

A Study on the Understanding of AIGC in the Concept Images of Chinese Poems

—Take PIKA and Aibrm for Examples

Peishan Cai^{1,a}, Yan Zhou^{2,b,*}

¹*School of Foreign Languages, University of Electronic Science and Technology Zhongshan Institute, Zhongshan, Guangdong, 528400, China*

²*School of Education Science, Hunan Normal University, Changsha, Hunan, 410000, China
a. 1290394305@qq.com, b. zhouyan1818@hunnu.edu.cn*

**corresponding author*

Abstract: With the rapid development of natural language processing (NLP) and computer vision technology, the technology of artificial intelligence creative generation (AIGC) has achieved automated conversion from text to video. However, it is still uncertain whether these technologies can provide a deeper understanding of the concept images in Chinese poetics while the videos are generated more efficiently and conveniently. This study analyzes 9 AIGC software based on 8 evaluation dimensions, involving the representative domestic software Aibrm and PIKA with high popularity from aboard. The authors designed a questionnaire survey of the matching rate between poem videos generated by AIGC and poem concept images, in terms of Chinese teachers. Also, researchers explore the concretization of abstract concepts in AIGC software. The research found that there are differences in the understanding of concept images of Chinese poetics by different AIGC tools. More than 50% of respondents prefer poetry videos generated by Aibrm, but some believe that both videos do not match poetic concept images. This is mainly due to stiff roles and contexts, insufficient integration, lack of video coherence, and insufficient emotional communication. Despite the development of the AIGC, the understanding of poetic concept images should be identified well and improved more. The application of the Chinese poem concept, emotional expression, and cultural integration still requires continuous exploration.

Keywords: AIGC, Chinese poem, concept images.

1. Introduction

Because of the improvement of natural language processing (NLP) and visual technology, the technology of artificial intelligence creative generation (AIGC) has achieved flexible conversion from text to video. These technologies help create videos more efficiently and conveniently, as well as decrease the content photographing conditions. Currently, it promotes the application of AIGC in the field of video generation with the improvement of hardware computing ability and the enhancement of large-scale model training levels. Numerous video generation platforms and tools pop up in the

market, like LUMA, VIDU etc. These platforms utilize AI technology to change the traditional way of video production, enabling users to easily automate the generation and editing of video content.

For example, the Sora model released by OpenAI can generate a 60-second coherent video only based on prompts, far exceeding the industry average. With the continuous development of large model technology, the application of generative AI in the field of video generation has been significantly improved. The foundation model can be trained with a large of unlabeled data and analyze features and rules, generating videos that are similar to the training results. Deep learning models such as Generative Adversarial Networks (GANs) and Variational Autoencoders play a crucial role in AIGC video creation. GANs continually ultimate the content of videos by adversarial training of generators and discriminators; VAE accurately catches high-level features by encoding input data into latent space and decoding to generate new data.

Recently, artificial Intelligence has made significant progress in its application between literary and artistic fields. Specifically, there is a huge potential for generating poem videos. Concept images of poetry are an extremely vital and unique concept in literature. It combines the poet's subjective feelings with objective scenery or events. Through refined language and rich imagination, an artistic realm of transcends concrete images and contains profound meanings can be created. All in all, concept images are the harmonious unity of emotion and scenery, meaning and images in poetry, and it is an aesthetic experience that readers can realize more than the text itself in the process of appreciation. Learners face difficulties because of implicative characteristics in Chinese and foreign poetry, including language comprehension barriers, the difference in cultural background, complexity of poetic form and structure, insufficient emotional experience, and lack of problem-solving skills. In order to solve these problems and achieve concretization from abstract concepts, AIGC helps learners analyze the concept images from the text and transit it into visual video.

This research studies 9 kinds of AIGC with 8 evaluation dimensions, the authors select Aibrm software among domestic AICG cetology and PIKA software from overseas. A questionnaire survey has explored which AICG performs better in matching poem concept images and video content, *the matching rate between poem video generated by AICG and poem concept images, in terms of Chinese teachers*. With the deep exploration, artificial intelligence applications can be promoted competitively in the field of poetry literature.

