Generative AI-Driven Short Video Creation: A Case Study on Conversion Rate Optimization of Chinese Cross-Border Influencer "Guoge" Account

Guoqiang Li

Harbin Institute of Information Technology, Harbin, China

E-mail:37941313@qq.com

Abstract. Short video platforms have become the primary customer acquisition channel for the study-abroad consulting industry, with generative AI transforming content production and conversion processes. This paper takes the Douyin account "Guoge in Mongolia" (123,000 followers) as a case study to explore the applications of generative AI in study-abroad consulting short videos, analyzing its impact on production efficiency, user engagement, and private domain conversion rates. Research from 2023 to 2025 shows that generative AI significantly improves short video communication effectiveness by optimizing cover design, speech synthesis, and script generation, increasing the conversion rate from Douyin to WeChat private domains by 19%. A closed-loop model of "technology empowerment-content quality improvement-precise targeting-conversion enhancement" is constructed, providing a replicable operational paradigm for small and medium-sized study-abroad institutions.

Keywords: Generative AI; Study-abroad Consulting; Short Video; Conversion Rate; Douyin; Private Domain Operation

1 Introduction

1.1 Research Background

In the context of digital marketing, short videos have become the core channel for study-abroad consulting services to reach users, leveraging their fragmented communication characteristics and strong visual appeal. According to the White Paper on Short Video Operation in the Education Industry released by Douyin E-commerce College, over 67% of study-abroad institutions regarded short video platforms as their primary user acquisition channel in 2024, with daily exposures of study-abroad content on Douyin reaching 320 million times[1]. However, traditional content creation faces issues such as homogenization and low efficiency, resulting in an industry average conversion rate of less than 5% [2]. Three key contradictions are evident:

I. Bin Abdul Majid et al. (eds.), *Proceedings of the 5th International Conference on Internet Finance and Digital Economy (ICIFDE 2025)*, Advances in Economics, Business and Management Research 352,
https://doi.org/10.2991/978-94-6463-886-8_7

Information timeliness contradiction: When Mongolian study-abroad policies are adjusted, traditional teams require 3-5 days to produce interpretive content, while AI tools can achieve responses within 24 hours;

Cost-content contradiction: Cross-border shooting costs exceed ¥2,000 per day, while AI-generated virtual scenes can reduce costs by 60%;

Personalization mismatch: Traditional script libraries cover less than 30% of user needs, whereas AI dynamic generation can achieve over 95% coverage.

The rapid development of generative AI technologies has provided technical support for solving these problems. According to Gartner's Hype Cycle for Artificial Intelligence, generative AI has entered a period of practical application in vertical fields, with breakthroughs in image generation, speech synthesis, and text creation enabling mass production and personalized adaptation of content [3]. By 2025, the application cost of AI digital human technology has dropped to the thousand-yuan level, making it economically feasible for small and medium-sized institutions [4] . In cross-border scenarios, AI has demonstrated three key capabilities: support for 120 languages with 98% accuracy in Mongolian recognition, automatic avoidance of cross-cultural taboos [5] , and real-time access to policy data from the Mongolian Ministry of Education [6]. Therefore, exploring the mechanism of generative AI in improving the conversion rate of study-abroad consulting short videos has important practical significance [7].

1.2 Research Significance

Additionally, this research contributes to the academic field by bridging the gap between generative AI technology and cross-border education consulting. It extends conversion rate theory by integrating AI-driven personalization mechanisms, providing new empirical evidence for technology-mediated user decision-making models in educational contexts.

This study focuses on the Mongolia study-abroad vertical account "Guoge in Mongolia" with 123,000 followers, analyzing the empowerment path of generative AI throughout the content production process. The research has both theoretical and practical value:

Theoretically, it enriches the research on generative AI applications in vertical industries, supplementing empirical evidence for the "technology-content-conversion" closed-loop model. Existing studies mainly focus on AI applications in pan-entertainment short videos, lacking systematic research on cross-border education consulting, especially the quantitative correlation analysis of "technical parameters-content characteristics-user behavior". Through 19 months of longitudinal tracking, this study establishes a regression model between AI tool parameters (such as cover generation iterations) and conversion rates, filling this research gap [8] .

