

Breaking Boundaries and Reconstructing Value: Innovative Paths and Value Breakthroughs of Multimodal Large Models and Generative AI Empowering the Translation Industry

Chenyang Liu*

Wuhan University of Engineering, Wuhan 430000, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: The iterative advancement of multimodal large models and generative artificial intelligence has transcended the constraints of conventional text-to-text translation, propelling the translation sector into a new era of global translation characterized by multimodal input and multimodal output. Moving beyond the conventional research lens of technological tool application, this paper centers on crossmodal semantic alignment and cultural transcoding efficiency as core entry points. The paper systematically examines the disruptive restructuring of translation practice, communication logic, and industrial ecology brought by multimodal large models and generative artificial intelligence, probes deeply into their innovative value in addressing the last-mile challenge of cultural communication, and underscores their enabling role in the high-quality development of the translation field. This study focuses on two innovative application dimensions of multimodal large models and generative artificial intelligence. First, the mechanism of crossmodal semantic symbiosis underlying multimodal large models is analyzed, detailing how synchronous translation and emotional alignment of multidimensional information—including text, images, audio, video, and speech—are achieved, thereby mitigating information loss inherent in single-modal translation. Using short-video translation and virtual digital human translation and broadcasting as illustrative cases, it validates the immersive communication strengths of multimodal large models in cultural going global initiatives. Second, the dynamic adaptation and ethical governance of generative artificial intelligence are explored, departing from the traditional perception of AI-assisted translation. The study investigates innovative applications of generative artificial intelligence in personalized translation, real-time cross-cultural transcoding, and low-resource language adaptation, while confronting ethical risks arising from technological alienation in translation. A tripartite governance framework of technological empowerment, humanistic calibration, and normative constraint is proposed accordingly. This paper argues that multimodal large models and generative artificial intelligence are not merely auxiliary tools for enhancing translation efficiency, but value reconstructors that drive the transformation of the translation industry and facilitate cross-cultural communication. Through precise crossmodal semantic translation, they enable the intact transmission of cultural connotations; via generative innovation, they adapt to the reception habits of diverse overseas audiences; and through technological empowerment, they push the translation sector to upgrade profoundly from language conversion

toward cultural interpretation and value transmission. This research offers a fresh perspective for the construction of translation disciplines, innovation in the language service industry, and translation practice for cultural going global, supporting the highquality development of translation and the deep advancement of crosscultural exchanges and mutual learning in the digital intelligence era.

Keywords: Multimodality; Artificial intelligence; Translation technology; Generative AI; Multimodal large models; Crosscultural communication

Online publication: May 25, 2026

1. Introduction

In the digital intelligence era, the technological breakthrough of multimodal large models and generative artificial intelligence has penetrated the translation industry with unprecedented depth and breadth. Traditional translation research and practice have long been dominated by texttotext language conversion, which struggles to meet the rising demand for multimodal content—such as images, audio, and video—in contemporary crosscultural communication. A core question of this paper is whether technological iteration constitutes a mere upgrade of translation tools or a paradigm shift reshaping the nature, process, function, and value of translation.

A multimodal large model refers to a largescale neural network model capable of processing and integrating heterogeneous information across text, images, audio, video, and other modalities, with its technical essence lying in crossmodal joint representation and understanding. Distinct from traditional discriminative artificial intelligence, generative artificial intelligence acts as an emerging content creator, marked by creativity, interactivity, and dynamic adaptability. In the translation context, boundary breaking denotes technologydriven dismantling of content boundaries, media boundaries, and subject boundaries, whereas reconstruction signifies the reshaping of practice models, communication logics, industrial ecologies, and value systems by emerging technologies ^[1].

Departing from the microperspective of instrumental application, this paper takes crossmodal semantic alignment and cultural transcoding efficiency as core entry points to investigate the systemic restructuring of translation enabled by technology. It concentrates on two innovative dimensions: crossmodal semantic symbiosis and dynamic adaptation with ethical governance, aiming to reveal the deep mechanisms and innovative values of technological empowerment in translation.

2. Technological foundations: From singlemodal conversion to crossmodal semantic alignment

2.1. Limitations of the traditional translation paradigm

The texttotext paradigm has long prevailed in translation theory and practice. Its inherent deficiency lies in severe information loss caused by singlemodal processing: cultural imagery, visual metaphors, emotional ambience, and other nonlinguistic elements are largely lost during conversion, hindering the full transmission of sourcelanguage cultural connotations. Traditional translation also faces bottlenecks in bridging the last mile of cultural communication. The lack of situated and immersive delivery often leaves translated products at the readable level rather than perceptible and resonant, failing to resonate with the aesthetic expectations and emotional needs of mass overseas audiences.

