

Large Language Model-Based Translation Agents: A Review and Future Perspectives

YiHeng Qi

School of Foreign Studies, Yangtze University, Jingzhou 434000, China

Email: 924243915@qq.com

How to cite this paper: Qi, Y. H. (2026). Large Language Model-Based Translation Agents: A Review and Future Perspectives. *Literature, Language and Cultural Studies*, 4(3), 80 - 87. ISSN Print: 3079 - 5095; ISSN Online: 3079 - 5109.

<https://doi.org/10.63313/LLCS.9146>

Published: 2026-03-24

Copyright © 2026 by author(s) and

Erytis Publishing Limited.

This work is licensed under the Creative

Commons Attribution International

License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

Large language models (LLMs) have enabled general-purpose agents to expand into vertical domains such as translation. Accordingly, translation tools are evolving from passive programs into translation agents capable of autonomous planning, memory management, and external resource use. This paper reviews recent studies on translation agents from the perspective of Translation Studies rather than from a purely algorithmic viewpoint. It traces their technological evolution and theoretical mapping, with particular attention to single prompting with retrieval-augmented generation, multi-agent collaboration, and multi-level memory mechanisms. These developments partly simulate and reconstruct the cognitive strategies and collaborative networks of human translators. At the same time, the integration of translation agents into complex workflows reveals persistent problems, including limitations in quality assessment, blurred ethical accountability, and fluctuations in translators' cognitive load. The paper argues that future research should further examine evaluation mechanisms, ethical norms, and responsibility allocation within human-AI collaborative translation while also strengthening translator prompt literacy, digital resilience, and curriculum reform in translation education. The significance of translation agents lies not in replacing human translators, but in reshaping the translation ecosystem and its collaborative models.

Keywords

Large language models; translation agents; human-AI collaborative translation

1. Technological Evolution of Translation Agents

In the field of artificial intelligence, an agent is generally defined as an entity capable of perceiving its environment, making autonomous decisions, and executing actions to achieve specific goals (Xi et al., 2023). When this concept is introduced into Translation Studies, technological tools are no longer passive "source-in, target-out" programs. A translation agent (TA) is therefore defined as a virtual worker equipped with autonomous planning, memory management, and external resource invocation capabilities (Zhao and Li, 2024). Building on this definition, the

technological progression of TAs reflects a process through which machines increasingly approximate professional human translation workflows. This evolution has advanced from early single-model optimization to more complex forms of multi-agent collaboration and long-term memory management.

1.1. Technological Evolution of Translation Agents

In the early stages of development, researchers primarily sought to improve translation quality by optimizing instructions and integrating external knowledge. Zhao and Li (2024) show that well-designed instruction templates can guide a TA to establish an internal "translation-reflection-improvement" self-loop, enabling it to overcome the limitations of literal translation more actively. To address the persistent challenge of terminology inconsistency in specialized texts, TAs can also query external databases through retrieval-augmented generation (RAG) (Lewis et al., 2020). For example, when processing legal texts, a TA may retrieve authoritative domain glossaries or existing translation memories before generating the target text, thereby substantially reducing the hallucination of specialized vocabulary.

1.2. Multi-Agent Collaborative Mechanisms

When processing ultra-long literary texts or highly complex political discourse, a single agent may suffer from information overload. This limitation gave rise to multi-agent collaborative mechanisms. Essentially, this approach assigns specific roles to different agents in order to simulate the division of labor in a human translation agency. The TransAgents framework exemplifies this by assigning virtual roles such as CEO, editor, translator, and localization expert. For extensive literary translations, these roles sequentially handle initial translation, localization-oriented rewriting, and final review (Wu et al., 2024). In the translation of political discourse, the MAGIC-PTF framework introduces agents across four dimensions - style, translator, reviewer, and reader. By simulating the reading psychology of foreign audiences to adapt the target text, this framework significantly improves cross-cultural communicative accuracy (Liu, 2024).

Table 1. Comparison of collaborative mechanisms in typical multi-agent translation frameworks.

Framework	Application Domain	Agent Roles	Simulate Workflow Features
TransAgents	Ultra-long literary translation	CEO, editor, translator, localization expert, proofreader	Mirrors agency-level workflow; uses two-stage preparation and pipeline quality control.
MAGIC-PTF	Political discourse translation	Style, translator, reviewer, reader	Emphasizes cross-cultural acceptability; uses a reader agent for blind evaluation.
TACTIC	General/high-context texts	Drafter, polisher, evaluator, scorer, context, researcher	Models cognitive strategies; adds parallel trial translations and internal evaluation loops.