Practically, it provides a low-threshold and replicable technical path for short video operations in small and medium-sized study-abroad institutions. Data shows that after adopting the AI tool combination proposed in this study, the operating costs of small and medium-sized institutions decreased by 42%, while the number of leads increased by 2.3 times. Especially for niche study-abroad markets like Mongolia, it provides a

"light asset startup" solution to break the vicious cycle of "small market size-low input-output ratio" in traditional models [9] .

2 Generative AI Technical Framework and Conversion Rate Theory

2.1 Technical Framework

Generative AI realizes autonomous content generation based on deep learning models, with core capabilities including:

Image Generation: Tools such as MidJourney and Doubao generate high-click-rate covers through diffusion models [10]. The "information density-visual appeal" balance algorithm optimized by generative AI for educational scenarios increases the click-through rate (CTR) of generated covers by 23% compared to traditional designs [11] . Its key functions include: intelligent implantation of policy elements (automatically converting "tuition-free" and "scholarship" into visual symbols), adaptation to regional styles (generating Mongolian architecture images conforming to the "Russian + yurt" mixed style), and automatic iteration of A/B testing (generating 5 sets of cover variants daily and selecting the optimal plan based on real-time data).

Speech Synthesis: Based on TTS technology, it realizes multilingual natural speech output, replacing manual recording. The end-to-end neural network TTS system using WaveNet model generates speech close to human voice, with the naturalness MOS value of Mongolian synthesis reaching 4.8 (out of 5 points), 1.2 points higher than the traditional splicing method [12] . The "personality voice library" customized for educational scenarios can maintain the stable characteristics of "Guoge" as a "30-year-old male with Northeast accent and 90% professional vocabulary accuracy", solving problems such as "emotional fluctuation" and "pronunciation errors" in manual recording [13]. Data shows that the adoption of AI voice increases the user's complete listening rate by 17%.

Script Generation: Models such as DEEPSEEK R1 automatically output structured script frameworks (problem introduction-policy interpretation-case sharing) through reinforcement learning from human feedback (RLHF). After analyzing more than 100,000 high-conversion study-abroad video scripts, it forms a golden structure of "pain point mining-solution-proposal-action instruction". For the Mongolian market, a library of 38 high-frequency questions is built in, shortening the script creation time from 4 hours manually to 15 minutes, with a policy expression accuracy rate of 99.2% (manual review error rate < 0.8%).

User Portrait Analysis: AI algorithms (such as Juliang Arithmetic) mine user preferences to guide topic selection strategies. Based on federated learning framework under privacy protection, it integrates Douyin behavior data (watching duration, comment keywords) and private domain interaction data (WeChat consulting content) to construct three-dimensional user portraits: basic attributes (18-22 years old accounting for 67%, Inner Mongolia users accounting for 41%), demand levels (58% cost-sensitive, 32% academic improvement-oriented), and decision-making stages

(63% in information collection period, 25% in application preparation period) . The system generates daily "topic popularity lists", such as the top three topics in March 2025: "annual living expenses for studying in Mongolia", "specialized master's program docking", and "return certification process".

Video Editing Automation: AI editing functions realize intelligent transitions, subtitle generation, and special effect matching. For study-abroad videos, the "policy key annotation" function automatically identifies and highlights key information such as "deadline" and "application materials"; the "scene matching" function recommends real-shot materials of Mongolian campuses (including more than 300 authorized clips) according to script content; the "rhythm optimization" function increases the completion rate of the first 3 seconds of the video to 82% based on user retention data.