2. Literature Review

The spark of collision between artificial intelligence (AI) and poetry has shown enormous potential, and opens up new paths for the diversified expression of culture and art. However, the quality evaluation of Chinese poems and foreign poetry videos generated by AI is a complex and multidimensional problem. They are not only technical indicators such as the clarity and fluency of video concept images, but also subjective feelings such as the communication of poetic concept images and cultural background. At present, people are constantly exploring it.

Nowadays, many scholars and experts have made researches related to AIGC, which shows that artificial intelligence can analyze the rhythm of poetry and generate matching visual elements. Through AI deep learning and natural language processing (NLP), human can interact with artificial intelligence [1]. These videos enhance the expressive power of poetry through visual art, providing learners with a brand-new aesthetic experience [2,3]. In the education field, poetry videos created by AI have been proved to effectively enhance students' interest in learning, which has injected new vitality into the traditional teaching mode [4].

Nevertheless, current studies mentioned that AI still faces many challenges when generating poetry videos, such as the accuracy of emotional expression, the compatibility between visual and textual elements [5]. Future research needs to focus on optimizing the algorithm model, and improving AI's ability to understand and reproduce the deep meaning of poetry. Hoping to provide more videos with

higher quality [6]. Lately, among the generative artificial intelligence software, there are quite a lot of tools that can generate videos efficiently, and they have their own characteristics. For example, Vidu, which is jointly launched by Shengshu Technology and Tsinghua University, is one of the leading AI video generation tools in China. By integrating Diffusion and Transformer technologies, it is possible to carry out video content with long duration, high consistency, and high dynamism. Ying AI, Kling, Aibrm and other AI products, have all achieved technological breakthroughs in converting text into video. Exploring more application scenarios of AI technology in literary education and cultural dissemination, it is also an important direction for future research [7].

In China, experts in the educational field are positively exploring the application of AI technology, especially in literary appreciation, and cross-cultural communication [8]. Videos produced by AI systems has become a new style of teaching resource, which can arouse students learning interest and help learners understand the poem connotation with different cultural backgrounds [9]. In addition, domestic researchers are also concerned about how to use artificial intelligence technology to improve teaching quality. These efforts not only enrich teaching methods, but also promote the educational quality [10].

In terms of foreign academic research, AI application research in the field of education started earlier. In the field of poetry video generation, foreign researchers not only pay attention to the optimization and innovation of technology, but also discuss its application effect and potential value in teaching practice [11]. At the same time, foreign educational institutions encourage learner to participate in the creation and evaluation process, and cultivate their innovative thinking and critical thinking skills [12].

Remarkably, a variety of fields based on generative AI have obtained significant results and breakthroughs. Whereas, average tools have an understanding differences of poetry, then the authors invest the application exploration in poetry. Software with high user quantity, good reputation, advanced technology, and outstanding research teams on the market have been considered. The authors input romantic ancient poems into the system for comparison, so as to choose the software that is more in line with the research purpose.

3. Method

3.1. Models Selection

3.1.1. Software Sample Selection

In this study, the authors select 9 AIGC products from home and abroad by comparing the technology level of AIGC tools which are the latest and frequently used. The reasons why the authors consider are that they possess subscribers' recognition, a high comprehensive score and published in the market in these three years. Notably, they are representative.

Among these 9 software, there are 3 foreign ones, namely LUMA, which focuses on realistic 3D model technology and realizes the rapid conversion from the real world to digital; with high definition and high resolution, whether it is color, detail or dynamic effects, Runway Gen 2 has reached the standard of film and television; with millions of users and millions of videos generated every week, Pika is developed by a technical team of former AI researchers from well-known tech companies such as Google, Meta and Uber. There are six models in China, which are China's first self-developed video model VIDU with long duration, high consistency and high dynamics. It can understand and generate unique China elements, such as pandas and dragons. Users can customize the character images, action and scene description, provide a variety of different styles of characters for users to choose from, freely adjust the layout of the mirror during the creation process, and integrate the background music and narration and dubbing functions. Dreamina has an intelligent painting system

based on Generative Adversarial Network (GAN) that provides a variety of painting styles, such as watercolor, drawing, oil painting, etc., and can perform secondary editing of the generated paintings. Kling has a 3D spatio-temporal joint attention mechanism, which helps to accurately capture the complex spatio-temporal motion in the video, and the modeling ability of self-developed model architecture and Scaling Law stimulation. Reproduce the real physical world characteristics in the virtual scene, such as light reflection, fluid dynamics under the action of gravity, etc., and deeply understand the three technical highlights of text semantics based on the powerful concept combination ability of Difficulty Transformer architecture. Ying AI is excellent in the speed and efficiency of video generation. It adopts 3D VAE technology to effectively compress video data and improve the ability to capture the relationship between frames, so as to ensure the natural smoothness of the generated video. With its easy-to-understand interface, efficient creation capabilities and rich material library, WHEE provides creators with convenient and efficient creation solutions. It reduces the number of complex instructions that users need to input, making the creative process smoother and more intuitive whether or not with a rich technology background.

