Journal of Literature and Art Studies, August 2026, Vol. 16, No. 8, 660-665
doi: 10.17265/2159-5836/2026.08.004

Ethical Dilemmas and Paths to Reconstruction in the Era of AI-Assisted Translation

CUI Wei-xia
Xi'an International Studies University, Xi'an, China

The shift toward Large Language Model (LLM)-driven "human-machine synergy" has enhanced translation efficiency but triggered unprecedented ethical dilemmas. This paper systematically examines four major crises in AI (artificial intelligence)-assisted translation: the diffusion of responsibility, privacy vulnerabilities stemming from digital extractivism, cultural flattening driven by algorithmic hegemony, and translators' cognitive deskilling. To resolve these challenges, we propose reconstructing the ethical framework across four dimensions: implementing explainable AI architectures, establishing granular liability and fair data policies, reaffirming the "Human-in-Command" principle to restore human agency, and fostering a multi-stakeholder governance ecosystem. Ultimately, this study provides actionable pathways for building a trustworthy, culturally responsible translation ecosystem that balances computational efficiency with social accountability.

Keywords: AI-assisted translation, generative AI, ethical dilemmas, human-machine synergy

Introduction

In the wake of the rapid advancement of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs), translation has shifted from a simple bilingual conversion skill to a collaborative human-machine cognitive process. Driven by high efficiency and extensive language coverage, AI-assisted translation has significantly lowered the costs of cross-linguistic communication, becoming a core infrastructure for cultural exchange and global commerce. However, this technological boom has also sparked unprecedented ethical concerns. While it greatly boosts translation productivity, it introduces a series of ethical issues, such as ambiguous accountability, algorithmic bias, and the erosion of the translator's subjectivity (Lu, 2024; Wu & Chen, 2023; Moorkens, 2022). These challenges actively test the limits of traditional translation ethics. In light of the paradigm shift in the AI era, this paper systematically examines the ethical dilemmas facing AI-assisted translation and explores practical ways to rebuild its ethical framework across technological, institutional, humanistic, and governance dimensions. Ultimately, this study aims to balance technological utility with human values, fostering a translation ecosystem that is both trustworthy and culturally responsible.

Acknowledgments: This research was supported by Humanities and Social Sciences Research Project funded by the Ministry of Education of China [Grant No. 23YJAZH139].
CUI Wei-xia, master degree, associate professor, School of English, Xi'an International Studies University, Xi'an, China.

The Paradigm Shift of Translation in the AI Era

The transition from traditional Computer-Assisted Translation (CAT) and standard Neural Machine Translation (NMT) to Generative AI and Large Language Models (LLMs) marks a watershed moment in the history of translation (Wang et al., 2023). LLMs are no longer mere bilingual dictionaries or sentence-level mapping tools; they function as context-aware cognitive agents capable of stylistic adaptation, zero-shot learning, and logical reasoning. This technological leap has driven the industry from a human-dominated, tool-assisted model to a new paradigm of human-AI machine translation. This paradigm shift fundamentally restructures the translation ecosystem across three dimensions.

From Solitary Creator to Distributed Ecosystem

Under the traditional framework, the human translator acted as the sole creator of meaning, holding complete control over textual interpretation, terminology choices, and cultural adaptation. In the AI era, however, the translation process is no longer centralized. It has evolved into a distributed ecosystem involving algorithm developers who design the core logic, massive training corpora that supply the data, AI models that generate drafts, prompt engineers who provide contextual constraints, and post-editors who refine the final output (Wang & Liu, 2026). As LLMs take on the role of primary text generators to rapidly produce fluent drafts, the human translator's focus shifts from word-by-word conversion to higher-level tasks. They now act as arbiters of meaning and ethical gatekeepers—verifying cultural appropriateness, correcting machine hallucinations, and ensuring tone consistency. This shared agency breaks down the traditional single-translator structure and blurs the lines of authorship. Consequently, it sets the stage for subsequent accountability dilemmas, such as determining who is legally responsible when a machine-generated mistranslation leads to financial or reputational damage.

