

Design of AIGC Commercial Digital Human Video Generation Link

Yingyin Huang^{1,a}, Yunzhen Ye^{1,b,*}

¹*School of Design and Architecture, Zhejiang University of Technology, Liuhe Road, Hangzhou, China*

a. 1070065080@qq.com, b. jenniferheydrich311@gmail.com

**corresponding author*

Abstract: This study aims to explore and analyze the application of AIGC technology in digital human video creation, in order to achieve the goal of efficient conversion generation and personalized content production. This study adopted a systematic analysis method to integrate and apply various technologies of AIGC. On this basis, an innovative digital human video generation link was designed by combining deep learning technology and multimodal data fusion methods. Taking commercial digital human image videos as an example, this study combines the digital human image features favored by target users with AIGC technology to create generative digital human images, and summarizes an innovative digital human video generation chain. The design and application of the AIGC digital human video generation chain provide new impetus for the development of the digital human industry. By continuously exploring and optimizing the application of AIGC technology in digital human video creation, designers will create richer, more vivid, and personalized digital human video content in the future, promoting the further development and application of digital human technology in commercial promotion and media communication.

Keywords: AIGC, digital people, video generation, link design.

1. Introduction

According to the 2023 Global Virtual Digital Human Industry Report, the number and investment scale of digital human enterprises continue to grow, driving the rapid development of the digital human industry. With the development of technology companies such as ChatMindAi and DeepMind, the application ecosystem of AIGC technology is gradually taking shape, and its application scope and influence are significantly expanding in multiple fields, especially in the field of digital human generation links, which are gradually becoming the focus. [1]. In commercial promotion, the application of digital humans significantly improves user experience and brand affinity, and can carry more multi-dimensional service tasks. The generation link can ensure the stability and consistency of digital humans and quickly respond to different scenarios' needs. AIGC technologies such as Midjourney and Stable Diffusion enable designers and developers to quickly generate high-quality digital human images through link optimization, providing personalized and intelligent solutions and enhancing technical support for digital human video creation.

This study focuses on the innovative design path of integrating AIGC technology with digital human video, aiming to explore new forms and applications to improve the quality of digital human

video content. The research will include the following steps: analyzing the development status of existing AIGC technology and digital human videos, and sorting out their advantages and disadvantages in content production, dissemination, and user experience; Explore technology integration, new interactive methods, and video creation; Research on how to use AIGC technology to enhance the authenticity, artistry, and personalization of digital human videos; Build an experimental platform for production and testing, optimize fusion schemes based on experimental results, and improve content quality.

2. Current Status of Digital Human Development

Digital humans refer to virtual entities constructed using digital technology that possess human characteristics and can be created through techniques such as computer graphics rendering, motion capture, deep learning, and speech synthesis. With the continuous influx of capital into this field, digital human technology has developed rapidly, and its application scenarios have become rich and diversified. The development of digital humans can be traced back to 1982, when the world's first digital human, Lin Mingmei, appeared in "Space Fortress", marking the beginning of people's attempts to combine digital humans with real life. In the 21st century, China has also begun to conduct research in the field of digital humans. Zhejiang TV in China has launched "Gu Xiaoyu", the National Museum has launched "Ai Wenwen", and Zhonghua Book Company has launched "Su Dongpo". These digital humans showcase the increasingly mature status of China's digital human industry. At the same time, digital humans are widely used in the business field, such as AYAYI becoming the first digital employee of Alibaba Group, serving as the digital manager of Tmall brand, and establishing cooperative relationships with internationally renowned brands such as Louis Vuitton and Guerlain. The above cases fully demonstrate the enormous potential of digital humans in business applications. [2]. Under the development trend of AIGC, digital humans are moving towards intelligent production to simplify the production process and shorten the cycle, thus being applied more quickly in the commercial field. For example, in 2023, Tencent launched "Tencent Zhiying", which directly provides digital human creation functions.

Overall, digital human technology is advancing day by day, bringing new opportunities and challenges to multiple fields. With the increasingly widespread application of digital humans in daily life, it will attract more and more companies and scholars to study the generation paths and achievements of digital humans, in order to promote the sustainable development of digital human generation technology.