3.1.2. Text Sample Selection

Chinese poetry occupies a prominent position in Chinese traditional culture. It is not only a literary form, but also an imperative carrier of human emotions, thoughts and culture. Concept images is an unchangeable part of poetry, so for the selection of texts, Chinese poems with romanticism should be selected for text input, for instance, "A thousand times I search for her in the crowd. And, suddenly turning my head, discover her where the lantern lights are dim.", from *The Lantern Festival Night-to the tune of Green Jade Table*, Qiji Xin, China, translated by Xianyi Yang and Naiyi Dai. The first sentence shows the protagonist's persistent pursuit and unremitting efforts to the goal. No matter how many difficulties and challenges people face, never give up easily. At the moment of " suddenly turning my head ", it became the turning point of the whole concept images. After a long search, the protagonist began to doubt whether his persistence was meaningful. However, at this moment, an inadvertent turn back revealed the person he had been looking for. This unexpected reunion brought great surprise and satisfaction. Secondly, " discover her where the lantern lights are dim " reveals the unique temperament and environment of the person being sought. On the bustling and brightly lit night of the Lunar New Year, most people are enjoying the excitement and joy of the festival, but he chose a relatively deserted and quiet place. This distinguished attitude towards fame and fortune, detached from the outside world, forms a sharp contrast with the hustle and bustle around him, making him even more unique and extraordinary. This poem tells people that sometimes the things people desperately pursue may not be in bustling places, but in those seemingly ordinary and even overlooked places, and also reminds them that in the process of pursuing their goals, they should keep a normal mind and not be too impatient and utilitarian. Sometimes, if they slow down and look back, they may find different scenery and gains.

Table 1: Prompt poetry list.

Grade	Poetry name
first grade	<i>Thoughts on a Tranquil Night</i>
second grade	<i>Fishing in Snow</i>
fifth grade	<i>Mooring by Maple Bridge at Night</i>
fifth grade	<i>Everlasting Longing</i>
Seventh grade	<i>Tian Jing Sha: Autumn Thoughts</i>
Seventh grade	<i>The Dream Like Order</i> <i>Last night, the rain was sparse and the wind was strong</i>

Table 1: (continued).

Eighth grade	<i>Drinking alcohol (part five)</i>
Eighth grade	<i>The Lantern Festival Night-to the tune of Green Jade Table</i>
Eighth grade	<i>Linjiang Immortal · Dream Back Tower High Lock</i>
Ninth grade	<i>Drunken Flower Yin · Mist Thick Clouds Sorrow Eternal Day</i>
Ninth grade	<i>The Water Tune Song-When will the bright moon appear</i>
Ninth grade	<i>Yu Meiren-When is it time for spring flowers and autumn moon</i>
High school sophomore	<i>Preface to Tengwang Pavilion</i>

This questionnaire survey selects 9 ancient poems that primary and secondary school students must memorize (Table 1). These Chinese poetries have famous literary sentences that has been passed down through the ages like *The Lantern Festival Night-to the tune of Green Jade Table*. It is the content that teachers and educator need to lead learners to focus on learning, understanding, feeling in Chinese textbook and literature. The responders are Chinese language teachers in the related questionnaire, they have more authority and ideas about assessment of does the video generated by AIGC conform to the concept images.