This language-symbol-centered paradigm reduces translation to a technical exercise of code switching, overlooking the synergistic effects of multimodal components in cultural communication. As translation objects expand from static texts to emerging media such as short videos, games, and virtual reality, the limitations of the traditional paradigm become increasingly acute, rendering it unfit for the complexity and diversity of digital-age cross-cultural communication.

2.2. Underlying logic of multimodal large models: The mechanism of crossmodal semantic symbiosis

Multimodal large models establish a shared semantic space to support joint representation and understanding of diverse modal data, including text, images, audio, and video. Their technical core is to break modal barriers and enable precise alignment and integrated fusion of heterogeneous information within a unified representational framework. Alignment technology ensures that elements referring to the same semantic unit across modalities correspond accurately; fusion technology allows aligned information to complement and reinforce one another, forming a more comprehensive semantic representation.

On this basis, the concept of crossmodal semantic symbiosis reveals a deeper operational principle: modal inputs in translation are not simply superimposed or converted, but mutually generative, coemerging into a richer and more cohesive holistic meaning. For instance, in translating a commercial with visual metaphors, the model can simultaneously comprehend the compositional intent of visuals and the rhetorical features of voice-overs, identifying equivalent expressive combinations in the target culture to generate a synergistic effect greater than the sum of parts.

2.3. Role evolution of generative artificial intelligence: From translation assistant to content coconstructor

The traditional notion of AI-assisted translation frames artificial intelligence as an efficiency-enhancing tool. However, the evolution of generative artificial intelligence has endowed it with autonomous functions, including content generation, stylistic rewriting, and contextual adaptation, marking a role shift from assistant to coconstructor [2].

As a virtual translator, generative AI can independently complete translation tasks in specific scenarios, demonstrating unique advantages in high-real-time, high-volume contexts. As a creative collaborator, it engages in deep synergy with human translators: inspiring multiple expressive alternatives in the creative conception phase, undertaking laborious draft generation and iterative refinement in production, and rapidly adjusting stylistic features based on user feedback during optimization. This human-machine coconstruction redefines translation workflows and division of labor, freeing human translators from repetitive tasks to focus on cultural creativity and aesthetic quality control.

3. Innovative path I: Global multimodal translation driven by multimodal large models

3.1. Immersive restructuring of short-video translation

Multimodal large models exhibit powerful integrated capabilities in short-video localization, enabling simultaneous subtitle translation, voice cloning, visual element replacement, and cultural adaptation of background music. Taking intangible cultural heritage short videos as an example, the model preserves the

original aesthetic style while converting Chinese narration into natural targetlanguage speech with cloned timbre, aligning virtual dubbing with the emotional tone of the original speaker. Onscreen text, such as street signs and product labels, can be automatically recognized and replaced with localized versions, and background music can be adjusted to match the aesthetic preferences of the target culture ^[3].

This fullmodal translation achieves textvisualaudioemotion synchronous rendering, overcoming the fragmented watchingreading experience caused by conventional subtitle translation. It enables overseas audiences to engage with content more naturally, substantially improving the acceptance and dissemination efficacy of cultural going global products.

3.2. Virtual digital human translation and broadcasting: From language conduit to cultural interpreter

The integration of virtual digital human technology and translation transcends the physical constraints of human translators, supporting multilingual, roundtheclock, and situated cultural communication. Translation moves from behind the scenes to the foreground, building emotional connections with audiences through personalized virtual personas.

Cultural adaptation in nonverbal modalities—including lip synchronization, facial expressions, body language, costumes, and makeup—visualizes cultural transcoding. Virtual avatars can dynamically adjust expressive ranges, gestural patterns, and overall temperament in line with the nonverbal communicative norms of the target culture. When addressing audiences from different cultural backgrounds, virtual translators can adjust visual styling to respect local aesthetic preferences while conveying core sourceculture meanings. Beyond the conventional role of a passive language conduit, the personaized cultural interpreter serves as a relatable and trustworthy intermediary for crosscultural communication.