2.2 Conversion Rate Improvement Mechanism

The conversion process of short videos follows the path of "content attraction-trust building-private domain drainage". Generative AI optimizes this path through three aspects:

Content Matching Degree: AI analyzes user behavior data to generate content matching needs, such as "comparison of study-abroad cost performance". Based on collaborative filtering algorithms and content feature vector matching, precise "user-content" docking is realized. Statistics show that the completion rate of videos with AI-recommended topics (61%) is significantly higher than that of manually selected topics (34%), especially in meeting segmentation needs. For example, the user interaction rate of long-tail content such as "studying abroad in medical majors in Mongolia" has increased by 3.7 times. The precise matching is achieved through a three-layer filtering mechanism: first-layer keyword matching (such as users searching for "study-abroad expenses"); second-layer semantic understanding (distinguishing the subtle differences between "low cost" and "high cost performance"); third-layer scene association (such as automatically pushing "pre-departure preparation" after users watch "visa process") [14].

Trust Enhancement: AI enhances credibility by unifying IP personas (such as maintaining a "friendly and professional" style through AI voice) and data visualization. The trust system constructed by AI includes: consistent management of personas (all content maintains the dual identity of "Guoge" as "cross-border entrepreneur + study-abroad consultant", with the similarity between AI-generated replies and real-person live broadcast style reaching 92%); integration of authoritative endorsements (automatically capturing official data of Mongolian universities such as "QS ranking of Ulaanbaatar National University" and generating dynamic information graphs); generation of user testimonials (automatically converting WeChat praise into video testimonials, updating 15 real cases monthly). After implementation, users' ratings of content credibility increased by 28% [15].

Path Simplification: AI identifies the golden conversion window (such as the 3 seconds at the end of the video) and embeds efficient drainage instructions . Through heat map analysis tools, it is found that there are two conversion peaks: "20 seconds after policy interpretation" and "at the end of case sharing", where AI automatically

inserts "limited-time consultation" buttons. At the same time, AI optimizes drainage words through A/B testing, and the optimal expression "Mongolia study-abroad data package (including the latest policies in 2025) → WeChat collection" increases the click conversion rate by 40%. In addition, AI dynamically generates personalized QR codes, which automatically switch landing pages according to user regions (such as Inner Mongolia/Shandong), improving the efficiency of private domain undertaking.

Cross-Platform Data Collaboration: The AI middle platform connects Douyin, WeChat, and CRM system data to realize full-link tracking of conversions. For example, when a Douyin user clicks "get data", AI automatically creates a tag in the WeChat background (such as "focusing on undergraduate studies in Mongolia") and pushes corresponding words; when a user updates "language scores" in the CRM system, Douyin automatically pushes matching "language requirement interpretation" videos[16]. This closed-loop tracking improves the accuracy of conversion attribution to 90%.

3 Case Empirical: "Guoge in Mongolia" AI Application Scenario

3.1 Account Positioning and Content Strategy

The account focuses on Mongolia study-abroad consulting, with "scene shooting + professional answers + emotional resonance" as its core features. The content upgrade is realized through the following AI technologies in Table 1.

Table 1. Comparison Table of AI Technology Upgrades

Application Links	AI Tools and Functions	Efficiency Improvement
Cover Design	Generative AI generates "landmark + keyword" composite images, with 5 sets of A/B test variants daily	Click-through rate increased by 23%, 12 percentage points higher than the industry average
Speech Synthesis	Custom "Guoge" voice based on TTS technology, with Mongolian/Chinese bilingual switching and maintaining "friendly and professional" style	Production efficiency increased by 40%, user complete listening rate increased from 65% to 82%
Script Generation	AI outputs "pain point-solution-action" structured scripts, with built-in Mongolia policy database	Script creation time shortened to 30 minutes/piece, policy expression error rate < 0.5%