3.2. Software Comparison

This study compares the video works generated by AIGC from multiple dimensions, including video style type, number of camera movements, selectivity of background music, output video ratio, mode, clarity type, generation time of individual videos, duration of works, etc. According to these dimensions, the authors conducted statistical analysis. (see Table 2 and Table 3)

Table 2: Performance comparison of overseas AIGC

Evaluation items	Luma	Runway Gen 2	Pika
Quantity of styles	N/A	33	7
Quantity of camera movement	N/A	12	8
Quantity of background music	N/A	N/A	1
Quantity of aspect ratio	N/A	6	6
Quantity of resolution	N/A	2	N/A
Time spent	240s	240s	180s
Video duration	3s or 5s	3s or 5s	3s
Characteristics	1.Video duration can be lengthened	1.Lip-sync 2.Adjustable strength of motion	1.Various tones 2.Adjustable frames fer second 3.Adjustable strength of motion 4.Consistency with the text

3.3. Questionnaire Survey

This survey focuses on the concept images of ancient poetry. It is necessary to examine whether AIGC integrates the life picture described in the poem with the thoughts and feelings expressed, forming an artistic picture. The authors examine whether the video makes a reasonable association

with the artistic images in the poem and the art it triggers, so as to depict an emotional atmosphere or spiritual realm that can touch people's hearts. There will be a huge subjective potential during the assessment period, accordingly, the authors design a questionnaire of *the matching rate between poem video generated by AICG and poem concept images, in terms of Chinese teachers*.

Table 3: Performance comparison of domestic AIGC

evaluation items	Vidu	Dreamina	Kling	Ying AI	Aibrm	Whee
Quantity of styles	N/A	N/A	N/A	4	11	N/A
Quantity of camera movement	N/A	15	5	4	N/A	N/A
Quantity of back ground music	N/A	3	N/A	45	39	N/A
Quantity of aspect ratio	N/A	6	3	N/A	N/A	5
Quantity of resolution	N/A	2	2	N/A	N/A	N/A
Time spent	35s	240s	300s	125s	180s	600s
Video duration	4s or 8s	3s/6s/9s/12s	5s	6s	8s	2s
Characteristics	N/A	1.Adjustable strength of motion	1.Adjustable creativity degree 2.Negative prompt	1.Emotional keynote	1.Character customizing 2.Storyboarding 3.Show title	1.Intelligent association

The questionnaire was distributed to Chinese language teachers in primary and secondary schools, regardless of geographical location or teaching experience. It was designed by the authors themselves, including three aspects of information, the basic information of the respondents, the choice of videos that conform to the poetic concept images and the choice of videos that conformed to a specific concept image. The research consisted of text fill-in-the-blank and multiple-choice questions, but in each multiple-choice question, except for the specified options, when the respondent chose "I don't think they match", responders needed to briefly explain the reasons. The questionnaire was distributed online for one week. A total of 87 questionnaires were collected, and the author excluded 34 invalid questionnaires based on the completion time and completeness of the questionnaire questions. The exclusion criteria were completion time within 100 seconds or failure to complete all questionnaire questions. There was a total of 53 valid questionnaires, with a validity rate of approximately 60%.

This questionnaire has 16 questions, which are composed of text fill-in-the-blank and multiple-choice questions. (Table 4)

Table 4: Questionnaire information

Item	Question	Question type
Interviewee basic information	What age group do you teach Chinese?	Text input
	How long have you been teaching Chinese?	
Concept-image (Yi-Xiang) Fit degree of videos (Chinese poems)	Which video do you think is more in line with <i>Thoughts on a Tranquil Night</i> concept image?	Multiple choice
	Which video do you think is more in line with <i>Fishing in Snow</i> concept image?	
	Which video do you think is more in line with <i>The Lantern Festival Night-to the tune of Green Jade Table</i> concept image?	
	Which video do you think is more in line with <i>Mooring by Maple Bridge at Night</i> concept image?	
	Which video do you think is more in line with <i>Everlasting Longing</i> Concept image?	
Concept-image (Yi-Xiang) fit degree of videos (specific Concept-image)	Which video do you think is more in line with “loss” concept images?	
	Which video do you think is more in line with “desolation, loneliness ” concept images?	
	Which video do you think is more in line with“ missing, loneliness” concept images?	
	Which video do you think is more in line with“ peace, appreciation” concept images?	
	Which video do you think is more in line with“ farewell, wish” concept images?	
	Which video do you think is more in line with“ anguish, grief ” concept images?	
	Which video do you think is more in line with“ wander, solitude” concept images?	
	Which video do you think is more in line with“ peace, naturalism” concept images?	
	Which video do you think is more in line with“ regret spring, naturalism, inner anguish” concept images?	