3.3. Humanmachine collaborative fullmodal translation workflow

The future of translation points to a humanmachine collaborative fullmodal workflow, in which human translators oversee creativity, emotion, and ethics at a high level, while multimodal models handle largescale information processing and multimodal content generation, forming a highly complementary collaborative network.

Human translators are liberated from repetitive labor to concentrate on cultural creativity and aesthetic judgment—for example, evaluating the appropriateness of creative transformations of cultural imagery and verifying that the emotional tone of translations aligns with communicative goals. Multimodal models undertake scalable production: simultaneous multilingual version generation, adaptive multimodal modification, and rapid response to user feedback. This synergy markedly improves the acceptability and transmissibility of translated products in target markets, allowing cultural content to retain core values while resonating more effectively with audiences, thus genuinely solving the lastmile challenge of cultural communication.

4. Innovative path II: Value creation and risk governance enabled by generative artificial intelligence

4.1. Dynamic adaptation: Personalized and situated generative translation

Generative AI can dynamically produce differentiated translation versions according to the cultural

background, age cohort, and aesthetic preferences of target audiences, realizing a fundamental shift from standardized products to personalized services. For Generation Z, translations can integrate internet slang and fastpaced expressions; for middleaged and elderly audiences, a more elegant and restrained style can be adopted. Such audienceprofilingbased dynamic adaptation allows translations to reach distinct groups precisely ^[4].

In lowresource languages, endangered languages, and specialized vertical domains, generative AI excels at rapidly building translation capacity through lightweight finetuning on limited corpora, effectively filling market gaps. For linguistically marginalized communities, this opens new channels for their cultural expressions to enter global communication, positioning generative translation as a vital force for safeguarding linguistic diversity and promoting equitable crosscultural dialogue.

4.2. Enhanced cultural transcoding efficiency: From literal accuracy to vivid resonance

Traditional translation pursues faithfulness, expressiveness, and elegance primarily at the linguistic level. In contrast, generative AI prioritizes cultural equivalence and emotional resonance, shifting evaluative criteria from translation accuracy to communicative vitality.

With robust knowledge association and contextual comprehension, generative AI proactively conducts creative transformation and interpretation of cultural imagery, rendering culturespecific items that resist literal translation into functionally equivalent elements in the target culture. When handling texts containing culturespecific allusions, for example, the model identifies their communicative functions and replaces them with culturally analogous resources in the target context rather than merely adding explanatory footnotes. Such dynamic transcoding preserves core information while infusing translations with vitality and dissemination power, lifting crosscultural communication from information transmission to emotional connection.

4.3. Technological alienation and ethical governance: A tripartite governance framework

The deployment of generative artificial intelligence in translation entails multiple ethical risks. Foremost are cultural biases and stereotypes: biases embedded in training data may be amplified by models, reinforcing misunderstandings or discriminatory representations in outputs. Cultural value dilution and misinterpretation also pose risks: excessive adaptation may oversimplify or distort deep cultural connotations, turning translation into a source of cultural discount rather than a bridge for dialogue. Furthermore, authorship disputes and threats to translator subjectivity persist, as copyright boundaries of AIgenerated content remain unclear and the core value of human translators risks marginalization ^[5].

To address these risks, this paper proposes a tripartite governance framework of technological empowerment–humanistic calibration–normative constraint: Technologically: develop culturally sensitive algorithms and establish a multicultural alignment evaluation system; Humanistically: strengthen the central role of translators as cultural gatekeepers and ethical arbiters; Normatively: advance industry ethical guidelines and legal regulations to clarify copyright ownership and liability boundaries of AIgenerated content.

5. Value reconstruction: Deep logic of translation industry transformation

5.1. Reconceptualizing translation: From language conversion to cultural interpretation and value transmission

Driven by technology, the core mission of translation has expanded beyond linguistic matching to fulltext

cultural interpretation encompassing visual signs, auditory cues, and interactive modalities, fundamentally redefining the meaning of translation. The ultimate goal of translation has evolved from information accessibility to value identification; translated products are no longer mere information carriers but media for conveying cultural values, aesthetic ideals, and emotional identity, undertaking the strategic role of shaping national image and facilitating civilizational mutual learning.

Within this renewed framework, translators work with cultural products rather than static texts; translation evolves from code switching to cultural interpretation; and outputs advance from readable texts to perceptible experiences. Translation thus moves from the periphery of language services to the core of cultural communication, serving as a critical bridge connecting civilizations and fostering mutual understanding.