1.3. Multi-Level Memory Mechanisms

Maintaining consistency in terminology and pronominal reference during the translation of lengthy documents is a persistent technical challenge. The DelTA agent (Wang et al., 2024) addresses this barrier by incorporating a multi-level memory mechanism. Equipped with an internal proper noun log and both short-term and long-term memory modules, the agent extracts and retains key entities and macro-contexts translated earlier in real time. This mechanism simulates a human translator's note-taking process, allowing the machine to maintain strong textual coherence while translating long documents sentence by sentence.

2. Theoretical Mapping and Collaborative Reconstruction of Translation Agents

From the perspective of Translation Studies, quality bottlenecks in machine translation often stem from the lack of human-like cognitive reasoning. Current frontier agent frameworks attempt to translate abstract translation theories into concrete computational collaboration protocols.

2.1. Modular Reproduction of Cognitive Strategies

Cognitive Translation Studies holds that translation is a psychological process involving strategy selection, meaning construction, and deep context processing. The TACTIC framework (Li et al., 2025) offers a systematic modular simulation of this human cognitive behavior. Human translators often test different expressive strategies before drafting; in TACTIC, the drafter agent mirrors this process by generating three parallel initial drafts - literal translation, free translation, and creative adaptation. The polisher agent and evaluator agent then function as a self-monitoring mechanism, deconstructing and refining the initial drafts. Crucially, the system also includes a context agent to infer the psychology of the potential audience and a researcher agent to simulate consultation of parallel corpora, thereby endowing the machine with rare context awareness.

2.2. Sociological Reconstruction of Human-AI Collaborative Networks

Agent technology fundamentally reconstructs the sociological collaborative relationships of translation. From the perspective of Actor-Network Theory (ANT), technological devices can be understood as non-human actors in social practice (Iskanderov and Pautov, 2020). Traditional machine translation systems lacked autonomy and therefore functioned merely as passive tools restricted to one-way output. By contrast, the deep contextual understanding, interactivity, and intelligent features of generative AI enable agents to process complex instructions actively and provide multidimensional reference support.

Liu and Zhang (2025) argue that this evolution is driving a reconstruction of

translation production relations, pushing the industry from the traditional machine translation post-editing model toward artificial intelligence post-editing and human-AI collaborative translation paradigms (Wang and Liu, 2025). In this newly configured network, the agent functions as a co-worker, while the human translator shifts from a terminal error corrector to a global planner and quality gatekeeper who directs the process through prompts.

3. Multidimensional Challenges and Opportunities in Human-AI Collaborative Translation

Although multi-agent systems demonstrate considerable potential in laboratory settings, their integration into complex translation workflows introduces unavoidable challenges, including blind spots in quality assessment, ambiguous ethical responsibilities, and heightened cognitive demands on human translators.

3.1. Dilemmas in Quality Assessment and Multi-Agent Evaluation Approaches

Traditional automated metrics often fail to assess high-quality translations accurately when the target text involves extensive free translation or localization. To address this issue, researchers have attempted to employ large language models directly as evaluators, but single-model judgments often suffer from ambiguous criteria and evaluative bias (Feng et al., 2025). Confronted with this gap, scholars have begun to explore more diverse evaluation frameworks. For instance, the CHORUS framework (Wang et al., 2026) and the M-MAD mechanism (Feng et al., 2025) deploy agents representing distinct perspectives, such as fluency, accuracy, and terminology, to cross-examine one another and synthesize feedback in pursuit of more reliable assessments. Nevertheless, these efforts remain preliminary. When dealing with deep cross-cultural metaphors, existing evaluation agents may still omit substantive cultural imagery while pursuing absolute target-language fluency.

3.2. Ethical Risks and Accountability Challenges Arising from Increased Autonomy

The application of generative AI in translation inherently involves ethical controversies related to data privacy and algorithmic bias (J. Chen, 2025). As workflows move toward the agent paradigm, the growing autonomy of these systems further amplifies such risks. Santoni de Sio and Mecacci (2021) argue that AI systems can create responsibility gaps and blur accountability boundaries among stakeholders, confining human participants to isolated stages and causing them to overlook the broader consequences of system actions. Within translation workflows, when multiple agents operate and interact independently without real-time human supervision, politically misleading or culturally offensive output can make it difficult to determine whether responsibility lies with the algorithm developer, the corpus

provider, or the end user. Furthermore, empirical work in high-pressure contexts such as geopolitical conflict zones shows that current AI translation systems still retain built-in biases when processing highly sensitive culture-loaded terms (Alfaify, 2025). The more agents pursue automated decision-making, the more likely they are to overlook these subtle yet critical cultural frictions, thus underscoring the indispensability of human review mechanisms.

