Applications of Artificial Intelligence in Video Game Creation: A Case Study of Visual Novels

Authors: Hong Yipeng

Date: 2024-10-21T23:42:00+00:00

Abstract

The application of artificial intelligence technology in video game creation, especially in the domain of visual novels, has gradually become a new trend in industrial development. The utilization of AI technologies such as natural language processing, image synthesis, and sound generation significantly improves the efficiency of visual novel creation, lowers the barrier to entry, and expands creative possibilities. Furthermore, although the personalization capabilities of AI technology have driven innovation in visual novel creation methods, the application of this technology has also sparked widespread discussion regarding artistic merit, copyright, and social impact; therefore, video game creators should carefully address the potential issues arising from AI technology.

Full Text

Preamble

Application of Artificial Intelligence in Video Game Creation—Taking Visual Novels as an Example

HONG Yipeng
(Shanghai Maritime University, Shanghai 200135, China)

Abstract

The integration of artificial intelligence technologies in video game creation, particularly in the realm of visual novels, has emerged as a new trend in industry development. The use of AI techniques such as natural language processing, image synthesis, and sound generation has significantly enhanced the efficiency of creating visual novels. These technologies lower the barriers to entry and expand the possibilities for creativity. Moreover, the personalization capabilities of AI have revolutionized methods of creating visual novels. However, the application of these technologies has also sparked widespread discussions on

artistic integrity, copyright issues, and societal impacts. Consequently, video game creators must carefully address the potential issues raised by the use of AI technologies.

Keywords: artificial intelligence; video games; visual novels; art creation

0 Introduction

With the rapid advancement of artificial intelligence technology, its application in video game creation has become increasingly widespread, particularly in visual novels. Visual novels are audio-visual games characterized primarily by "visuals" and "narrative," featuring extensive text accompanied by diverse elements such as music, sound effects, animations, and videos, while possessing qualities of gameplay, visual aesthetics, and interactivity [1]. Traditional visual novel creation relies heavily on extensive script writing, manual drawing, and voice acting work. Now, however, with the aid of machine learning algorithms and rich datasets, AI can mimic artists' styles and techniques for painting creation. By employing advanced natural language processing, image synthesis, and sound generation technologies, developers can quickly and efficiently obtain various resources needed for visual novel production [2]. This not only substantially lowers the barrier to visual novel creation but also significantly enhances the productivity of the visual novel game industry.

1 Introduction to Visual Novel Creation Process

The creation process of visual novels typically involves multiple creative and technical stages. The main workflow is introduced below.

Story conception and scriptwriting mark the beginning of visual novel creation. At this stage, developers must determine the game' s theme, setting, character relationships, and main plot, establishing the tone and framework for the visual novel. Scriptwriting encompasses not only dialogue and descriptive text but also requires planning player choice points and how these choices influence story direction [3]. Once the story script is finalized, character design and background drawing commence. Character design must reflect personalities and story backgrounds while capturing player interest. Background drawing must connect closely with story settings and scenes to create visual environments that match the narrative atmosphere. This stage demands high creativity and meticulous artistic skill from artists [4]. To enhance immersion, visual novels often include character voice acting. Voice actors perform based on character personalities and plot requirements, a process demanding rigorous casting and professional recording equipment. Excellent voice acting can greatly enhance character dimensionality and story expressiveness. Music must be produced according to story emotional changes to strengthen mood and atmosphere, while sound effects enhance actions and interactions in the game, such as audio feedback for button clicks or scene transitions. Finally, programmers integrate all created elements—text, images, and audio—into the game engine through coding. This

stage includes writing game logic, processing player input, and ensuring stable operation across different platforms. Visual novels typically employ specialized engines such as TVP-KR, Ren' Py (shown in [Figure 1: see original paper]), and Artemis, which are primarily open-source, easy to learn, and support rapid development.

2 Application of AI in Visual Novel Creation

[Figure 1: see original paper] Development interface of the Ren' Py game engine

With the assistance of AI technology, certain steps in visual novel creation can now be automated or simplified using various deep learning models.

2.1 GPT Model

The GPT (Generative Pre-trained Transformer) model is a natural language processing model based on the Transformer architecture, developed by OpenAI, primarily used for text generation. The Transformer architecture is a model structure that relies on self-attention mechanisms, utilizing stacked self-attention and feed-forward networks. It comprises an encoder and a decoder. The encoder reads input data and produces a continuous representation, while the decoder uses this representation to generate output sequences. Both encoder and decoder consist of multiple identical layers, each containing self-attention layers and feed-forward neural networks that process all input data points, enabling the model to understand the relationship between each word and other words in the context [5].

Since natural language processing technology can automatically generate creative text such as storylines, dialogues, and background descriptions, developers only need to input basic plot concepts or keywords, and AI can expand them into detailed story frameworks, dramatically improving creative efficiency and breadth. This technology provides creators with unlimited inspiration and possibilities during the story conception and scriptwriting phases.

