

AI-Enabled Content Optimization Mechanisms and Ethical Norms for the Translation and Dissemination of Chinese Culture in Japan

Yan Xie, Jian Chen

School of Foreign Languages, Hebei University of Technology, Tianjin 300130, China

Corresponding author: Yan Xie (e-mail: xylunwen2026@163.com).

ABSTRACT The translation and dissemination of Chinese culture in Japan face persistent challenges involving cultural imagery attenuation, weak contextual adaptation, and insufficient ethical review of AI-generated output. To address these problems, this study constructs a Chinese-Japanese cultural parallel corpus and knowledge graph, vectorizes culturally loaded expressions, and introduces a cognitive-anchor remapping strategy to reduce metaphorical shift in cross-cultural translation. On the basis of Japanese natural-language corpora, an audience cognitive preference model is trained to optimize register softening, honorific hierarchy, and the narrative structure of translated content. A multimodal style cooptimization procedure is further combined with a three-dimensional ethical review module, in which rule-based interception and manual verification form a closed control loop. Experimental results show improved cultural imagery targeting, audience reach, and control of sensitive expressions. The study frames cultural translation as an AI-assisted semantic transmission and communication optimization task, offering technical support for multilingual digital dissemination systems and multimodal content propagation.

INDEX TERMS artificial intelligence translation, semantic mapping, multimodal communication, content optimization, ethical review

I. INTRODUCTION

TRANSLATING Chinese culture into Japanese for identification outside the East Asian cultural community and enhancing China's cultural soft power are crucial. Currently, translation for external publication in Japan encounters three structural problems that have existed for a long time. First, the problem of semantic consistency in cultural imagery is particularly prominent. The semantic connotations of cultural imagery expressed in classical poetry, idioms, philosophical terms, etc., are often attenuated and even lost during translation into Japanese due to the fact that the two languages have a surface different but deep-similar meaning. For instance, idioms such as "surrounded by enemies on all sides" and "learning to walk in Handan" have historical and geographical coordinates and will lose these attributes after being translated into idiomatic Japanese by target language readers directly translated by simple machine translation. The spiritual core of Chinese culture is lost in the process of cross-linguistic transmission. Second, there is a serious problem in audience awareness in terms of context adaptation. Official texts for external publication are habitually written in macro-narrative and political commentary styles, which form a sharp contrast to target readers'

expectation of being interpreted in more relaxed, subtle and concrete ways in Japanese. Poor performance in target language level formality and fixed narrative style of translation has led to the poor reception of external propaganda content translated into Japanese for a very long time, resulting in greatly reduced effectiveness. Thirdly, there is a lack of ethical review. Translation for external publication covers highly sensitive applications such as historical narratives, political terms and socially taboo language, but current process of translating Chinese into Japanese by AI generally lacks an effective ethical filtering link for target language. The target language translation is not controlled, which may trigger the risk of touching on Japanese social and cultural taboos, and even become a security risk for external propaganda [1-3].

Existing research has explored several directions related to AI-enabled cultural external propaganda translation. For example, Wang et al., starting from the relationship between AI-enhanced translation and crosscultural communication, pointed out that in the context of rapid technological evolution, the humanities play an irreplaceable fundamental role. [4] Xiao discussed the mediating function of translation in international knowledge dissemination from the

perspective of knowledge translation studies, emphasizing that translation is not only a language conversion. [5] Hu, combining the translation of Chinese classics with the international dissemination of Chinese culture, revealed the close relationship between discourse reconstruction, cultural interpretation and communication context. [6] Yu et al. further proposed key operational paths to reduce "cultural discount" from the perspective of communication studies, explaining whether the content of external publicity can achieve effective reach. [7] However, the above studies mostly focus on the technical optimization or theoretical discussion of a single link, lacking a systematic integration of semantic fidelity, contextual adaptation and ethical review into a unified technical framework. Existing methods are difficult to respond to the practical needs of "accurate translation", "widespread dissemination" and "safeguarding" in the external publicity of Chinese culture translated into Japanese.

