systems selected for global prominence, distinct architectures, and widespread use among both non-specialists and professional translators. The pool comprised Google Translate, Baidu Translate, DeepL Translator, and Wordvice AI (NMT/MT services), alongside the newer strong-LLM: GPT-4, Vertex AI API, and DeepSeek.

To keep comparisons meaningful across genres, we adopted a strand-specific inclusion policy. All seven systems were screened on the classical Chinese strand. Because Google, Baidu, and DeepL exhibited floor effects there, subsequent strands (contemporary novels and non-literary texts) were restricted to the strong-LLM trio, which are closer peers in capability.

2.3. Evaluation Framework

In line with recognized practices for translation quality assessment, this study employed a multi-dimensional framework, adapting elements from Castilho et al. (2018) and Son & Kim (2023). The criteria were chosen specifically to illuminate whether an AI system could handle more than mere surface-level accuracy—namely, how it performs in domains of idiomatic usage, cultural context, and expressive fidelity.

1. Idiomatic Accuracy

Particular attention was given to the manner in which AI systems interpreted set phrases (成语), proverbs, or culturally contingent metaphors. An accurate rendering of idiomatic language serves as a key indicator that the system grasps not only lexical correspondences but also the underlying cultural or rhetorical function of such expressions.

2. Cultural References

As both classical and modern texts frequently employ historical or social allusions, translations were scrutinized for evidence that the cultural essence was preserved or rendered intelligible in the target language. Cases of mistranslation often arose from literal or decontextualized output, in which the immediate cultural or historical connotation was lost.

3. Contextual Understanding

Effective translations require recognizing contextual cues that inform word sense disambiguation, register shifts, and pragmatic subtleties. These range from understanding the domain (technical vs. literary) to detecting implicit references within a dialogue (e.g., forms of address or shifts in politeness).

4. Emotional Tone

Especially relevant for literary and journalistic texts, emotional tone encapsulates the tenor or emotive force intended by the original author. Whether conveying reverence in a Confucian passage or the urgency of a public health notice, capturing tone ensures the reader perceives the text's affective dimension authentically (Venuti 1995).

5. Overall Fluency

Finally, fluency describes the extent to which translations read as coherent, grammatical, and stylistically appropriate in the target language. While fundamental to any translation evaluation, fluency alone does not guarantee cultural or contextual fidelity (Castilho et al., 2018), hence its integration alongside other criteria.

3. Results

3.1. Classical literature

A careful appraisal of five classical literature excerpts reveals significant divergences in how current AI translation tools handle cultural nuance, idiomatic patterns, and contextual accuracy. These results echo earlier observations indicating that deeper semantic modeling is indispensable when facing philosophically and culturally rich material.

Comparative readings of *The Analects* (verses 2.1 and 2.4) highlight the inherent challenge of conveying Confucian moral teachings in translation. Expressions such as the "North Star" metaphor exemplify pitfalls for literal or decontextualized machine processing. For instance, Google Translate handles English syntax with relative ease but sometimes falters in Romanian by presenting "Beichen" untransformed, diluting Confucius's intended cultural resonance. Baidu Translate likewise produces syntactically coherent yet semantically imprecise renderings, occasionally misrepresenting key life stages in Confucius's self-reflection.

The cosmological image "居其所而众星共之" means the North Star stays fixed and other stars align/revolve by it. Literal renderings flatten cosmology into *location* ("lives there") and misconstrue "共之" as *share it*, losing the idea of moral centripetal order. Google (RO) leaves "Beichen" unassimilated and literalizes structure („Beichen rămâne la locul său, iar stelele îl împart.") which sounds like property division, not alignment. Baidu (EN/RO) similarly yields "*resides in its place and the stars share it*." By contrast, GPT-4, Vertex AI, DeepSeek more effectively retain the metaphorical sense of governing "like the

North Star," integrating context awareness in line with hermeneutical demands for interpretive depth (Gadamer 1960). These models anchor to the spatial frame evoked by "居" ("reside") and to the high-frequency English verb "share," rather than the Classical Chinese astronomical-moral frame ("orient / revolve / pay court").