5.2. Reconstructing translation subjects and objects: Humanmachine symbiosis and modal fusion

The translation subject has expanded from individual human translators to a hybrid subject of human translator + AI agent, forming a collaborative and cocreative relationship that distributes translation subjectivity. Human translators contribute cultural judgment, aesthetic discernment, and ethical governance; AI agents provide scalable processing, multimodal generation, and realtime responsiveness. Their complementary strengths continuously expand the boundaries of translation production ^[6].

The translation object has evolved from static texts to dynamic, interactive, and generative multimodal cultural products. Translation no longer merely converts predefined content but participates in the entire lifecycle of cultural production. This dual reconstruction of subject and object calls for a redefinition of translator competence, translation education, and translation ethics. Translators in the new era must master integrated capabilities, including multimodal comprehension, AI tool literacy, and crosscultural creativity.

5.3. Restructuring translation ecology: Value chain transformation of the language service industry

Technology has disrupted the linear and closed industrial chain of traditional language services, spawning a highly integrated ecosystem of technology platforms–content creators–translators–distribution channels, with blurred industrial boundaries. In the conventional model, language service providers mainly offer single translation services and occupy the lowvalue end of the chain. In the emerging ecosystem, newgeneration language service enterprises that integrate multimodal translation, cultural creative design, and technology platform operation occupy highvalue positions.

The core competitiveness of the future language service industry will shift from translation capacity to cultural creativity and technological integration capability. Pure speed and cost advantages will be superseded by culturalunderstandingcentered creative capacity. This shift unlocks new growth trajectories: industry value extends beyond translation services to endtoend solutions for cultural going global, including content innovation, localized adaptation, and crosscultural marketing.

6. Conclusion

Multimodal large models and generative artificial intelligence are not simple translation tools. Through crossmodal semantic alignment and dynamic cultural transcoding, they drive a fundamental transformation

of translation: from text conversion to global multimodal translation, and from information channel to value reconstructor. This research provides a new theoretical perspective for translation discipline innovation in the digital intelligence era, expands the boundaries of translation studies, and places multimodality, humanmachine collaboration, and cultural transcoding efficiency at the center of inquiry^[7].

Practically, this study offers guidance for technological adoption and strategic transformation in the language service industry, delivers actionable approaches for cultural going global institutions to leverage new technologies in narrating Chinese stories, and supports the reform of translation talent cultivation systems. This research is primarily theoretical and casebased; empirical investigations into implementation costs, largescale application feasibility, and longterm social impacts remain insufficient.

Looking ahead, as general artificial intelligence evolves, the translation industry may undergo even deeper changes. How to steer technology to serve equitable dialogue and mutual learning among civilizations will continue to challenge translation scholars and the broader academic community. Building an interdisciplinary research community to collectively explore a humanmachine collaborative, culturally oriented future of translation is both timely and imperative.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Ge MJ, 2024, Underlying Logic, Application Practice and Realistic Concerns of Multimodal Content Production Empowering the Audio Visual Industry in the General Artificial Intelligence Era. *New Media Research*, 10(24): 60–64.
- [2] Liu ZS, Zhou YL, 2024, Research on Archive Digital Narrative Model Empowered by AIGC. *Zhejiang Archives*, 2024(10): 18–22.
- [3] Chen AT, Lu J, Zhang XQ, 2024, Research on the Influence Mechanism of Generative Artificial Intelligence on Doctor Patient Shared Decision Making. *Chinese Medical Ethics*, 37(9): 1087–1092.
- [4] Wang YW, Shen T, Zhang SY, et al., 2024, Progress in Collaborative Evolution Technology of Large and Small Models at the End and Cloud. *Journal of Image and Graphics*, 29(6): 1510–1534.
- [5] Shao SM, 2024, Artificial Intelligence Platform Project Based on the Fusion of Decision Making and Generative Modes. Agricultural Development Bank of China, Beijing.
- [6] Wang Y, Song YX, Wang YF, et al., 2024, Ethics and Governance of Artificial Intelligence in the Health Field: A Guide to Multimodal Large Models. *Chinese Medical Ethics*, 37(9): 1001–1022.
- [7] Zhang XH, Ma Y, 2024, Generative Artificial Intelligence Technology Empowering the Emergence of New Quality Productive Forces: Value Implication, Operating Mechanism and Practical Path. *E Government*, 2024(4): 17–25.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.