3. Research Status of AIGC Video Generation Technology

AIGC is a way of automatically generating content through artificial intelligence technology, leading a digital revolution with the development of generative adversarial networks and large-scale pre trained models. Major enterprises have begun to adopt AIGC to automatically generate video and audio content, greatly improving the efficiency of enterprise creation. [3]. At present, AIGC's applications in the video field are mainly divided into three types of algorithms: text to video, image to video, and video to video. The advancement of algorithms has greatly expanded the possibilities of enterprise video content production, improved creative efficiency, and reduced costs. Therefore, AIGC video generation technology is receiving close attention from many well-known companies. For example, Runway provides a cloud based video editing and generation platform, where users can easily create video content through APIs or web interfaces. In China, tech giants such as Tencent, Baidu, and Alibaba are actively developing AIGC video generation technology and have launched multiple related platforms, such as PixVerse, Vidu, and Kling. These platforms utilize AIGC

technology to provide users with a new way of producing video content and foreshadow the main direction of future video production.

AIGC video generation technology has injected new vitality into design and become one of the hotspots in video production. For example, the movie "Under Strangers" combined live action performances and artificial intelligence style transfer technology to produce some animations, significantly improving the efficiency and quality of film production. The 2023 Oscar winning film Everything Everywhere All at Once also uses the AIGC technology of the Runway platform to achieve the video slowdown effect and the production of unlimited extension pictures, thus greatly enhancing the visual effects of the film. With the continuous advancement of multimodal AI technology, AIGC technology has been able to participate in multiple aspects of film and television production, including automatically generating scripts, creating visual effects, and so on. The combination of AIGC technology and digital humans not only endows digital humans with more "vitality", but also makes them more realistic and vivid in various application scenarios. This conversion technology from static images to dynamic videos further promotes the development and application of digital human technology.

4. Design Ideas for Digital Human Video Generation with AIGC Intervention

4.1. Analysis of Video Generation Issues with Digital People

In the current AIGC field, digital human video generation platforms such as Synthesia, Realusion, and Did have become innovative technologies that allow users to create customized digital humans. In addition, GitHub provides open-source models and code for designers to conduct secondary development; TurboSquid and Sketchfab provide a large number of 3D digital human models for users to purchase or download. However, these platforms have the following limitations in practical applications.

4.1.1. Lack of Image and Personalization

Most digital human video generation platforms provide relatively single images, lacking sufficient personalization and diversity. This may be due to limitations in algorithms and preset templates, resulting in a uniform digital human image that is difficult to meet users' needs for personalization and specialization. In addition, the customization functions provided by the platform are limited, making it difficult to meet users' specific customization needs, thereby limiting the personalized development of digital humans.

4.1.2. Limitations of Interaction and Expression

The actions and expressions of existing digital humans in videos are usually relatively simple, lacking multi angle images and rich action ranges. This not only limits the emotional transmission ability of digital humans, but also affects the authenticity of interaction between users and digital humans. In practical applications, users hope to interact with more expressive and realistic digital humans, but current platforms have not fully achieved this goal.

4.1.3. Scene Fusion Problem

The integration of digital humans and video scenes is an important factor affecting visual effects. At present, many digital human video generation platforms have shortcomings in scene fusion, resulting in digital humans appearing stiff and unnatural in videos. This is usually due to insufficient interaction between the digital person and the background content, as well as a lack of smooth transition processing between the scene and the digital person.

4.2. AIGC Video Generation Scheme

The research aims to explore the application of AIGC technology in assisting designers in creating virtual human videos, and introduce multimodal thinking to enhance the performance of AIGC technology in virtual human video production. In order to improve various generation problems of digital humans, it is necessary to re-examine and optimize the video generation chain of digital humans. In the process of designing co creation videos for AIGC, the following aspects need to be comprehensively considered.

4.2.1. Evaluation of Generative Narrative Scripts

Its evaluation dimensions include factors such as attractiveness, emotional influence, technical quality, role shaping, and dialogue quality. An excellent script should be able to attract the attention of the target audience, while generating positive impact through emotional resonance, and can achieve richer plot scripts and character performances through natural language processing and computer vision technology, achieving high-quality presentation effects.

4.2.2. Evaluation of Generative Digital Human Images

Its evaluation includes aspects such as image, mouth shape, facial expressions, movements, and character consistency. The image design of virtual digital humans needs to conform to character settings, with natural mouth shapes and expressions, smooth movements, and consistent with the character's personality. In addition to the designer's self-evaluation, a user feedback mechanism can also be introduced to use multimodal sentiment analysis tools to identify the audience's emotional reactions to different scenes, characters, and conversations during the viewing process in real time, and make timely strategic adjustments.