3.3. Cognitive Load Fluctuations and Human-Computer Interaction Friction

The shift in human-computer interaction paradigms has not reduced translators' workload in a linear fashion, as once anticipated. Eye-tracking research shows that, during human-AI collaborative translation of complex texts, translators' cognitive expenditure fluctuates across stages of the task (L. Chen, 2025). Because agent-generated translations often display exceptionally high surface fluency, any concealed errors in their internal reasoning may force translators to oscillate repeatedly between the machine's logical black box and their own cognitive networks. Tracing errors in agent output can therefore demand more intensive cognitive resources than pure human translation.

4. Future Perspectives on the Translation Ecosystem: Defending Subjectivity and Transforming Education

The structural transformation of translation production compels the field to reconsider the balance between human and machine. Future development must focus not only on continued technological refinement, but also on reinforcing human subjectivity and proactively reshaping translation education.

4.1. The Irreplaceable Subjective Value of Translators and Digital Resilience

Examining the limitations of translation agents ultimately reaffirms the core value of human translators. Translation in literature and the humanities is fundamentally a form of cross-cultural artistic reconstruction, shaped by the translator's aesthetic tendency, emotional experience, and cultural awareness (Guo and Moindjie, 2022). Existing research also indicates that, in literary translation, professional human translations still generally outperform large language model outputs, which often display a stronger tendency toward literalism and lower expressive diversity (Zhang et al., 2025).

As translation technology continues to evolve, prolonged technological dependence may also induce automation bias in translators (Parasuraman and Manzey, 2010), quietly eroding their linguistic sensitivity. For this reason, scholars increasingly emphasize the need for digital resilience. Translators must learn to use agent tools effectively while preserving critical distance from machine output. As the final

responsible subjects in the workflow, they remain indispensable in handling the most culturally complex passages and in safeguarding the ethical baseline of automated translation practice.

4.2. Constructing Translator Prompt Literacy for Agent Interaction

As workflows shift from traditional passive correction to deeper collaboration with advanced AI systems, conventional bilingual conversion skills are no longer the sole core competency. In response to the disruption caused by generative technology, Wang (2025) proposes the theoretical framework of translator prompt literacy. Although rooted in the broader use of large language models, this framework directly aligns with the capacity reconstruction required for operating translation agents. In an agent-collaborative environment, the competence of future translators must develop from static information retrieval to the ability to drive intelligent systems to acquire and reorganize knowledge through precise instructions. Professional translators therefore need to internalize a search-think-act methodology: identifying translation bottlenecks, converting complex stylistic and pragmatic demands into constraints that agents can interpret accurately, and continuously optimizing output through multi-turn negotiation and feedback.

4.3. Restructuring Translation Education Curricula for the Intelligent Era

The continued integration of agent technology is shifting translation education from conventional training in tool operation to the cultivation of higher-order human-computer collaboration skills. Wang and Liu (2025) argue that translation education in the era of artificial intelligence post-editing must redefine translation competence by constructing a hybrid training model of language, professional knowledge, and technology, while incorporating large language model principles, prompt engineering, translation agents, and intelligent translation project management into the curriculum. At the same time, Wang's (2025) framework of translator prompt literacy suggests that future translators' core competitiveness lies not simply in bilingual transfer, but also in the ability to formulate effective prompts, optimize instructions continuously, and evaluate output prudently. International studies likewise show that machine translation literacy and data literacy are increasingly entering translation pedagogy, helping learners understand technological boundaries, strengthen quality assessment, and develop responsible technology use (Bowker, 2021; Hackenbuchner and Kruger, 2023). Therefore, future curriculum reform should move beyond the simple addition of technological tools and instead systematically cultivate students' prompt design capabilities, quality assessment skills, and ethical awareness of technology.

5. Conclusion

Large language model-based translation agents have opened a new frontier at the intersection of computational linguistics and Translation Studies. From single-prompt systems to multi-agent collaboration and memory-enhanced frameworks, recent research has expanded the scope of machine translation while partially modeling the cognitive and collaborative processes of human translators. At the same time, the integration of translation agents into real workflows reveals persistent challenges in evaluation, ethical accountability, and human cognitive load. These limitations suggest that translation agents should not be understood as replacements for human translators, but as technologies that are reshaping the translation ecosystem. Future research should therefore pay closer attention to evaluation mechanisms, ethical governance, and responsibility distribution within human-AI collaboration, while also strengthening translator prompt literacy, digital resilience, and curriculum reform. Only by balancing technological capability with human judgment can translation agents contribute to the more reliable and sustainable development of translation practice.