2.2 Stable Diffusion Model

Stable Diffusion is an advanced image synthesis technology that can automatically generate scene images matching textual descriptions. Its principle involves gradually adding noise to data and then learning how to reverse this process to generate data. This model consists of a forward diffusion phase and a reverse diffusion phase. In the forward diffusion phase, the model progressively introduces Gaussian noise into original data, gradually transforming it into a pure noise state through multiple time steps. In the reverse diffusion phase, the model must learn how to gradually restore clean data from the noise state. This step is accomplished by training a neural network that attempts to predict the noise to be removed at each denoising step. Through this approach, the model progressively recovers samples close to the original data from pure noise [6].

The Stable Diffusion model can not only mimic different artistic styles but also be customized according to specific requirements to create unique visual effects. Beyond static images, AI can also generate dynamic scenes and character interactions using conditional Generative Adversarial Networks (GANs), enhancing the dynamism and immersion of visual novels.

2.3 VITS Model

VITS (Variational Inference with adversarial learning for end-to-end Text-to-Speech) is an end-to-end text-to-speech (TTS) model designed to directly generate speech waveforms from text, optimizing the multiple steps in traditional TTS systems into a unified model. The VITS model primarily comprises three key components: a text encoder, a generator (vocoder), and a discriminator. The text encoder converts input text into latent feature representations. Based on these features, the generator produces mel-spectrograms of speech through a Variational Autoencoder (VAE) structure, which are then converted into speech waveforms [7]. During this process, the model employs adversarial learning, where the discriminator attempts to distinguish generated speech from real speech, while the generator tries to fool the discriminator, making the generated speech as close to real speech as possible.

The VITS model can automatically generate natural and fluent speech from text, significantly reducing the manual recording and editing time required in traditional voice acting processes for visual novel creation, thereby substantially improving production speed. Simultaneously, compared to hiring professional voice actors and recording equipment, using the VITS model can significantly lower production costs, which is particularly beneficial for developers with limited budgets, enabling them to achieve high-quality voice acting at lower costs.

3 Implementation Cases of AI in Visual Novel Creation

Taking the "Visual Novel Generation System" as an implementation case, this system integrates multiple AI technologies to support automated visual novel creation. The framework design diagram is shown in [Figure 2: see original paper], with the main functional modules using AI being image resource generation and voice resource generation.

3.1 Visual Novel Image Resource Generation

[Figure 2: see original paper] System framework design diagram

When descriptive text is input, such as detailed scene descriptions or specific character feature explanations, the system first uses natural language processing technology to parse the text and extract key information and features. These extracted key elements are then passed as input to an image synthesizer based on the Stable Diffusion model. During this process, the system uses multiple preset and user-defined keywords to guide the image generation process. For example,

when the description "A tranquil lakeside, the afterglow of sunset shimmering on gently rippling water" is input, the system first extracts keywords such as "tranquil," "lakeside," "sunset," "afterglow," and "rippling water." These words are then used to guide the Diffusion model to focus on rendering these features during image generation. The final image resource generation result is shown in [Figure 3: see original paper], with generated resource files shown in [Figure 4: see original paper].

[Figure 3: see original paper] Image resource generation result

3.2 Voice Resource Generation

[Figure 4: see original paper] Image resource files

When character dialogue text is input, with specific voice character names selected and parameters such as speech rate and pitch set, the system employs the VITS model to output high-quality synthesized speech files, which serve as part of the visual work to provide voices for characters. For instance, to create a voice for a character, text such as "Wow! You look so beautiful today!" is input, and a cheerful tone with medium speech rate is selected. The system processes these inputs and generates a voice file that sounds cheerful and natural. The final voice resource generation result is shown in [Figure 5: see original paper], with generated resource files shown in [Figure 6: see original paper].

[Figure 5: see original paper] Voice resource generation result

Finally, by integrating these resources into the Ren' Py game engine, a visual novel game can be simply developed, as shown in [Figure 7: see original paper].

[Figure 6: see original paper] Voice resource files

4 Development and Impact of AI in Visual Novel Creation

[Figure 7: see original paper] Visual novel game interface

The integration of AI technology with visual novel creation has not only transformed the traditional production workflow but also exerted profound influences on industry development trends, creator roles, and user experience.

First, AI has significantly improved visual novel production efficiency. By automatically generating text, image, and audio content, AI reduces the most time-consuming parts of traditional creation, greatly shortening project cycles. Additionally, AI technology lowers creative barriers, allowing more creators without professional skill backgrounds to enter this field. For example, authors not proficient in drawing or music production can now use AI tools to generate such content, enabling individuals or small teams to independently complete complex visual novel projects.

Second, the introduction of AI into visual novel creation has greatly expanded creative possibilities. Using AI models such as GPT and Stable Diffusion, devel-

opers can generate rich and diverse creative content including text and images based solely on simple prompts. This technology encourages creators to explore new narrative techniques and visual expression methods, pushing the artistic and technical boundaries of visual novels.

Third, AI' s personalization capabilities have also brought new forms to visual novels, driving innovation in creative styles and narrative methods. AI can analyze user choices and interaction history to adjust story development directions, achieving dynamic changes in story content. This technology can not only adjust story details and endings according to user preferences but also introduce new branches generated based on user interests within the story, making each reading experience unique. This enhanced personalization not only improves user immersion but also increases user satisfaction and work replayability. Users feel that their choices truly influence story direction, a sense of participation difficult to achieve in traditional visual novels.