From Linguistic Decoding to Algorithmic Moderation

AI has triggered a fundamental shift in what it means to translate. Historically, translational competence was largely defined by deep bilingual proficiency and cross-cultural intuition, with translators focusing on manually decoding the source language and encoding the target language. Today, as AI efficiently handles the initial text generation, the core competency required of a human translator is shifting toward AI literacy, risk management, and critical moderation.

Specifically, AI literacy requires translators to understand how Large Language Models work—such as their reliance on statistical probabilities rather than genuine comprehension (Wang & Zhang, 2026). Consequently, translators are no longer primarily evaluated on their ability to write a sentence from scratch. Instead, their professional value lies in their capacity to detect subtle semantic deviations and correct machine hallucinations, where the AI generates text that reads fluently but completely distorts the original facts. Furthermore, risk management and critical moderation require translators to counter systemic biases in the AI's training data. For example, they must intervene when the algorithm over-simplifies culturally loaded terms or neutralizes sensitive historical contexts. Ultimately, translation has transformed from a purely linguistic endeavor into a collaborative practice of human-machine cognitive alignment, where the machine provides the computational processing power, and the human provides the necessary critical oversight, contextual awareness, and ethical judgment.

Ethical Dilemmas of AI-Assisted Translation

Despite the notable efficiency dividends, the paradigm shift toward LLM-assisted translation has exposed numerous normative failures in the ethical dimension. Unlike traditional tools, GenAI operates as a highly autonomous yet opaque system, engendering complex ethical crises that can be categorized into the following four dimensions:

Diffusion of Responsibility and Algorithmic Black Boxes

In traditional translation, the human translator holds undisputed authorship and full accountability for the final text. In the AI-driven ecosystem, however, accountability is severely obstructed by the diffusion of responsibility across a complex supply chain. When a severe mistranslation occurs—such as a critical error in a legal contract or a medical document—it becomes nearly impossible to isolate liability. Should the blame fall on the algorithm developers who designed the system, the data providers who supplied the flawed corpora, the platform operators, or the human post-editors who failed to spot the error? This dilemma is further complicated by the black-box nature of deep neural networks, which makes tracing the exact cause of a specific machine-generated word technically unfeasible. Consequently, a structural asymmetry emerges: AI platforms claim technological success, while human editors are often left to bear the legal and moral risks.

Digital Extractivism and Privacy Vulnerabilities

AI translation systems rely heavily on consuming massive amounts of existing bilingual texts. This heavy reliance leads to a major ethical issue often called digital extractivism. In practice, technology companies frequently harvest large databases of past translations created by human professionals to train their AI models. They often do this without asking for the translators' permission, giving them proper credit, or providing financial compensation, which has sparked ongoing copyright disputes. In addition to copyright concerns, there are significant privacy risks for users. When using AI translation platforms, people often upload highly sensitive materials, such as confidential business contracts, personal information, or internal government documents. Because LLMs can absorb and learn from the data users feed into them, the AI might memorize these private texts. Consequently, there is a real danger that the system could accidentally leak this confidential information when generating translations for other users later on, posing a severe threat to both corporate security and personal privacy (Wang, 2025).

Cultural Flattening and Algorithmic Hegemony

Traditional translation ethics emphasize respecting and preserving the unique cultural features—or cultural otherness—of the source text. However, because LLMs operate on statistical probabilities rather than genuine cultural understanding, they often produce highly standardized outputs. When translating culturally loaded terms, indigenous concepts, or specific historical references, AI systems tend to over-simplify them. They frequently substitute these culturally rich expressions with generic, emotionally neutral words in the target language. Additionally, because English texts heavily dominate the training data of most foundational models, AI systems frequently use English syntax and Western logical structures as a hidden intermediary when translating between other languages. This algorithmic bias often causes non-Western languages to conform to Western rhetorical standards. By standardizing translated discourse in this way, AI translation risks reinforcing a form of digital cultural hegemony in which diverse cultural identities are flattened into a single, dominant framework.