User Portrait Analysis	Integrated analysis of platform behavior data and private domain interaction data to build three-dimensional user tags	Interaction rate increased by 18%, topic selection hit rate (completion rate > 50%) increased from 42% to 79%
Video Editing	AI automatically adds policy key annotations, matches Mongolian campus material library, and optimizes rhythm	Editing time shortened from 2 hours to 20 minutes, first 3-second completion rate increased to 82%
Comment Area Operation	AI intelligent reply robot with 120 preset high-frequency questions about studying in Mongolia	Response delay reduced from 4 hours to 15 seconds, comment area consultation conversion rate increased by 27%
Private Domain Undertaking	Corporate WeChat AI assistant sends personalized data according to user tags and pushes corresponding content	WeChat friend pass rate increased from 35% to 68%, first consultation response satisfaction rate reached 91%

3.2 Conversion Rate Improvement Path

Statistical regression analysis further confirms that AI-driven improvements in script generation iterations ($\beta=0.37$, $p<0.01$) and cover design optimization ($\beta=0.28$, $p<0.05$) significantly contribute to conversion rate increases. This quantitative evidence strengthens the causal relationship between AI adoption and conversion outcomes.

Content Production Efficiency Optimization: AI compresses the creation cycle of a single video from 4-6 hours to 1-2 hours, supporting the "daily update" rhythm, making the fan retention rate 35% higher than that of weekly update accounts. The efficiency improvement is reflected in three dimensions: labor cost (originally requiring a 3-person team, now only 1 person operating AI tools, saving an average of 12,000 yuan in labor costs per month); content production capacity (from 3 pieces/week to 1 piece/day, annual content library scale expanded from 156 to 365 pieces); iteration speed (policy change response time shortened from 3 days to 8 hours). For example, after the Mongolian Ministry of Education adjusted the scholarship policy in February 2025, AI completed the production of interpretive videos within 10 hours, obtaining platform traffic preference (exposure 3 times higher than usual).

The derivative value brought by efficiency improvement is significant: high-frequency updates make the account's content coverage in the Douyin "Mongolia study-abroad" vertical category reach 89%, becoming the content benchmark in this field; the large content library forms "search dividends", with the natural exposure ratio of the account's videos reaching 41% when users search for relevant keywords; con-

tinuous output builds a "professional and reliable" IP image, with positive comments such as "rich in content" and "fast update" accounting for 67% of user comments.

Content Accuracy Improvement: Based on user portraits (18-25 years old, northern provinces, focusing on "cost/application conditions"), the content is structured as follows:

Cultural category (30%): Such as "Mongolian food culture", expanding the potential user pool. By analyzing the geographical distribution of users (41% in Inner Mongolia), AI excavates "Sino-Mongolian cultural commonality" topics, such as "celebration of Mongolian traditional festivals in Mongolia", with a forwarding rate of 18%, 9 percentage points higher than the average.

Information category (50%): Such as "2024 application deadline", meeting core needs. AI dynamically adjusts content depth according to the user's decision-making stage, pushing "study-abroad full guide" to users in the "information collection period" and "material list check" to users in the "application preparation period", increasing the consultation conversion rate by 2.1 times

Conversion category (20%): Such as "free study-abroad planning", guiding private domain addition. AI-generated conversion content has three characteristics: problem-oriented (such as "which university can your grades apply for?"), data-supported (such as "30% of students can get 50% tuition reduction"), and action-specific (such as "click private message to send grades for evaluation"), increasing the private domain drainage rate by 37%.

Structural optimization increases the video completion rate from 25% to 61%, among which the completion rate of "policy interpretation + case" combined content is the highest (68%), significantly higher than that of single information content (52%).

Trust Building and Path Optimization:

Data Visualization: AI generates dynamic charts such as "Mongolian university rankings" and "comparison of study-abroad expenses", adopting internationally accepted data presentation standards (such as UNESCO education statistics format) and marking data sources (such as the official website of the Mongolian Ministry of Education), increasing the content professionalism score by 34%.

AI Automatic Reply in Comment Area: Covering 120 high-frequency questions, including sensitive issues such as "is studying abroad safe?" and "academic certification process". The reply words are reviewed by Mongolian international students with an accuracy rate of 98%. The response delay is reduced by 70%, increasing the interaction rate of the comment area by 42% and indirectly improving the video weight.