Regarding the semantic fidelity problem of cultural proprietary items, the progress of knowledge graph technology in cross-language knowledge alignment and concept vectorization representation provides a new technical path for the accurate mapping of culturally loaded words. For example, Liu et al.'s research shows that structured knowledge organization and semantic matching mechanisms can significantly improve the accuracy of cross-language text retrieval and concept correspondence, which provides a methodological basis for the alignment of Chinese and Japanese cultural concepts. [8] At the audience reception level, Nettelbeck analyzed the differences in subtitle comprehension between Japanese and English audiences and pointed out that the cultural cognition habits, understanding paths and role perception methods of the target audience directly affect the reception effect of the translation. [9] At the same time, the historical AI translation system must establish corresponding constraint mechanisms at the levels of knowledge generation, expression boundaries and value orientation. [10] However, current research has obvious shortcomings: knowledge graph methods are mostly limited to general domains and lack specialized construction for the alignment of Chinese and Japanese cultural concepts; audience preference modeling has not been deeply coupled with the foreign propaganda translation generation process; ethical review rules lack a fine-grained classification system for Japanese-translated foreign propaganda scenarios. Based on this, this paper organically integrates the above three methods and proposes a chain optimization mechanism covering semantic mapping, context adaptation and ethical review to break through the limitations of fragmentation and singularity in existing research.

The main innovation of this paper can be summarized in the following three folds: Firstly, we build a parallel corpus and knowledge graph of Chinese and Japanese culture for Japanese-translated foreign propaganda, and propose an implementation path of culturally-loaded words from vectorized representation to semantically accurate mapping.

Secondly, based on a cognitive preference model of Japanese audience, we design a contextualized dynamic adaptation mechanism, including softening of register, adjustment of honorifics and reconstruction of narrative framework to make a pragmatic leap from "readable" to "readable". Thirdly, we construct an ethical review module including three dimensions of political red line, historical representation and social taboo, and build a compliance filtering closed loop based on a rule engine and manual calibration to form an operable solution of ethical governance of AI-assisted translation for external communication.

II. DEEP SEMANTIC MAPPING AND KNOWLEDGE ENHANCEMENT MECHANISM OF CULTURALLY LOADED WORDS IN SINO-JAPANESE RELATIONS

The corpus consists of three components. For classical Chinese texts, the dataset includes the Japanese translations of *The Analects* by Kinami Noriko and Kawatani Osamu, Makita Takeru's translation of *The Book of Songs*, and a curated selection of Tang and Song dynasty poetry accompanied by authoritative Japanese annotations to ensure the knowledge graph's coverage of diverse rhythmic and metaphorical patterns. The original texts and their translations are aligned by chapter, and the boundaries of culture-specific items are annotated. For official external propaganda, news releases and feature reports from the Japanese edition of *People's Daily Online* and the Japanese service of *China Radio International* over the past five years are collected, yielding approximately 80,000 sentence pairs. For Japanese academic sources, fragments of scholarly literature dealing with classical Chinese concepts are selected from the digital collection of the National Diet Library to serve as supplementary semantic background knowledge. Based on the above corpus, a topdown ontology design method is adopted to construct a Chinese-Japanese cultural knowledge graph, where culture-loaded terms serve as nodes, and edge relations cover four types: source text reference, homologous metaphor, Japanese translation equivalent, and contextual variation. In the semantic encoding stage, a Transformer-based pre-trained language model is introduced to perform deep bidirectional encoding of target words, extracting dynamic semantic vectors that integrate contextual and discourse information. Combined with sense clustering information from the knowledge graph, the activated sense in the current context is determined via cosine similarity calculation. Taking the character *wu* as an example, the model identifies the ethical resonance between the Chinese concept of "stopping violence through force" and the self-restraint (*katsujinken*) inherent in traditional bushidō, thereby mapping the translation to a shared cultural logic rather than a simplified binary opposition.

For allusion expressions without common historical experience, the mechanism extracts their narrative structure, i.e., "type of predicament, subject identity, and emotional tone," from the knowledge graph and retrieves isomorphic native Japanese allusions from a Japanese cultural repository

as cognitive anchors. For example, the narrative structure of “besieged on all sides” is “a hero at the end of his road and isolated without support,” which is structurally similar to the Japanese native allusion “the knight in white armor.” The target allusion marker is carried forward in the translation, while a generic guiding modifier is added to trigger the reader’s association. To mitigate the problem of Old Chinese texts being susceptible to mis-segmentation and mis-judgment of imagery in general-purpose models, we adopt low-rank adaptation for parameter-efficient fine-tuning. A bypass structure with trainable low-rank decomposition matrices is inserted into the selfattention layers of the base model. The model is then incrementally trained on classical text corpora using only aligned training data, allowing it to learn the distinctive rhetorical patterns and imagery system of classical Chinese texts. Before sending the semantic mapping results to downstream modules, they must first clear the confidence check. The system computes the co-occurrence probability match between the semantic vector of the target word and its candidate Japanese translation equivalents in large-scale Japanese corpora. Those that do not meet the threshold are automatically forwarded to a human proofreading interface. The proofreader only needs to confirm the result or correct it if necessary; the relevant record is then written back to the edge weight parameters of the knowledge graph. This is a continuously optimized data loop in which automated processing exerts efficiency over cultural details while experts exert authority.