A parallel challenge appears in rendering the *Dao Dejing*, whose aphorisms are elliptical and paradox-driven. Google Translate and Baidu Translate sometimes collapse paradox into tautology in both Romanian and English, thereby diluting the Daoist tension between "the nameless" and "the named." The line 「道可道，非常道；名可名，非常名。」 is a case in point: we repeatedly obtain outputs like "Calea care poate fi spusă este Calea fără nume," i.e., "The Dao that can be named is the nameless Dao," which erases the contrast "not the *constant* Way/Name." The error stems from mis-scoping negation: in 非常道 / 非常名, 非 modifies 常 ("constant"), yielding "The Way that can be spoken is not the constant Way; the Name that can be named is not the constant Name," but the systems attach 非 to 道 / 名 and default to the high-frequency collocation "nameless Dao," collapsing polarity. This is emblematic of what Searle (1980) calls "syntactic manipulation": fluent strings without semantic penetration. By contrast, Vertex AI and GPT preserve the paradox intact -*the Way that can be spoken is not the constant Way; the Name that can be named is not the constant Name*—suggesting that broader training corpora and stronger context modeling yield higher fidelity on culturally dense material.

The Art of War introduces yet another layer of complexity by combining specialized military terminology with succinct, declarative phrasing. While Google Translate performs reasonably in English, it slips into awkward literalism in Romanian (e.g., "Soldații sunt chestiuni importante ale țării"). Baidu Translate struggles to parse essential strategic vocabulary, distorting references such as "five constant factors." Tools like DeepL reach moderate clarity but sometimes soften Sun Zi's didactic tone. Wordvice AI, featuring domain-specific adaptability, produces concise and coherent outputs, yet strong LLMs again maintain the highest consistency by preserving both rhetorical gravity and lexical precision.

Turning to *Dream of the Red Chamber*, the emotive, colloquial style of passages such as "girls are made of water, boys are made of mud" demands sensitivity to figurative language. Google Translate occasionally yields disjointed Romanian phrasing, reducing the text's lyrical quality. Baidu Translate, for example, unwittingly introduces extraneous characters or familial references that do not exist in the original, skewing the novel's depiction of gender roles and social critique. Strong LLMs more successfully capture Jia Baoyu's nuanced perspective, conveying both freshness and repulsion in a manner that aligns with the text's heightened emotional register.

3.2. Contemporary novels

Across five passages drawn from Yu Hua, Mo Yan, and Ah Cheng, we treated the paired human translations (Romanian and English) as the benchmark and examined three systems (GPT-4, Vertex AI, DeepSeek) in both languages against five criteria: idiomatic accuracy, treatment of cultural references, contextual understanding, emotional tone, and overall fluency.

In the opening passage from *To Live*, the benchmark preserves both the blunt register and the rural material culture embedded in the Chinese. Key items such as "先生" (polite address), "拉屎" (vulgar), "马桶" (chamber pot), and "粪缸" (the manure vat for night soil) are carried over without euphemism, while the pace of "慢吞吞地" is rendered as a leisurely amble and the comic simile "打着饱嗝…像青蛙叫唤" becomes a frog-like belch. GPT-4 and Vertex match this well in English and in Romanian, keeping the low register and the specific rural container ("hârdăul cu fecale / manure vat"). By contrast, DeepSeek's Romanian dilutes the scene twice: "mergea la necazuri" is a euphemism that does not mean defecate, and "găleata de gunoi" generalizes the specialized "粪缸"; a small agreement slip ("râgâind ca un broască") further cools the voice. These micro-choices matter because the passage's effect depends on the friction between social prestige ("domnule / sir") and a bodily routine described in plain farm language.

The second Yu Hua excerpt magnifies the role of culture-specific references and sonic images. The source's unit of land "亩 (mu)"—the social titles "阔老爷/阔少爷," and the onomatopoeic line "钱在钱上面哗哗地流" together stage memory as material plenitude: copper coins clink ("铜钱碰来撞去"), money rushes over money, and the child's call-and-response—"'爷爷，你为什么动呀？' '是风吹的。'"—anchors the tenderness of the scene. The benchmark keeps *mu* in English and Romanian, lets the titles breathe as "rich old master / stăpânul bogat," and reproduces the money-soundscape rather than replacing it with abstraction. GPT-4 aligns most closely: it preserves *mu* transmits the dialogue verbatim and maintains the narrator's deadpan warmth. Vertex is equally strong in Romanian, but its English replaces *mu* with "acres" and over-converts the quantity; this looks helpful but is materially inaccurate and erases a visible Chinese cultureme. DeepSeek is broadly correct in both languages, though it tends to add explanatory phrasing that slightly formalizes the rustic narrator. Here the criteria pull together, as illustrated also in table 1: idiomatic accuracy is high across models, yet the best handling of cultural reference is the one that resists domestication, keeps *mu*, and trusts a brief gloss to do the intercultural work.