The Illusion of Fluency and Cognitive Deskilling

A unique ethical challenge introduced by LLMs is how they affect the working conditions and cognitive skills of translators. Unlike older machine translation systems that often produced obvious and awkward errors, modern AI generates text that is highly fluent, even when it contains factual mistakes or algorithmic hallucinations. This illusion of fluency makes it much harder for human reviewers to detect underlying errors. Driven by the industry's demand for high efficiency, professional translators are increasingly assigned to fast-paced post-editing of these machine-generated drafts. Working constantly under tight deadlines to correct the machine's plausible but sometimes flawed syntax poses a real risk of cognitive deskilling. Over time, translators may find their independent linguistic intuition weakening, shifting their role from active, creative language professionals to passive reviewers of algorithmic outputs. To address the aforementioned dilemmas, it is imperative to architect a collaborative governance framework defined by the principles of technologically controllable, institutionally clear, human-centric, and multi-stakeholder governance.

Paths to Reconstructing AI-Assisted Translation Ethics

To counteract the algorithmic erosion of translational values and resolve the aforementioned dilemmas, relying solely on technological determinism is insufficient. It is imperative to architect a holistic governance framework that translates abstract ethical principles into concrete, actionable mechanisms across technological, institutional, humanistic, and governance dimensions.

Explainable Architectures and Culturally Aware Pre-training

To address the black box problem and reduce cultural flattening, AI systems need to adopt Explainable AI (XAI) features designed specifically for Large Language Models. In practice, this means creating user interfaces that offer more than just a final translated sentence. Future systems should provide transparent background information, such as confidence scores for specific terms, alternative word choices, and the sources of the training data. This level of transparency makes it much easier for human reviewers to evaluate the output and catch potential errors. Furthermore, developers should actively build "cultural diversity" into the models during the pre-training and fine-tuning stages. Recent studies emphasize that aligning large language models solely with generic datasets often exacerbates cultural biases, making targeted human-AI alignment crucial (Wang et al., 2023). This can be effectively achieved by using Reinforcement Learning from Human Feedback (RLHF) guided specifically by cultural experts and local linguists (Ouyang et al., 2022). By fine-tuning models with high-quality, localized data, and cultural dictionaries, developers can create culturally conscious AI. These models would be better equipped to preserve unique cultural expressions rather than automatically replacing them with generic or Westernized terms.

Granular Liability Delineation and Fair Data Frameworks

Addressing the "diffusion of responsibility" and "digital extractivism" requires the establishment of detailed legal and industry frameworks. To resolve accountability issues, clear liability structures must be created through technological auditing. Industry standards should require the use of digital watermarks or metadata tagging to trace the origin of a translated segment. This would explicitly record whether a sentence was entirely AI-generated, drafted by AI and edited by a human, or fully human-translated. Such traceability allows for clear

Service Level Agreements (SLAs) regarding liability if severe mistranslations occur. Furthermore, fair data policies must be institutionalized to counter extractivist practices. Technology companies should implement robust "opt-out" protocols for data collection and establish transparent data licensing mechanisms. This would allow professional translators to be fairly compensated for their proprietary translation memories (TMs) rather than having their work harvested without permission. Finally, to mitigate privacy vulnerabilities, enterprise-level AI translation platforms must adopt strict zero-retention policies, ensuring that sensitive user inputs are not absorbed into the model's ongoing training data.

Critical Vigilance and Restoring Human Agency

To address the "illusion of fluency" and prevent cognitive erosion, the translation industry must reaffirm the "Human-in-Command" principle, moving beyond a basic "Human-in-the-Loop" approach. This ensures that in high-stakes domains, such as legal, medical, and diplomatic contexts, AI remains strictly an assistive tool, while human professionals retain final decision-making authority to cross-reference and validate machine outputs. To maintain this human agency, translation education and professional training must adapt. University curricula should transition from exclusively teaching bilingual conversion skills to also cultivating algorithmic literacy and critical vigilance (Kenny, 2020). This includes introducing training on automation bias, prompt engineering, and AI risk management. By equipping translators with the critical capacity to identify and correct machine hallucinations, the profession can shift its identity from passive reviewers of algorithmic outputs to proactive ethical gatekeepers and cross-cultural communication experts.