4. Result

4.1. Software Comparison Result

As shown in the Table 2 and Table 3, among the styles that can generate video presentations, RUNWAY, PIKA, and aibrm are the most diverse, with RUNWAY being the most prominent among them. The most commonly used auxiliary operations for customizing the final presentation content in the early stages of video generation, such as camera movement, background music, and output ratio, are RUNWAY, PIKA, Ying AI, Dreamina, and aibrm. Among them, Dreamina is the most prominent. During the video generation process, each software has a longer waiting time. PIKA, VIDU, Ying AI and aibrm have an oppositely shorter waiting time for video generation when people input the same poem. The analysis finds of the final video product that VIDU, Dreamina, Ying AI, and aibrm can achieve a video duration of more than 5 seconds. These dimensions are interrelated and influence each other, and the software to be compared is selected together.

After comprehensive comparison, the authors have decided to choose PIKA from three foreign software and aibrm from six domestic software for a more in-depth comparison.

4.2. Questionnaire Result

4.2.1. Basic Information

There are a total of 45 teachers who teach elementary school and 7 teachers who teach middle school, with 1 teacher teaching both elementary and middle school. Among them, there are 7 respondents with teaching experience of 30 years or more, 22 with 10-29 years, 11 with 5-9 years, 10 with 5 years of work experience, and 3 intern teachers who have not officially entered the teaching profession. (Figure 1)

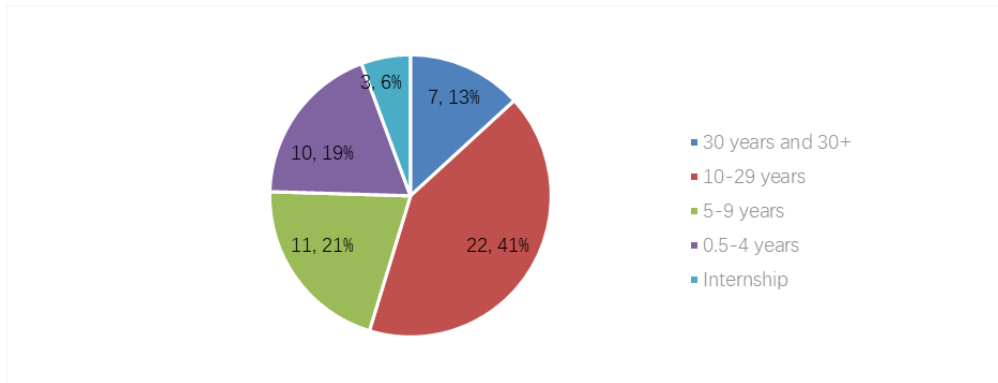


Figure 1: Analysis on the Teaching Duration of Respondents

4.2.2. Concept Images of Video

In the first part of the video artistic conception analysis of the questionnaire, more than 50% of the respondents in each group of videos agreed with the poetry videos generated by aibrm. However, the data that both videos do not conform to the poetry concept images has aroused public concern. (Figure 2) The respondents gave their own opinions on the reasons why the videos generated by the two AI tools did not conform to the concept images of poetry. The author classifies the reasons into six categories: 1. The integration of characters and situations is not enough. 2. The expression of the characters is stiff. 3. The video is not coherent enough. 4. The authenticity is not enough. 5. Emotional communication is not in place. 6. The elements in poetry are not fully reflected.