Table 1. Comparison of key source concepts in a contemporary literary excerpt (Yu Hua, To Live), contrasting human EN/RO baselines with GPT-4, Vertex, and DeepSeek.

Source concepts	Human EN baseline	Human RO baseline	GPT-4	Vertex	DeepSeek
打着饱嗝…像青蛙叫唤	"belch… like a frog's croak"	„râgâit… ca orăcăitul de broască"	Accurate, neutral cadence	Accurate, natural rhythm	Accurate; sometimes "echoing like" - stylistic
阔老爷/阔少爷	"rich old master / rich young master"	„stăpânul… tânărul stăpân"	EN/RO accurate	EN/RO accurate	EN/RO accurate
钱在钱上面哗哗地流	"money… rushing over money"	„banii curg… zornăie"	EN accurate, slightly prosy	RO very idiomatic („foşnesc curgând")	EN concise; RO idiomatic
亩	Noted in footnote or gloss	Often implicit	EN sometimes adds a short gloss in note	EN/RO keep „mu," no over-explanation	EN/RO keep „mu"; may add brief gloss

The first Mo Yan passage from *Life and Death are Wearing Me Out*—crisp skin in hot oil, "肌肉爆裂的噼啪声," the "贪官污吏," and the vow to be ground to powder under a "石磨" or pounded to paste in an "铁臼"—tests whether systems can balance grotesque physicality with folk-Buddhist cosmology ("阎王"). The human benchmark does so with unflinching concreteness ("moara de piatră / millstone," "piuă/mojar / mortar") and a sardonic, stoic voice ("To hell with them!"). GPT-4 and Vertex reproduce the tool pair accurately in both languages and keep "King of Hell / Regele Iadului," avoiding secular dilution. DeepSeek's Romanian slips on the implement—rendering "铁臼" as "piuliță" (a nut), not *piuă/mojar*—a precise example of how world-knowledge about material culture feeds idiomatic accuracy. The consequence is more than terminological: the image of bodily reduction into paste depends on the mortar and pestle pairing, not on a hardware fastener.

In the companion Mo Yan courtroom passage from the same novel (see table 2), the evaluation turns on three details: cadence, the courtroom object, and the semantics of injustice. The benchmark hears "连珠炮般的话语" as a machine-gun flow, keeps the magistrate's "惊堂木" as a specifically Chinese sounding block (the Romanian explains it succinctly as "plăcuța de lemn cu care se cere ordine în sală"), and renders "冤枉" as *wronged* rather than merely *innocent*. GPT-4 handles all three with poise; when it uses "gavel," it often

appends "(sounding block)," which is an acceptable compromise. Vertex matches the rhythm but occasionally paraphrases "冤枉" as "wrongful death," shifting the meaning from moral injustice to legal causation and thereby narrowing the text's karmic dimension. DeepSeek remains fluent but defaults to the generic "gavel," losing the cultural grain of the Chinese courtroom.

Table 2. Comparison of key source concepts in a contemporary literary excerpt (MO Yan, Life and Death are Wearing Me Out), contrasting human EN/RO baselines with GPT-4, Vertex, and DeepSeek.

Source concepts	Human EN baseline	Human RO baseline	GPT-4	Vertex	DeepSeek
焦干…油汪里…肌肉爆裂的噼啪声	"scorched… slick of oil… muscles bursting (popping)"	„pârjolit… baltă de ulei… mușchi care plesnesc"	EN vivid, tight cadence	EN/RO vivid; RO keeps "unsuros" later	EN sharp; RO "pocnetele mușchilor"
阎王 / 判官	"King of Hell / judges"	„Regele/Împăratul Iadului; judecătorii"	Terminology consistent	Terminology consistent	Terminology consistent
连珠炮般的话语	"rapid-fire barrage"	„ca un foc neîntrerupt / șuvoi de cuvinte"	EN "machine-gun-like torrent"	EN/RO good	EN good; RO good
惊堂木	"gavel / sounding block"	„bloc/plăcuță de lemn / masa judecătorească"	EN "gavel" (acceptable domestication)	EN "gavel," RO "blocul de lemn"	EN "gavel," RO "ciocănelul"
冤枉	"wronged / falsely accused"	„nedreptățit / nevinovat"	Accurate	Accurate	Accurate