III. CONTEXTUALIZED DYNAMIC ADAPTATION MECHANISM FOR JAPANESE AUDIENCE COGNITIVE PREFERENCES

Audience adaptation is based on accurately grasping the reading habits of the target group. This paper collects approximately 300,000 pieces of Japanese natural text on China-related topics. While this corpus size is focused, the study employs data augmentation and dropout regularization to prevent overfitting; furthermore, it utilizes cross-domain validation to ensure that the model captures the high-dimensional nuances of Japanese honorifics despite the specialized nature of the dataset. After cleaning, it uses a latent Dirichlet assignment topic model and word frequency-inverse document frequency statistics to extract three typical preferences of Japanese audiences when facing foreign cultural information: a tendency to use speculative tones rather than assertive statements, with the use of speculative auxiliary verbs effectively reducing reading resistance; high sensitivity to the interpersonal distance between the author and themselves, requiring the selection of honorifics to construct an appropriate scale of respect and intimacy; and a strong influence on the “introduction, development, transition, and conclusion” structure in information organization, with straightforward political commentary easily perceived as lacking structure. Based on these characteristics, a cognitive preference vector model encompassing stylistic features, honorific levels, and structural patterns is constructed to

provide a quantitative basis for adaptation operations. At the stylistic level, addressing the issues of dense long, fixed sentences and frequent affirmative judgments in machine-translated texts, the system performs three types of transformations: converting definitive judgments into objective inferences to soften the didactic tone; breaking down lengthy modifiers into independent short sentences based on the “one sentence, one meaning” principle in Japanese writing; and replacing evaluative vocabulary with factual statements to restore a neutral style to the translation.

Regarding honorifics control, we classify the intergovernmental external propaganda text into three types according to release context, namely, intergovernmental document, public media and grassroots cultural exchange, and match honorific level for each type. We also define the beneficiary judgment rule, which specifies how to judge the beneficiary direction of a description when switching honorifics from beneficiary-based ones to beneficiary-forced ones. As for the text, we rearrange the translation logic into the “introduction, development, transition, conclusion” structure: introduction explains the background information to make preparation for description; development transfers parallel listing into progressive sequence; transition introduces limiting viewpoint or counter-argumentative view to enhance subjectivity; and conclusion naturally concludes and does not use call-for-attention ending. In addition, the introduction part of long texts also designs transitional subheadings for browsing. Simultaneous adjustment of style and honorifics for multimodal text in external publicity product, such as image caption, poster title, and video subtitle, will be introduced. Image caption uses automatic compression into noun phrase to get brief description of information; poster title adjusts number of syllables into reasonable range to fit the rhythm of Japanese poster title and meanwhile keep enough information; video subtitle uses compression into long, medium and short three levels of options according to the maximum time length of display, to be consistent with style and interpersonal distance in other media.

IV. INTELLIGENT REVIEW AND COMPLIANCE MANAGEMENT MECHANISM FOR ETHICAL RISKS IN EXTERNAL PUBLICITY CONTENT

The effectiveness of the review mechanism depends on a systematic analysis of risk types. Based on an analysis of relevant regulatory documents from the external publicity authorities and past translation review cases, this paper divides the ethical risks in Japanese-translated external publicity scenarios into three dimensions. The knowledge base of the review mechanism consists of a sensitive word vector library and a rule knowledge base. Political red line content uses a whitelist mechanism for mandatory verification; any deviation from the standard translation is directly blocked from output. Historical and social taboo content uses a vectorized sensitive word library containing semantic variations to capture circumvention methods such as synonym substitution and voice conversion. The rule knowledge base

stores contextual judgment logic in a "condition-action" pair format to handle boundary issues dependent on specific contexts. Initially written by experts, it is continuously expanded based on review feedback.