References

- [1] Chen, J. (2025) A Comparative Analysis of Generative AI Translation Ethics at Home and Abroad. *Journal of Kunming University of Science and Technology (Social Sciences Edition)*, 25(5).
- [2] Liu, M. (2024) Research on Russian Translation of Chinese Political Discourse Empowered by Large Language Models: Creation and Application of a Multi-Agent Interactive Translation Framework. *Technology Enhanced Foreign Language Education*, No. 6.
- [3] Liu, S. and Zhang, Y. (2025) From MTPE to HACT: Research on Translation Workflow Innovation Driven by Large Language Models. *Foreign Language Education Research*, 13(1), 18-26.
- [4] Wang, H. and Liu, S. (2025) From MTPE to AIPE: The Evolution of Translation Models in the Generative AI Era and Its Implications for Translation Education. *Shandong Foreign Language Teaching*, 46(3), 111-121.
- [5] Wang, S. (2025) From "Search Quotient" to "Question Quotient": Constructing a Conceptual Framework for Translator Prompt Literacy in the Era of Generative AI. *Foreign Language Education Research*, 13(1), 1-8.
- [6] Zhao, J. and Li, X. (2024) Research on the Construction and Application of Translation Agents Driven by Large Language Models. *Technology Enhanced Foreign Language Education*, No. 5, 22-28.
- [7] Alfaify, A. (2025) Replacing the Irreplaceable: A Case Study on the Limitations of MT and AI Translation during the 2023 Gaza-Israel Conflict. In: Bouillon, P., Gerlach, J., Girletti, S., et al., Eds., *Proceedings of Machine Translation Summit XX: Volume 2*, European Association for Machine Translation, Geneva, 8-17.
- [8] Bowker, L. (2021) Machine Translation Literacy Instruction for Non-Translators: A Comparison of Five Delivery Formats. In *Proceedings of the Translation and Interpreting Technology Online Conference*, Held Online, 25-36.
- [9] Chen, L. (2025) Examining Cognitive Load in Human-Machine Collaborative Translation: Insights from Eye-Tracking Experiments of Chinese-English Translation. *Frontiers in Psychology*, 16, Article 1570929.
- [10] Feng, Z., Su, J., Zheng, J., Ren, J., Zhang, Y., Wu, J., Wang, H. and Liu, Z. (2025) M-MAD: Multidimensional Multi-Agent Debate for Advanced Machine Translation Evaluation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational*

Linguistics (Volume 1: Long Papers), Vienna, 7084-7107.

- [11] Guo, X. and Moindjie, M.A. (2022) A Perspective on the Translator's Subjectivity and Its Constraints in Literary Translation: A Critical Analysis. *Journal of Positive School Psychology*, 6(4), 3118-3123.
- [12] Hackenbuchner, J. and Kruger, R. (2023) DataLitMT - Teaching Data Literacy in the Context of Machine Translation Literacy. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, Tampere, 285-293.
- [13] Iskanderov, Y. and Pautov, M. (2020) Agents and Multi-Agent Systems as Actor-Networks. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, Valletta, 179-184.
- [14] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W.-T., Rocktaschel, T., Riedel, S. and Kiela, D. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.
- [15] Li, W., Chen, J., Li, B., et al. (2025) TACTIC: Translation Agents with Cognitive-Theoretic Interactive Collaboration. *arXiv preprint arXiv:2506.08403*.
- [16] Parasuraman, R. and Manzey, D.H. (2010) Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors*, 52(3), 381-410.
- [17] Santoni de Sio, F. and Mecacci, G. (2021) Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them. *Philosophy & Technology*, 34, 1057-1084.
- [18] Wang, G.X., Hu, J. and Qian, J. (2026) Who Has the Final Word? Designing Multi-Agent Collaborative Framework for Professional Translators. *arXiv preprint arXiv:2602.19016*.
- [19] Wang, Y., Zeng, J., Liu, X., Wong, D.F., Meng, F., Zhou, J. and Zhang, M. (2024) DelTA: An Online Document-Level Translation Agent Based on Multi-Level Memory. *arXiv preprint arXiv:2410.08143*.
- [20] Wu, M., Yuan, Y., Haffari, G. and Wang, L. (2024) (Perhaps) Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts. *arXiv preprint arXiv:2405.11804*.
- [21] Xi, Z., Chen, W., Guo, X., et al. (2023) The Rise and Potential of Large Language Model Based Agents: A Survey. *arXiv preprint arXiv:2309.07864*.
- [22] Zhang, R., Zhao, W. and Eger, S. (2025) How Good Are LLMs for Literary Translation, Really? Literary Translation Evaluation with Humans and LLMs. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Albuquerque, 10961-10988.