Ah Cheng's *The Chess Master* paragraph tests institutional literacy (for reference, table 3). The source is dense with Cultural Revolution nomenclature: "插队" (being sent to production brigades in the countryside), "运动" (the Movement), "留城政策" (policy allowing only children to remain in the city), and "机关的铝牌编号" (work-unit inventory tags). The benchmark names the period explicitly in Romanian, retains the administrative texture in English, and keeps the lupine self-image of "野狼似的转悠." All three systems perform competently here in both languages. GPT-4 and Vertex keep the policy nuance and the tag detail; DeepSeek's English usefully mirrors the agentive "我送走" ("sent off by me"), but its Romanian ends one sentence with the slightly meta "într-un mod numit corect," where the benchmark's "cel puțin în acte, pe deplin justificat" is more idiomatic and bureaucratically faithful.

Table 3. Comparison of key source concepts in a contemporary literary excerpt (Ah Cheng The Chess Master), contrasting human EN/RO baselines with GPT-4, Vertex, and DeepSeek.

Source concepts	Human EN baseline	Human RO baseline	GPT-4	Vertex	DeepSeek
插队 (sent-down to countryside)	"sent off to the countryside to join production brigades"	„trimişi la brigăzile de producţie"	EN/RO correct	EN/RO correct	EN/RO correct
污点 (political blemishes)	"black marks / stains"	„pete politice"	Accurate	Accurate	Accurate
运动 (the Movement)	"the movement" with CR implicature	„Mişcarea" (implicit CR)	EN "Movement (运动)" clarifies	RO keeps implicit; context clear	EN/RO clear
机关的铝牌编号	"work-unit aluminium inventory tags"	„plăcuţe de inventar ale instituţiei"	EN preserves state ownership nuance	RO preserves nuance	EN/RO preserve nuance
留城政策	"policy allowing… to remain in the city"	„politica de a rămâne în oraş"	Accurate	Accurate	Accurate

Taken together, these passages show consistent strengths and bounded weaknesses. Idiomatic accuracy is now high for all three systems in English and, with occasional dips, in Romanian. The most frequent failures are small nouns that carry a great deal of world knowledge. Cultural references are best served when systems resist over-domestication. Contextual understanding is generally robust—the prestige/earthiness irony in Yu Hua and the karmic justice frame in Mo Yan are correctly read—though the "wrongful death" paraphrase shows how semantic neighborhoods can mislead. Emotional tone is where small register choices have outsized effects: a single euphemism ("mergea la necazuri") or a lyrical embellishment that the source does not license can shave the rustic deadpan or the grotesque stoicism that define these narrators. Overall fluency remains strongest in English; Romanian is solid but more vulnerable to agreement errors and to the loss of rural or craft vocabulary.

For the purposes of this study, these examples do more than illustrate score differences; they explain them. When a system names the right object,

keeps the right unit, and holds its nerve on register, it reads as culturally literate and emotionally faithful—and its scores rise across all five criteria. Where it converts away a cultureme, generalizes a tool, or softens the register, the loss propagates: idiomatic accuracy drops, cultural reference weakens, tone shifts, and fluency begins to feel generic. The upshot is practical. With light, targeted post-editing GPT-4 and Vertex routinely approach the benchmark translations in both languages, while DeepSeek benefits most from Romanian register control and concrete-noun verification Their reliance on extensive multilingual corpora and robust neural architecture seems to support greater flexibility with archaic forms and metaphorical references**.** Figure 1 illustrates the scores across the five criteria obtained by each of the LLM tools.