After the translation enters the review process, the rule engine handles the first interception. Rule engine executes mandatory whitelist verification and sensitive word similarity scanning in parallel. To resolve potential conflicts, the system follows a hierarchical priority logic: mandatory standardized translations issued by authorities (the whitelist) strictly override individual expert-written rules to ensure political consistency. Obviously non-compliant expressions will be directly replaced by compliant expressions and uncertain expressions will be forwarded to manual review queue. For manual review, the interface shows warning statements and preliminary analysis reason. Reviewers choose to approve / modify / reject the application and provide textual explanation for their decision. All the interception decisions are recorded in case log database and sorted by risk level, namely to handle the uncertain expressions which have related to political red lines firstly. To sustain the above mechanism in the long run, we designed a closed-loop feedback path. For each review decision, extract it into "source expression - contextual features - compliant expression" triple and anonymize and periodically inject into rule base and sensitive word database. For each recurring boundary case type, system will automatically generate a draft rule and submit for expert confirmation and add it to the first round interception logic. Terminology update issued by competent authorities will also synchronize to rule base and make the dynamic adaptation to review standards and policy guidance.

V. VALIDATION OF THE OPTIMIZATION EFFICIENCY AND ETHICAL COMPLIANCE OF JAPANESE TRANSLATION OF CHINESE CULTURE FOR EXTERNAL PUBLICITY

A. EXPERIMENTAL SETUP AND PREPARATION

This study built a three-level heterogeneous evaluation dataset in total, including excerpts from the two classic canon works (200 excerpts from *Analects*, *Classic of Poetry*, etc.), contemporary official external publicity press release (50 randomly selected articles from Japanese version of *People's Daily Online*) and manually annotated ethical risk test set (containing 500 control sample set of historical expressions and social taboos). We conducted the experiment in controlled variable comparison design and designed three processing flows: A) baseline group (directly translating from a general large language model without any adaptation intervention), B) ablation group (only loading knowledge graph and disabled ethical rule mechanism), and C) full-mechanism group (enabling the full-link mechanism "semantic mapping—contextual adaptation—ethical review" presented in this paper). All translations were finally double-blinded human evaluated by Japanese native-speaker review panel (including 3 translation scholars and 20 general

audience). Cultural imagery comprehension test adopted 0-5 Likert scale, and ethical compliance interception test was finally confirmed based on the latest "Handbook for Japanese-Related Expressions" published by State Council Information Office.

B. FIDELITY COMPARISON EXPERIMENT OF DEEP SEMANTIC MAPPING FOR CULTURE-LOADED WORDS

We randomly chose 200 passages from *Analects*, *Classic of Poetry* and poems from Tang and Song dynasties including typical culturally specific terms, and then segmented them into three parts according to the type of allusion, i.e., philosophical concepts, historical figures, and ritual objects, each of which contains 66 passages. The baseline group (GPT-4o general translation interface) and the experimental group (loading Sino- Japanese cultural knowledge graph constructed in this paper and performing LoRA domain fine-tuning) were constructed, and then Japanese translation on the test set was conducted. Finally, we invited three native Japanese proofreaders who were familiar with classical Chinese literature to score the translation independently. The evaluation granularity includes two aspects: the accuracy of cultural imagery and the fluency of translation.

TABLE 1. Fidelity Assessment of Deep Semantic Mapping of Culturally Loaded Words

Evaluation Metric	Baseline Group (General LLM)	Experimental Group (Knowledge-Enhanced + Fine-Tuned)
BLEU Score	31.4	42.7
Cultural Image Fidelity Rate (%)	58.3	85.6
Translation Fluency Score (1–5 scale)	3.9	4.4

In Table 1, the BLEU score increased from 31.4 to 42.7, indicating a significant improvement in surface fidelity of the translation. The accuracy of cultural imagery increased from 58.3% to 85.6%, demonstrating that the

introduction of knowledge graphs effectively curbed semantic shift and imagery dissolution in the Japanese translation of allusions and metaphors. The translation fluency score also improved from 3.9 to 4.4, reflecting the positive transmission effect of semantic mapping accuracy on sentence naturalness. Overall, the deep semantic mapping mechanism proposed in this paper has a clear advantage in maintaining the fidelity of Chinese cultural information.

C. CONTEXTUAL ACCEPTABILITY EXPERIMENT BASED ON AUDIENCE COGNITIVE PREFERENCE ADAPTATION

The experiment picked out five typical press releases to be used for external publicity in the following types: promotion of policy, promotion of culture and tourism, and people-to-people exchanges. A baseline group (a direct translation with use of a general model) and an experimental group (adapted versions after the softening of the tone, honorifics,

and the narrative framework) were thus produced. Twenty native Japanese speakers of the net were invited to take part in the experiment. Each subject was randomly asked to read one of the two translations and fill in a questionnaire immediately after reading it. The items on the questionnaire were a five-levels rating of the naturalness of the text and a question of willingness to share.