Multi-stakeholder Governance and Digital Linguistic Sovereignty

To address the concentration of power among major technology companies, it is necessary to build a robust, multi-stakeholder governance ecosystem. A collaborative governance model should be established, integrating the efforts of AI developers, global professional translation associations (such as the International Federation of Translators, FIT), and governmental regulatory bodies. Under this framework, AI platforms should be required to open necessary APIs, enabling independent linguistic experts to conduct third-party audits on algorithmic bias. On a macro level, international organizations like UNESCO should facilitate cross-border dialogues to draft an inclusive International Convention on AI Translation Ethics. Crucially, to safeguard digital linguistic sovereignty against the dominance of English-centric models, governments and academic institutions must support open-source, community-driven AI initiatives for low-resource languages. Because AI systems frequently struggle with low-resource languages due to sparse and inconsistent training data, this targeted support is vital. Ultimately, this collective governance is essential to prevent algorithmic dominance and ensure the sustainable diversity of the global linguistic landscape.

Conclusion

The ethical dilemmas surrounding AI-assisted translation highlight a fundamental tension between the rapid expansion of technological capabilities and the need to safeguard core humanistic values. In the era of Large Language Models, translation can no longer be viewed merely as the mechanical conversion of linguistic codes. Instead, it has evolved into a complex socio-technical practice that requires critical vigilance, cultural awareness, and robust ethical oversight. While the shift toward human-AI collaboration is technologically irreversible, the

ethical baselines of the translation profession must be firmly upheld. As this paper argues, resolving these challenges requires moving beyond passive technological determinism. By implementing explainable AI architectures, establishing detailed legal and fair data frameworks, reaffirming the "Human-in-Command" principle, and fostering a multi-stakeholder governance ecosystem, it is possible to reconstruct a resilient ethical framework. This system will effectively balance computational efficiency with social accountability, integrating algorithmic intelligence with human judgment.

Future research and professional practice must transcend the binary narrative of human versus machine. Rather, stakeholders must continuously clarify the boundaries of accountability, counter digital cultural hegemony, and preserve human agency within the translation process. Ultimately, the goal of this ethical reconstruction is to ensure that artificial intelligence functions not as an opaque instrument of cultural flattening but as a transparent, equitable, and empowering bridge for genuine cross-cultural understanding.

References

Kenny, D. (2020). Technology and translator training. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 498-515). London: Routledge.

Lu, Y. (2024). Construction of translation technology ethics in the era of artificial intelligence. *Chinese Translators Journal, 45*(1), 117-125.

Moorkens, J. (2022). Ethics and machine translation. In *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 121-140). Berlin: Language Science Press.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... Lowe, R. (2022). Training language models to follow instructions with human feedback. In *Advances in neural information processing systems* (Vol. 35, pp. 27730-27744). New York: Curran Associates, Inc.

Wang, B. (2025). A survey-based analysis about consequential risk of errors and ethical complexities in the use of AI-powered machine translation in high-stakes situations. *Digital Studies in Language and Literature, 2*(2), 179-201.

Wang, H., & Liu, S. (2026). Deconstruction and reconstruction of translation ethics in the era of generative artificial intelligence. *Foreign Language Research, 49*(1), 15-24.

Wang, J., & Zhang, Z. (2026). Translators' prompting literacy in the era of human-machine collaborative translation: Framework construction and improvement paths. *Foreign Languages in China, 23*(3), 89-95.

Wang, J., Xiao, W., & Cui, Q. (2023). Technology-driven translation models in the era of artificial intelligence: Evolution, motivations, and implications. *Shanghai Journal of Translators, 38*(4), 14-19.

Wang, Y., Zhong, W., Li, L., Mi, F., Zeng, X., Huang, W., ... Liu, Q. (2023). Aligning large language models with human: A survey. arXiv:2307.12966.

Wu, M., & Chen, H. (2023). Ethical issues of machine translation in the artificial intelligence era. *Foreign Language Research, 46*(6), 13-18.