4.2.3. Generate Evaluation of The Scene

Establish a standardized evaluation system that covers aspects such as composition, texture, theme consistency, and emotional communication. To ensure that the scene design has good structure, authentic texture, and interesting interaction, which can visually align with the video theme and effectively convey emotions. [4].

4.2.4. Other Details

Consider other details such as dubbing, subtitles, special effects, and market positioning. Dubbing and subtitling should match the content and style of the video, and special effects should enhance the visual experience, making the video production meet the basic needs of the market. At the same time, attention should be paid to the video content of competitors, identifying differentiated market opportunities, and optimizing the video format and style based on the audience characteristics of different platforms.

4.3. AIGC Video Generation Technology Flow

The creation of short videos for digital humans should have memorable points. Taking individual merchants as an example, the design of commercial digital human promotion videos mainly covers the following three key aspects. (Figure 1)

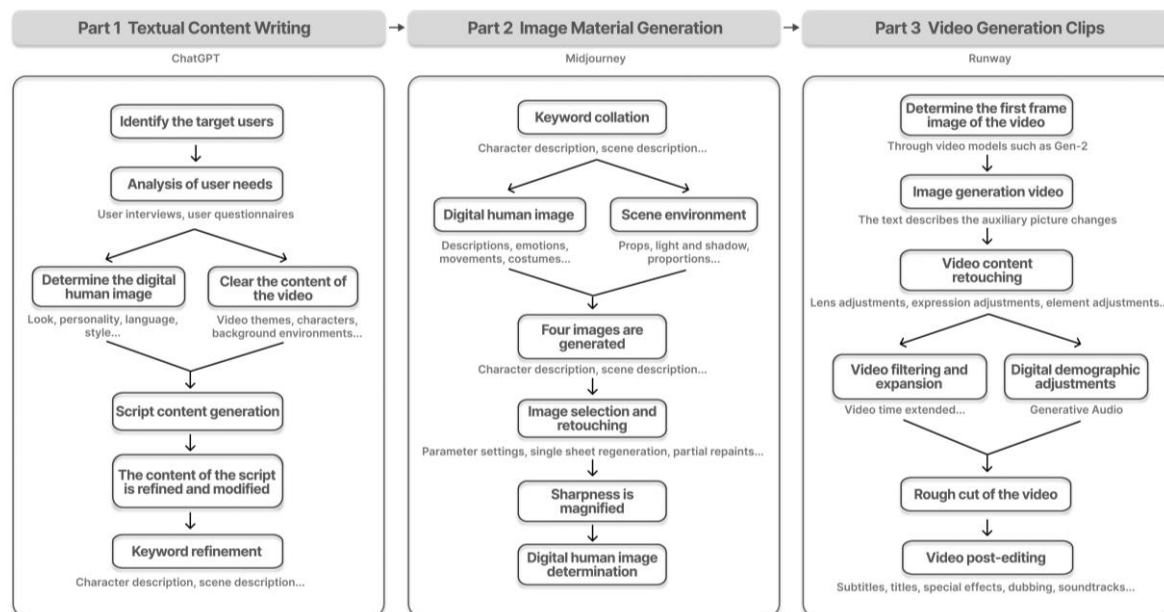


Figure 1: AIGC video generation technology flow.

4.3.1. Conceptual Conception and Plot Script Creation

Identify target users and analyze their needs, clarify the theme, style, etc. of the video clip, and ensure that the video content is consistent with the brand image and market positioning. Use ChatMindAi to quickly generate initial script drafts and refine the content, expression, wording, and other aspects of the script according to actual situations. At the same time, extract keywords based on script requirements to provide guidance for image material generation and post production special effects design.

4.3.2. Image Material and Visual Effect Generation

After organizing the keywords extracted from the script, multiple static image groups with character images are generated using Midjourney and Stable Diffusion. Through meticulous manual screening and fine-tuning of details, the images are further improved to visualize the script content until the image materials are determined and the clarity is increased to improve the image quality, providing rich visual elements for video production.