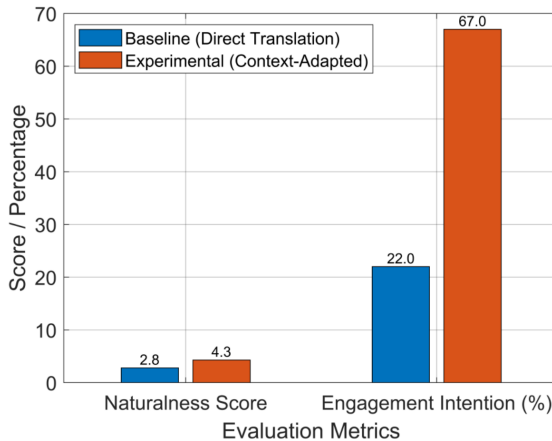


FIGURE 1. Contextual Acceptability Assessment Based on Audience Cognitive Preference Adaptation

In Figure 1, the text naturalness score of the experimental group increased from 2.8 points to 4.3 points, while engagement intention rose from 22.0% to 67.0%. While these preliminary results from our 20-person pilot group indicate a significant positive trend (up to 204.5% in this sample), further large-scale, stratified studies are required to generalize these findings to the broader Japanese population across different age cohorts. These data indicate that the adapted versions, after softening the tone, adjusting honorifics, and reconstructing the narrative framework, achieved higher reading comfort and a greater willingness to share among the Japanese audience. This difference reflects a significant tension between the macro-level tone of general literal translation and the restrained, concrete reading expectations of Japanese audiences, while contextualized dynamic adaptation can effectively bridge this gap. This experiment verifies the necessity and practical effectiveness of audience preference modeling in the production of Chinese cultural propaganda content for Japan.

D. EXPERIMENT ON THE EFFECTIVENESS OF COMPLIANCE INTERCEPTION OF ETHICAL NORMS IN EXTERNAL PUBLICITY

The experiment compiled a test library including 500 expert-annotated "risk-safety" comparative translations classified by violation type into 3 types: political red lines, historical expressions, and social taboos. Each type has 3 sub-levels: high-risk, borderline, and safe. We input this test library into the ethics review module designed in this paper. Subsequently, the rule engine carried out the first-level hard

interception, and then delivered the cases with ambiguous hard interception results to manual review. Finally, according to the results of manual cross-review, we calculated the precision, recall, and miss rate of the module in intercepting different types of illegal expressions.

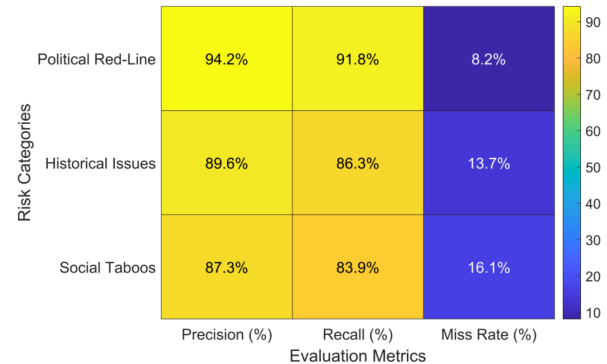


FIGURE 2. Evaluation of the Effectiveness of Compliance Interception of Ethical Norms in External Publicity

In Figure 2, the review module showed the most ideal interception effect for political red line expressions, with precision and recall rates reaching 94.2% and 91.8%, respectively, and a miss rate controlled at 8.2%. The accuracy rates for expressions concerning historical issues and social taboos were 89.6% and 87.3%, respectively, while the recall rates decreased to 86.3% and 83.9%, and the miss rates correspondingly increased to 13.7% and 16.1%. The latter two types of expressions are more context-dependent, with a higher proportion of cases having ambiguous boundaries, increasing the difficulty for the rule engine to identify them. Overall, the dual-verification mechanism has been significantly effective in intercepting high-risk content, and key political risks are now under control.