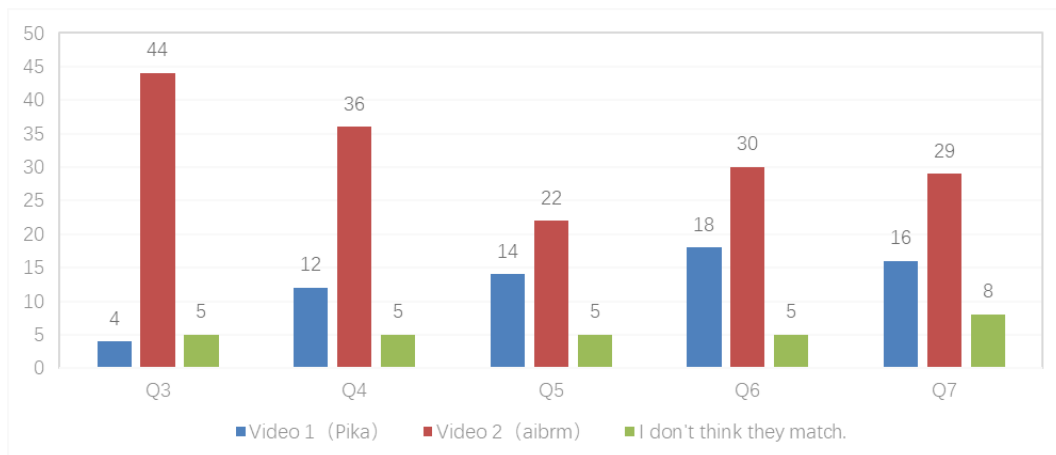


Figure 2: Videos analysis (Poem concept images)

In the second part of the survey questionnaire, the vast majority of respondents agree that poetry videos generated by aibrm tools are more representative of the designated concept images. (Figure 3) According to the respondents' opinions that the two videos do not conform to the specific images, the main reasons are classified into four categories: 1. Lack of authenticity. 2. Did not convey the specified artistic conception and emotion. 3. Some keywords such as "flow", "fall", "sorrow" and "mourning" are not reflected. 4. Lack of overall experience, such as lack of aesthetic feeling, insufficient overall presentation and strong sense of AI.

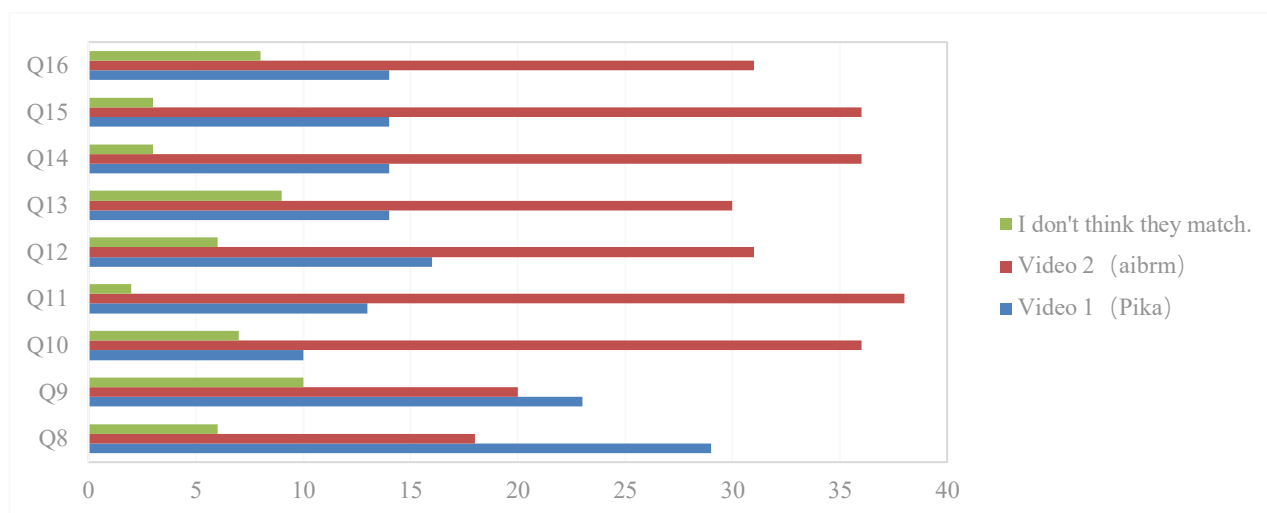


Figure 3: Videos analysis (Specific concept images)

5. Conclusion

Through an in-depth exploration of the application and potential of AIGC technology in understanding the concept images of Chinese poetry, people have successfully enabled machines to capture the rich emotions and unique aesthetic conception contained in ancient poetry to a certain extent. This study tests and verifies AI's abilities in text analysis, emotion analysis and image correlation during this process. What's more, it shows the great potential of cross-cultural communication, education popularization and cultural heritage protection.

The research demonstrated several challenges when AIGC tools are working for generating poem videos. The concept images of ancient poetry often integrate multiple elements, including natural scenery, historical allusions, personal emotions, etc., and their depth and breadth far exceed the understanding ability of current AI. Furthermore, a lack of in-depth cultural background knowledge makes AI ineffective in interpreting certain images with specific historical or regional colors. The perception and appreciation of artistic conception are highly subjective, and AI's understanding is usually based on data statistics and pattern recognition, making it difficult to fully resonate with human emotions and creative thinking.