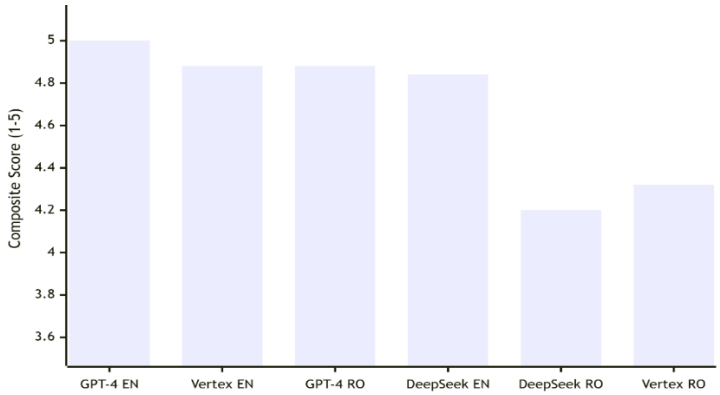

Figure 1. Performance of strong LLM Translation Tools throughout five criteria: Idiomatic Accuracy, Cultural Sensitivity, Contextual Understanding, Emotional Tone, Overall Fluency

3.3. Non-Literary texts

Against the same criteria, results converge tightly on the human benchmark: Vertex and DeepSeek are marginally stronger in Romanian, while GPT-4 remains highly competitive in English. Performance clusters around four pressure points: the semantics of policy metaphors, the precision of numerals and units, the handling of institutional terminology, and low-level negation.

The grading-reform notice (The Chairman's Bao 2024b) typifies the metaphor constraint. The source signals a methodological shift to "decouple" (脱钩) evaluation from GPA arithmetic while adopting an A–F letter scheme. The best outputs anchor 脱钩 to GPA—"abandons GPA in favor of letter grades"—and avoid the misleading calque "decuplează studenţii de la note," which would imply abandoning grades altogether. Vertex and DeepSeek handle this anchoring

most consistently in Romanian; GPT-4 is solid in English, preserving policy intent and newsroom neutrality.

The bereavement/re-entry vignette (The Chairman's Bao 2024c) shows how a single noun can skew context. Whereas one human draft misread 旅游文化班 as a swimming class, all three systems correctly render a travel/tourism culture course, maintain kinship address (阿婆 → "Grandma"/„bunica"), and carry the closure metaphor 走出了人生的阴霾 into natural Romanian (e.g., „a ieşit din umbra vieţii") and fluent English. Here, the models matched a baseline human draft on a terminology point.

Data-dense reporting further narrows the gap. In the excerpt on AI and jobs (The Chairman's Bao 2023), all three models preserve sample size (5,300 workers), scope (seven member countries), the headline statistic (60%), and the conditional stance attributed to Mathias Cormann, with no numeric drift. Likewise, the Beijing vaccination notice (The Chairman's Bao, 2021) tests polarity and units; the LLMs retain the two-dose schedule, keep the fee exact (93.5 yuan ≈ $14.2 per dose), and correctly scope eligibility around 医保 ("enrolled in Beijing medical insurance" / „înscrişi în sistemul de asigurări medicale"), avoiding a polarity flip present in one human draft.

Finally, the cultural-heritage brief on the 200-year-old Chinese wallpaper (The Chairman's Bao 2024a) probes entity chains and chronology. The models consistently maintain Westport / County Mayo, the 1730 build date, the Browne/Hughes lineages, and the conservation quote. Occasional enrichments (e.g., "Westport House") are contextually plausible but should be flagged when strict source fidelity is required. Register remains appropriately restrained.

In short, when terminology is conventional and figures are explicit, GPT-4, Vertex AI, and DeepSeek meet—or slightly exceed—human baselines under light post-editing. Residual, correctable risks are (i) semantic drift in policy metaphors, (ii) context-skewing noun misreads, and (iii) polarity flips in conditional statements. A brief editorial protocol—lock numerals/units verbatim, verify negation, gloss institutional acronyms once, and prefer precise reformulations—consistently lifts all three systems to publication-ready quality across non-literary texts.

4. Discussions

As Fig. 2 illustrates the strong LLMs attain the highest aggregate scores for cultural sensitivity, idiomatic fidelity, contextual alignment, and tonal control, while Google, DeepL, and Baidu are serviceable on straightforward content yet more prone to figurative loss in archaic or philosophically dense texts. This mirrors prior work linking superior outcomes to larger, more diverse corpora and stronger architectures (Castilho et al. 2018; Son & Kim 2023).