VI. CONCLUSION

This paper aims to solve the problems of cultural imagery degradation, weak contextual adaptation and lack of ethical review in the process of Chinese culture into Japanese propaganda translation, and proposes a chain-like optimization mechanism combining accurate semantic mapping, audience cognitive adaptation and ethical review. The experimental results show that the mechanism has achieved the expected results in improving the accuracy and fluency of propaganda translation, improving the willingness to read of Japanese audience and controlling the ethical risk of propaganda, and has initially achieved the synergic effect of "accurate translation, widespread dissemination and effective protection" of propaganda content dissemination technology. The limitations of this paper are that the coverage of knowledge graph is still mainly classical text, and its ability to handle contemporary online cultural terms is weak; the audience preference model does not consider the reading differences between audience of different generations. In

the future, this research will further expand the dimensions of the corpus, explore personalized adaptation strategies for granular audience, and improve the dynamic update mechanism of ethics rule base to enhance the robustness and generalization ability of ethics rule base in external communication scenario.

AUTHOR CONTRIBUTIONS

Conceptualization –Yan Xie and Jian Chen; methodology – Yan Xie and Jian Chen; investigation – Yan Xie and Jian Chen; writing-original draft preparation – Yan Xie and Jian Chen. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was funded by the First Batch of University-Level Project-Based Curriculum Construction

Projects of Hebei University of Technology, for the course Translation Theory and Practice (Official

Document No. [2026] 57 of Hebei University of Technology).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] K. Hu and X. Li, "The creativity and limitations of AI neural machine translation: A corpus-based study of DeepL's English-to -Chinese translation of Shakespeare's plays," *Babel*, vol. 69, no. 4, pp. 546-563, 2023, doi: 10.1075/babel.00331.hu.
- [2] X. Chen, "Translation Audience: Reflections from a Rhetorical Perspective," *Contemporary Foreign Language Studies*, vol. 21, no. 6, pp. 27-36, 2021, doi: 10.3969/j.issn.1674-8921.2021.06.003.
- [3] L. Chen, "Ethical Misconceptions and Governance Breaks in Generative Artificial Intelligence Translation," *Contemporary Foreign Language Research*, vol. 25, no. 6, pp. 134-141, 2025, doi: 10.3969/j.issn.1674-8921.025.06.015.
- [4] J. Q. Wang, G. Liu, and Y. Lei, "Reaffirming the importance of traditional humanities in the era of AI-enhanced translation and transcultural communication," *International Communication of Chinese Culture*, vol. 12, pp. 277-311, 2025, doi: 10.1007/s40636-025-00334-1.
- [5] H. Xiao, "International Knowledge Dissemination from the Perspective of Knowledge Translation Studies," *Contemporary Foreign Language Studies*, vol. 23, no. 3, pp. 14-24, 2023, doi: 10.3969/j.issn.1674-8921.2022.03.002.
- [6] A. Hu, "Archaeology of Knowledge, Discourse Reconstruction, and International Dissemination of Chinese Culture: A Study on the English Translation of Chinese Classics by American Sinologist Huazheng," *Contemporary Foreign Language Studies*, vol. 22, no. 5, pp. 12-21, 2022, doi: 10.3969/j.issn.1674-8921.022.05.002.
- [7] G. Yu, M. Wang, and Z. Yang, "Four key operations of communication and exchange in "mutual learning of civilizations"- On reducing and eliminating "cultural discounts" from the perspective of communication studies," *International Communication of Chinese Culture*, vol. 12, pp. 335-345, 2025, doi: 10.1007/s40636-025-00327-0.
- [8] S. Liu, C. Mao, and Y. Zhang, "A Han Vietnamese Cross border Ethnic Text Retrieval Method Based on Domain Knowledge Graph and Comparative Learning," *Journal of Nanjing University (Natural Science Edition)*, vol. 59, no. 4, pp. 610-619, 2023, doi: 10.13232/j.cnki.jnju.2022.04.008.
- [9] H. W. Nettelbeck, "Perceptions of Characters in Subtitled Film: Strategies to Achieve Similarity in Japanese-and English-Speaking Audience Understanding," *Interpreting and Translation Studies: The Journal of the Japan Association for Interpreting and Translation Studies*, vol. 22, no. 0, pp. 31-51, 2022, doi: 10.50837/its.2203.
- [10] Y. Shi, "Knowledge Ethics and Knowledge Construction in Artificial Intelligence Translation," *Contemporary Foreign Language Research*, vol. 23, no. 6, pp. 156-165, 2023, doi: 10.3969/j.issn.1674-8921.2022.06.016.