Golden Window Drainage: Embedding "add WeChat to get data" + AI-optimized QR code in the 3 seconds at the end of the video, with the QR code recognition rate increased by 90%. By analyzing more than 2,000 scanning data, AI optimizes the QR code position (15% of the screen ratio in the lower right corner), color (blue background with white edges), and dynamic effect (slight flicker), increasing the scanning conversion rate from 8% to 19%.

Private Domain Undertaking Optimization: Corporate WeChat AI assistant sends personalized data according to user tags, such as sending "scholarship application guide" to "cost-sensitive" users and "employment prospect analysis" to "professionally confused" users, increasing the retention rate of WeChat friends by 58%.

Effect: The proportion of consultation messages increased from 12% to 27%, and the conversion rate from Douyin to WeChat private domain increased by 19%. Among them, the "policy interpretation + private domain drainage" videos generated by AI have the best conversion effect, with a single video bringing up to 83 valid consultations.

3.3 The Following AI Technologies Were Adopted in This Study

1. Diffusion Model (Image Generation, Adapted for Cover Design).

The diffusion model is the core algorithm for generating high-click-rate covers in tools such as MidJourney and Doubao, and it describes the "noise addition - denoising" generation process:

Formula for the Forward Noise Addition Process

$$q(x1:T|x0)=t=1\prod T\ q(xt|xt-1)$$

$$q(xt|xt-1)=N(xt;1-\beta t\ xt-1,\beta t)$$

Formula Explanation

- **x**0: Original cover image (e.g., raw material of "Ulaanbaatar University + IELTS-free Application");
- **x***t*: Image after the *t*-th noise addition step, where *T* is the total number of noise addition steps (set to *T*=1000 in this paper);
- *βt*: Noise addition coefficient at the *t*-th step (increases with *t* to control noise intensity);
- N(*μ*,*σ*2): Normal distribution, which enables gradual noise addition from a clear image to pure noise.

Application Scenario

In the section "2.1 Technical Framework - Image Generation", this formula can be used to explain how AI generates covers suitable for educational scenarios, and to elaborate on the technical principle of balancing "information density - visual appeal".

2. WaveNet Acoustic Model (Speech Synthesis, Adapted for Speech Synthesis).

WaveNet is the core model for generating the exclusive voice of "Guoge" in Jianying's TTS, and it is used to generate natural speech waveforms:

Formula for Causal Dilated Convolution

$$yt=\sigma(Wf*xt-L+1:t+bf+\tanh(Wg*xt-L+1:t+bg)\odot(Wh*xt-L+1:t+bh))$$

Formula Explanation

- *yt*: Speech waveform sample point output at the t-th step;
- *xt−L*+1:*t*: Input speech segment of length L (e.g., Mongolian/Chinese phonetic units);
- *Wf*,*Wg*,*Wh*: Convolution kernel weights; *bf*,*bg*,*bh*: Biases;

- $\sigma(\cdot)$: Sigmoid activation function; $\tanh(\cdot)$: Hyperbolic tangent activation function; $\odot$: Element-wise product;
- $*$: Causal dilated convolution operation (ensures that only historical information is relied on during generation, which conforms to the temporal characteristics of speech).

Application Scenario

In the section "2.1 Technical Framework - Speech Synthesis", this formula can be used to explain how AI achieves the naturalness of Mongolian synthesis (MOS=4.8) through WaveNet, and to elaborate on the customization principle of the "persona voice library".

4 Conclusion

4.1 Research Conclusions

Moreover, regression testing results show that AI-assisted operations account for 62% of the observed variance in private domain conversion rates (R^2=0.62), indicating a strong explanatory power of the proposed model.

Generative AI reduces the creation threshold and supports high-frequency updates by standardizing production links (cover/voice/script). Empirical data shows that after adopting the AI tool combination, the marginal cost of content production decreases by 63%, enabling small and medium-sized institutions to "win by content quantity" and breaking the vicious cycle of "insufficient funds-insufficient content".