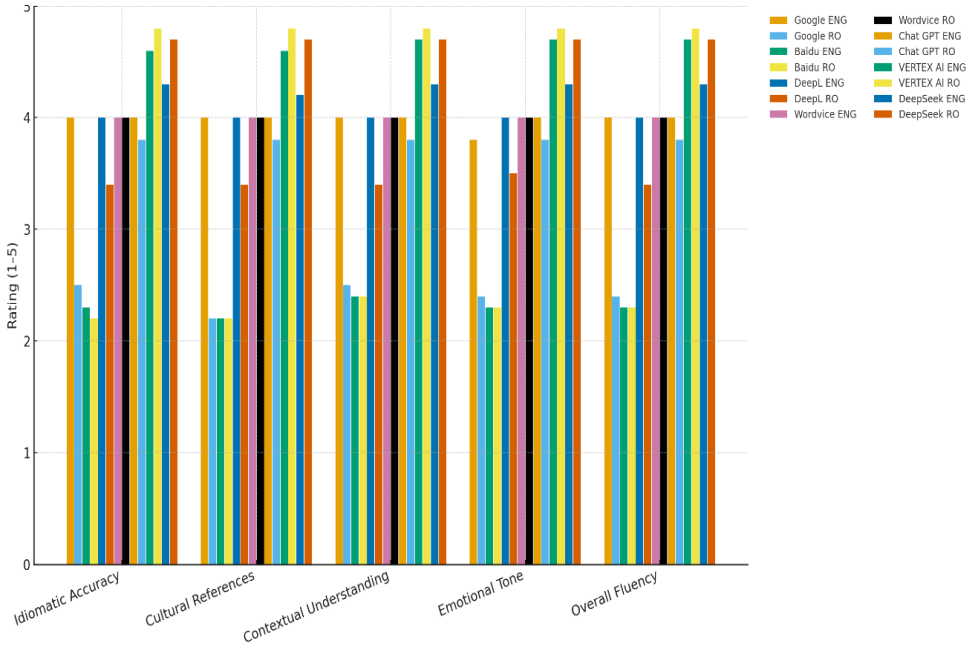

Figure 2. Comparison of Ratings of 6 Translation Tools across Different Criteria According to the Present Case Study

In the classical strand, top systems preserve metaphor scope and culture-specific references; weaker systems default to literal, monosemous renderings. In contemporary fiction, the triad separates under literary pressure along three axes: (i) culturemes/units, (ii) concrete-noun specificity, and (iii) register control. GPT-4 and Vertex more reliably retain culture-bound items and precise implement nouns, sustaining imagery and tone; DeepSeek's Romanian occasionally domesticates units or generalizes objects, softening stylistic effects. In non-literary news, outputs converge on the human benchmark when terminology is conventional and numerals/units explicit.

From a hermeneutic vantage point, these limitations reflect a larger philosophical tension: advanced NMT systems demonstrate remarkable facility at pattern recognition, yet they lack the historically and existentially grounded understanding that shapes deep literary or moral interpretation (Searle 1980; Schleiermacher 1838; Gadamer 1960). As Schleiermacher and Gadamer (1960) argued authentic comprehension transcends syntactic mapping, necessitating lived insight and a horizon of experience. AI, by contrast, operates within algorithmic constraints, capturing "surface equivalences" without fully embodying the cultural memory or ethical vision that animate classical texts and moral teachings.

Even with advanced context modeling, the existential dimension—encompassing self-reflection, intentionality, and moral agency—remains beyond neural mimicry. Accordingly, human oversight is focal: light, targeted post-editing—locking numerals/units, verifying negation, glossing culture-specific artefacts once, and resisting over-domestication—consistently restores rhetorical and cultural fidelity.

Taken together, these three strands suggest a stable division of labor. On classical passages, top systems lead but still require expert post-editing for metaphor scope, diachronic lexicon, and culturemes. On non-literary news, the strong-LLM triad reaches near-parity with humans under a light checklist. On contemporary novels, fine-grained cultural literacy separates GPT-4/Vertex from DeepSeek in Romanian, with GPT-4 generally most stable in English and Vertex strongest in Romanian cadence.

From a broader vantage, domain adaptation—whether via specialized subcorpora, glossaries of archaic Chinese, or bilingual commentary—could mitigate many of these diachronic mismatches. Yet the consistent need for philological discernment confirms that bridging centuries of linguistic evolution demands more than raw computational might. While scaling up corpora or refining contextual embeddings can somewhat lessen anachronisms, the interpretive subtlety inherent in classical Chinese—encompassing allusions, metaphorical frameworks, and shifting script reforms—ultimately requires the informed human–AI partnership described throughout this research. By weaving philological insight into large-scale machine translation pipelines, the field may progress toward a more sensitive, historically resonant handling of premodern Chinese texts, fully honoring the language's diachronic richness. In short, large models narrow—but do not erase—the gap between structural equivalence and cultural resonance; targeted human intervention remains the decisive step from fluency to fidelity.