Fourth, although AI demonstrates excellent performance in improving efficiency and reducing costs, its application in artistic creation has also sparked discussions about the perception of "true artistry." While AI-generated content is precise and efficient, it may lack the emotional depth and originality of human artists. Therefore, future visual novel creation may tend toward a human-AI collaboration model, where AI is responsible for generating large amounts of creative material, and human artists perform aesthetic selection and emotional processing on this foundation [8, 9].

Fifth, with the widespread application of AI technology in visual novels, the market may see a large number of low-cost, rapidly produced works, which will change the market dynamics of game and narrative content. On one hand, this enables users to access more diverse works [10]; on the other hand, it may also lead to market oversaturation, affecting content quality and originality. Additionally, copyright and intellectual property issues arise, particularly when AI uses existing artistic and literary works without authorization to create new content.

Sixth, the use of AI in visual novel creation raises legal considerations, especially regarding intellectual property and copyright [11]. For example, AI-generated characters and stories may inadvertently imitate or use existing literary works, film characters, or public figures, creating risks of copyright infringement. Therefore, when using AI technology, it is essential to ensure all content is generated within legal and ethical boundaries to avoid infringing on copyright or other intellectual property rights.

Seventh, AI content creation may bring ethical, moral, and social impacts. When using AI technology, it is necessary to ensure generated content does not contain discriminatory, insulting, or other inappropriate information to protect consumers and the public from potential negative effects. Consequently, establishing clear usage guidelines and conducting strict review and supervision of AI-created visual novels are also key to ensuring the healthy development of AI

technology and visual novel creation.

The application of AI in video game creation, particularly in the visual novel domain, provides developers with unprecedented tools and methods, greatly improving creative efficiency and innovation. This technology promotes the fusion of game art and technology, expanding the expressive forms and narrative techniques of visual novels. However, we must also recognize that AI cannot replace traditional video game creation tools but rather serves as a complementary force. Only through appropriate use and management can it become an important tool for promoting creative expression and multicultural development.

References

[1] Li Bin. Visual Culture and Game Characteristics: Story Construction and Future Development of Visual Novels. Urban Cultural Studies, 2023, (02): 485-495.

[2] Sun Weiping, Liu Hangyu. Generative Artificial Intelligence and the Revolution in Knowledge Production Modes—Starting from the Shockwave of Sora. Ideological and Theoretical Education, 2024(05): 12-18. DOI:10.16075/j.cnki.cn31-1220/g4.2024.05.001.

[3] Wang Ruocun. Interactive Reading Games as Hypertext Art: Focusing on "Cheng Guang" . Journal of Aesthetic Education, 2024, 15(02): 97-105. DOI:10.20073/j.cnki.2095-0012.2024.02.013.

[4] Lu Chunxiu. The Fusion of Literature and Games: A Narrative Study of Visual Novels. Dissertation. Northeast Normal University, 2023. DOI:10.27011/d.cnki.gdbsu.2023.000658.

[5] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 6000–6010.

[6] Chen Ying, Ma Hongtao. AIGC as a Powerful Assistant in the Field of Art and Design—Taking Stable Diffusion as an Example. Fashion Designer, 2024, (01): 73-84. DOI:10.20100/j.cnki.cn11-4548/ts.2024.01.008.

[7] Wang Jiawen, Gao Dingguo, Ni Qiong, et al. Research on Tibetan Kham Dialect Speech Synthesis Based on the VITS Model. Computer Knowledge and Technology, 2024, 20(04): 8-10+15. DOI:10.14004/j.cnki.ckt.2024.0226.

[8] Chen Yongwei. Beyond ChatGPT: Opportunities, Risks, and Challenges of Generative AI. Journal of Shandong University (Philosophy and Social Sciences Edition), 2023, (03): 127-143. DOI:10.19836/j.cnki.37-1100/c.2023.03.012.

[9] Yang Minran, Zhang Xinxing, Tao Rongxiang. Current Status and Trends: A Perspective on Domestic Artificial Intelligence Generated Content (AIGC) Research. Library Theory and Practice, 2024, (02): 56-65. DOI:10.14064/j.cnki.issn1005-8214.20231225.001.

[10] Wang Nuo, Bi Xuecheng, Xu Xin. Sharpening the Tools First: AIGC and Its GLAM Application Opportunities in the Metaverse Scenario. Library Tribune, 2023, 43(02): 117-124.

[11] He Zhe, Zeng Runxi, Qin Wei, et al. The Social Impact and Governance of New Generation AI Technologies such as ChatGPT. E-Government, 2023, (04): 2-24. DOI:10.16582/j.cnki.dzzw.2023.04.001.

Funding: Shanghai College Student Innovation and Entrepreneurship Training Program Project (S20240601)
Author: Hong Yipeng (2003–), male, Han ethnicity, Shanghai native, undergraduate student, research interests: computer science and technology, artificial intelligence

Note: Figure translations are in progress. See original paper for figures.