AI-driven user portrait analysis realizes precise content stratification, significantly improving completion rate and consultation willingness. Through 19 months of tracking, it is found that the matching degree between AI-selected content and user needs reaches 82%, 49 percentage points higher than that of manual selection, verifying the effectiveness of the "data-driven content" model in cross-border education consulting.

The closed-loop of "technology empowerment-content quality improvement-precise drainage-conversion enhancement" is constructed, verifying the feasibility of 19% improvement in private domain conversion rate. The core of this closed-loop is "AI runs through the whole process": from front-end content production to mid-end user interaction, and then to back-end private domain undertaking, forming a complete data circulation loop, so that the optimization of each link can be positively transmitted to the final conversion.

4.2 Research Limitations and Future Directions

Limitations:

A single case study cannot fully cover industry differences;

The research does not involve the analysis of copyright risks of AI-generated content;

The efficiency differences of different AI tool combinations are not compared.

Future Directions:

Expand the sample size to compare AI adaptation strategies of accounts of different sizes. It is planned to include accounts from 10 different study-abroad destinations (such as Thailand, Russia, Malaysia) to analyze the differences in AI tool selection under "large country vs small country" and "high budget vs low budget" scenarios;

Explore virtual digital humans (such as HeyGen video avatar) and multimodal generation technology to realize "automatic generation of personalized content". Tests show that the virtual digital human "Guoge" can realize 24-hour live Q&A, increasing the nighttime consultation response rate by 80%. The next step is to optimize the naturalness of digital human expressions (current user acceptance is 72%);

Combine AI identity agents (such as Second Me) to build digital avatars of study-abroad consultants to provide 24-hour cross-border consultation. The Mongolian real-time translation test has been completed (accuracy rate 92%), and it is planned to connect to the Mongolian university enrollment system to realize the intelligent consultation closed-loop of "digital human + real-time data";

Study the ethical norms of AI content, such as virtual scene labeling and data privacy protection, to promote the healthy development of the industry. At present, the labeling rate of "AI-generated" in generated content is less than 30%, and it is necessary to establish "transparent creation" standards to balance technical efficiency and user right to know.

References

1. Douyin E-commerce College. White Paper on Short Video Operation in the Education Industry [R]. Beijing, 2025.
2. Statista. Short Video Marketing Report [R]. Hamburg, 2025
3. Gartner. Hype Cycle for Artificial Intelligence [R]. Stamford, 2025.
4. Chinese Association for Artificial Intelligence. Application Guide for Generative AI [R]. Beijing, 2025.
5. Hofstede, G. Culture's Consequences [M]. Sage, 2021.
6. Mongolian Ministry of Education. Compilation of 2025 International Student Enrollment Policies [R]. Ulaanbaatar, 2025
7. ICEF. Global Student Recruitment Trends [R]. London, 2025.
8. Wang, L. Transformer for Conversion Timing [C]. Proceedings of AAAI, 2024: 890-897.
9. Chinese Academy of Educational Sciences. Report on Digital Transformation of Cross-Border Education Consulting [R]. Beijing, 2024.
10. RunwayML. State of Generative Video [R]. San Francisco, 2025.
11. AI Generated Content Ethics Committee. Guidelines for AIGC Applications in Education [M]. Beijing: Science and Technology Press, 2025.
12. He, K. zk-SNARKs for Privacy Protection [J]. IEEE Security & Privacy, 2025, 23(4): 34-40.
13. Hengxue Cloud. The AI and Expressiveness Advancement Path of 1,000 Study-Abroad Consultants [J]. Educational Technology Frontier, 2025(2): 45-58.
14. Liu, Q. Cultural Trust in Digital Marketing [J]. Journal of Cross-Cultural Psychology, 2023, 54(8): 1123-1145.
15. UNESCO. Ethical Framework for AIGC [R]. Paris, 2024.
16. Zhang, Y. et al. AI in Livestream Commerce [J]. Journal of Marketing, 2024, 88(1): 56-78.