Conclusion

Advances in neural architectures and large language models have propelled AI translation to new heights of fluency and efficiency, yet this study affirms that fully capturing the cultural, idiomatic, and contextual richness of Chinese texts remains elusive. In our design, six systems were evaluated on classical texts; after observing floor effects for Google Translate, DeepL, and Baidu Translate, we restricted the contemporary-novel and non-literary strands to the strong-LLM triad (GPT-4, Vertex AI, DeepSeek). While GPT-4, Vertex AI, and DeepSeek demonstrate notably higher proficiency, they still fall short of the interpretive depth that human translators provide. News articles or straightforward expository content typically benefit most from automated output, requiring minimal post-editing. By contrast, specialized or historically layered genres underscore the need for domain adaptation and expert

intervention, highlighting the persistent gap between surface accuracy and the deeper hermeneutic engagement that brings texts vividly to life.

Limitations include a finite, unidirectional corpus and the absence of targeted fine-tuning experiments. Future work should expand genre coverage, add bidirectional evaluation, quantify inter-annotator agreement on error typologies, and test domain-adapted models for archaic lexicon and cultureme retention.

In sum, large models narrow—but do not erase—the gap between structural equivalence and cultural resonance. The decisive step from fluency to fidelity remains human: translators ensure that both classical aphorisms and contemporary journalism are rendered not only with linguistic precision, but with the cultural memory and ethical vision that give these texts their enduring force.

We advocate a stance of critical openness toward AI in translation. Welcoming the efficiencies of LLM-assisted workflows does not diminish the translator's centrality; rather, it intensifies professional obligations. The human translator retains ultimate ethical and moral responsibility for the published text and must demonstrate heightened linguistic, cultural, and methodological competence—capable of auditing LLM outputs for hallucinations, conceptual drift, inappropriate domestication, register shifts, and other category errors. The question, therefore, is not whether to oppose LLMs, but how to integrate them responsibly. This entails principled protocols for their use—transparent prompts, documented post-editing, verifiable decision trails, and clear lines of accountability—so that AI functions as an assistive companion under informed human supervision, and fidelity to meaning, culture, and context remains securely human-guaranteed.

WORKS CITED

Ah Cheng. 2005. *《棋王》[The Chess Master. Chinese-English Bilingual Edition].* The Chinese University of Hong Kong Press.

Bahdanau, D., K. Cho, and Y. Bengio. 2014. "Neural Machine Translation by Jointly Learning to Align and Translate." *arXiv* preprint arXiv:1409.0473. https://arxiv.org/abs/1409.0473.

Bapna, A., et al. 2022. "Multilingual Mix: Example Interpolation Improves Multilingual Neural Machine Translation." In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*: 266–280.

Bender, E. M., and A. Koller. 2020. "Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*: 5185–5198.

Bender, E. M., et al. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency:* 610–623.

Brown, T., et al. 2020. "Language Models Are Few-Shot Learners." In *Proceedings of the 34th Conference on Neural Information Processing Systems*.

Cao Xueqin. 2025. "Hongloumeng 紅樓夢 (Dream of the Red Chamber)." *Gushiwen*. Accessed March 6, 2025. https://m.gushiwen.cn/guwen/bookv_ca3ee0e7ebd7.aspx.

Castilho, S., et al. 2018. "Evaluating MT for Professional Use." In *Translation Quality Assessment: From Principles to Practice*, edited by J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty, Springer: 241–262.

Chowdhery, A., et al. 2022. "PaLM: Scaling Language Modeling with Pathways." *arXiv* preprint arXiv:2204.02311. https://arxiv.org/abs/2204.02311.

Confucius. n.d. "Lunyu 論語 (The Analects), Book 2 'Wei zheng 為政,' 2.1 and 2.4." *Chinese Text Project*. Edited by Donald Sturgeon. Accessed March 2, 2025. https://ctext.org/lunyu-zhushu/wei-zheng/ens.

Floridi, L. 2013. *The Ethics of Information*. Oxford University Press.

Gadamer, H.-G. 1960. *Truth and Method*. Seabury Press. (Original work published 1960.)