In order to strengthen the ability of AIGC's understanding of poem concept images, related departments should strengthen interdisciplinary cooperation, combine linguistics, literature, computer science, psychology and other multidisciplinary knowledge, and build a more comprehensive and in-depth understanding framework, and enrich training data, collect and organize more diverse and high-quality ancient poetry data, especially those works that can reflect the characteristics of different eras and schools of thought. For interpreting the poetic concept images accurately in a wider cultural text context, the departments are required to enhance the generalization ability of the AI model, introduce cultural contextual information and develop AI models that can integrate cultural background knowledge.

AIGC should be applied in poetry teaching which can help learners understand and feel unplaceable concept images as well as aesthetic feelings from poetry directly and visually. At the beginning of the class, teachers can offer videos generated from different AIGC videos to give students a preliminary understanding of the artistic conception and stimulate their interest. In the mid-term of learning, under the guidance and inspiration of the teacher, students independently choose videos that are more in line with the poetic mood, and analyze the differences in groups to feel the subtle differences in mood under different expression. In the later stage of homework design, students independently think about the artistic conception contained in ancient poetry, and use AIGC to create a video of their understanding for submission, for teachers to judge whether students' mastery of the ancient poetry meets the standard. Learners take part in the decoding of poetry concept images and creation, and upgrade the learning interest and result with interactive style AIGC tools

However, this study also has limitations. The survey questionnaire was distributed online and took only 5 days. The survey was only conducted in a few cities such as Changsha, Nanchang, and Zhongshan, and the distribution format, time, and geographical sources of the respondents were not comprehensive enough.

AIGC's research on the understanding of the concept images in Chinese poetry is not only a brave exploration of current technology, but also a vision for the harmonious coexistence, common inheritance and cultural value creation of humans and computers in the future. With the continuous advancement of technology and the deepening of research, people believe that AI will inject new vitality and brilliance into ancient poetry, a treasure of human civilization, in a broader field.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Ying L., Songlin L. (2022) *Creation of Artistic Conception of Ancient Poetry from the Perspective of VR Narrative Experience*, *Industrial Engineering Design* 4 41-46.
- [2] Xiaolu Y., (2022) *The Application of Chinese Painting Imagery Expression in Digital Illustration*, *Beijing Institute Of Fashion Technology*,
- [3] Yifan C., Xinyi S., Jeanhun C. (2024) *A Feasibility Study on RUNWAY GEN-2 for Generating Realistic Style Images*, *International Journal of Internet*.
- [4] Xue D., Zhilan M. (2020) *Research on the Application of Artificial Intelligence in Middle School Chinese Poetry Teaching*, *Journal of Hubei Normal University (Philosophy and Social Sciences)* 40 105-107.
- [5] Zhen Z. (2024) *Research on Innovative Design and Visual Communication Path of Traditional Chinese Poetry Culture in the AIGC Intelligent Creation Era*, *Shoes Technology and Design* 4 88-90.
- [6] Mingda Z. (2024) *From Text to Video: The Future and Risks of Sora in Enhancing Advertising Communication Effectiveness*, *Literature and Art Development and Innovation*.
- [7] Karaarslan E., Aydın Ö. (2024) *REVIEW OF OPENAI SORA, STABLE DIFFUSION, LUMIERE AND COMPARABLE MODELS*, Karaarslan & Aydın.
- [8] Dan W. (2024) *Research on the Strategy of Artificial Intelligence Assisted Chinese Language Education and Teaching*, *HANYU CULTURE*.
- [9] Chen C., Zhuofen H., Yihui C. (2024) *Research on the Interactive Design of Virtual Reality for the Healing Space of Poetry and Artistic conception in Zhuligan*, *Packing*.
- [10] Yun D., Ang L., Jianjun Q. (2023) *Collaborative construction of artificial intelligence curriculum in primary schools*, *Journal of engineering education (Washington, D.C.)* 112 23-42.
- [11] Thomas C., Qi X., Xinyan Z., Ching C., Miaoting C. (2023) *Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education*, *Computers and Education: Artificial Intelligence*.
- [12] Darvishi A., Khosravi H., Sadiq S., C D.G., Siemens G. (2024) *Impact of AI assistance on student agency*, *COMPUT EDUC.*