Gentzler, E. 2014. "Translation Studies: Pre-Discipline, Discipline, Interdiscipline, and Post-Discipline." *International Journal of Society, Culture & Language* 2 (2): 13–24.

Green, S., J. Heer, and C. D. Manning. 2014. "The Efficacy of Human Post-Editing for Language Translation." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems:* 439–448.

Guerreiro, Nuno M., et al. 2023. "Hallucinations in Large Multilingual Translation Models." *Transactions of the Association for Computational Linguistics* 11: 1500–1517.

Laozi. n.d. "Daodejing 道德經, chap. 1." *Chinese Text Project*. Edited by Donald Sturgeon. Accessed March 4, 2025. https://ctext.org/wiki.pl?if=gb&chapter=492483&remap=gb.

Lonergan, B. 1958. *Insight: A Study of Human Understanding*. Philosophical Library.

Marais, K., ed. 2022. *Translation beyond Translation Studies*. Bloomsbury.

Mo Yan. 2006. *《生死疲劳》[Life and Death are Wearing Me Out. Chinese Edition].* Zhejiang Publishing House.

Nergaard, S., and S. Arduini. 2022. "Translation: A New Paradigm." *Translation. A Transdisciplinary Journal* 1 (1): 8–17.

Rundle, C. 2014. "Theories and Methodologies of Translation History: The Value of an Interdisciplinary Approach." *The Translator* 20 (1): 2–8.

Schleiermacher, F. 1998. *Hermeneutics and Criticism and Other Writings*. Edited and translated by A. Bowie. Cambridge University Press.

Searle, J. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417–424.

Son, J., and B. Kim. 2023. "Translation Performance from the User's Perspective of Large Language Models and Neural Machine Translation Systems." *Information* 14 (10): 574.

Strubell, E., A. Ganesh, and A. McCallum. 2019. "Energy and Policy Considerations for Deep Learning in NLP." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*: 3645–3650.

Sunzi. n.d. "Sunzi bingfa 孫子兵法 (The Art of War), chap. 1 'Ji 計 (Laying Plans).'" *Chinese Text Project*. Edited by Donald Sturgeon. Accessed March 2, 2025. https://ctext.org/wiki.pl?if=gb&chapter=934017&remap=gb.

The Chairman's Bao. 2021. "Beijing Opens Vaccination Applications to Foreigners." *The Chairman's Bao*. Accessed October 24, 2025. https://www.thechairmansbao.com/beijing-opens-vaccination-applications-to-foreigners.

The Chairman's Bao. 2023. "Study Finds Up to 27% of Jobs Could Be Lost to AI." *The Chairman's Bao*, July 13. Accessed October 24, 2025. https://www.thechairmansbao.com/study-finds-up-to-27-of-jobs-could-be-lost-to-ai.

The Chairman's Bao. 2024a. "200-Year-Old Hand-Painted Chinese Wallpaper Undergoes Restoration at Irish Home." *The Chairman's Bao*, May 16. Accessed October 24, 2025. https://www.thechairmansbao.com/200-year-old-hand-painted-chinese-wallpaper-undergoes-restoration-at-irish-home.

The Chairman's Bao. 2024b. "Chinese Universities Discuss Dropping GPA Grading for Applicants." *The Chairman's Bao*. Accessed October 24, 2025. https://www.thechairmansbao.com/chinese-universities-discuss-dropping-gpa-grading-for-applicants.

The Chairman's Bao. 2024c. "Elderly Woman Defies Expectations with Active Lifestyle." *The Chairman's Bao*, May 23. Accessed October 24, 2025. https://www.thechairmansbao.com/elderly-woman-defies-expectations-with-active-lifestyle.

Vaswani, A., et al. 2017. "Attention Is All You Need." In *Proceedings of the 31st Conference on Neural Information Processing Systems*: 5998–6008.

Venuti, L. 1995. *The Translator's Invisibility: A History of Translation*. Routledge.

Yu Hua. 2017. *《活着》 [To Live. Chinese Edition].* Beijing: Beijing October Literature and Art Publishing House.

Zaretskaya, A. 2018. "Machine Translation Post Editing at TransPerfect." In *Translation Quality Assessment: From Principles to Practice*, edited by J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty, Springer: 139–160.

Zheng, B., S. Tyulenev, and K. Marais. 2023. "Introduction: (Re-)Conceptualizing Translation in Translation Studies." *Translation Studies* 16 (2): 167–177.