4.3.3. Post Optimization and Synthesis Video Processing

Using video generation platforms such as Runway, adjust visual effects such as image content, camera movement, and lighting according to needs to make the video more attractive. This step requires a close integration of the plot and brand appeal to ensure the creativity and practicality of the visual effects. Manually splice the video clips generated by AI, re edit them according to the plot development and logical order, and complete the rough cut. Subsequently, software such as Jianying was used to add elements such as video music, dubbing, and subtitles to enhance the video's appeal and information transmission effect, ultimately exporting the complete video.

5. Generation of Commercial Digital Human Image Videos Using Individual Merchants as An Example

By researching the digital character images and video types favored by relevant users, specific information such as video themes, digital character images, environmental scenes, dubbing and music can be determined. Taking individual merchants as an example, the preferred digital persona of users is usually set between the ages of 30-50, which conveys a mature, professional, and trustworthy image. In terms of scene selection, users prefer scenes that are related to their daily living environment, such as street shops. Therefore, video scripts need to be more realistic, using humorous language and delicate emotions to showcase the diverse needs of daily life. This type of content creation can better meet the promotion needs of commercial digital human videos. By innovating AIGC's digital human generation process, we can more effectively control the design requirements and video animation content of digital human images, thereby solving problems such as homogenization and rigid expression of digital human images.

5.1. Generation of Video Script Content

When using ChatMindAi for video script design, designers first need to determine the digital character information and theme, and then generate the complete script.

Firstly, designers can use ChatMindAi to generate corresponding user interview outlines and questionnaires to collect target users' preferences for digital human images, including appearance, personality, language, scene, and style, and further understand users' preferences for digital human video styles, such as brand image, news hotspots, real-life cases, or creative plots. (Figure 2) This user driven creative process ensures that the video content and presentation can truly resonate with the target audience from the beginning, and the resulting digital human videos will be more in line with the needs of the target users. Secondly, the designer will integrate the results of user research, clarify the theme and type of the video, such as attracting the target audience through specific creative storylines. After clarifying the core content information of the video, input the required video theme, character image, background environment, and script outline into ChatMindAi to efficiently generate corresponding and detailed script content. After receiving the script, it is crucial to review and optimize it, as this is a necessary step to ensure its coherence, accuracy, and alignment with the predetermined goals. Designers need to fine tune or further assist in writing the unreasonable content generated by ChatMindAi, flexibly adjust it according to actual needs, and ultimately determine the script. The designer will extract character images and key words for each shot based on the final script after renovation, and begin preparing to use AI tools to generate scene images for the subsequent video production stage. (Figure 3)

Through the above process, the design of video scripts can not only meet user preferences, but also have a high degree of creativity and professionalism, avoiding the problems of monotony or mismatch that may occur when directly applying models.

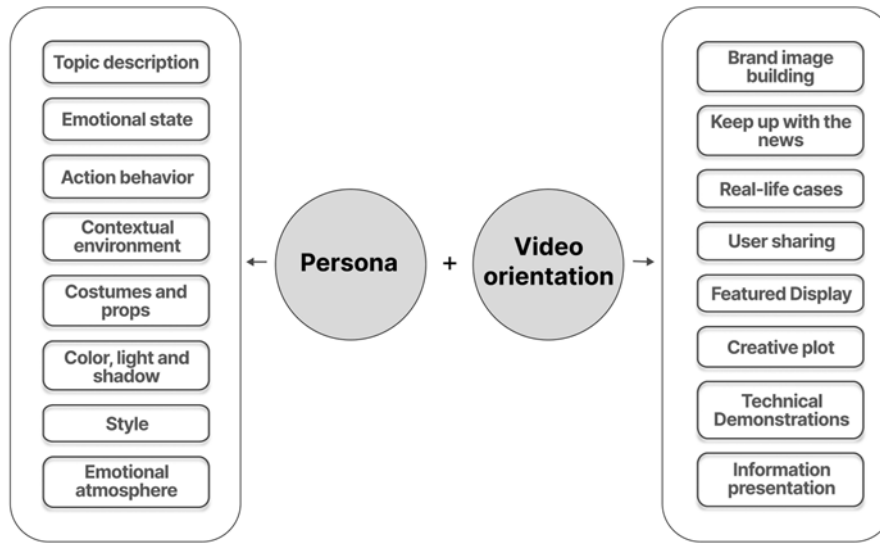


Figure 2: Positioning digital human image and video direction.

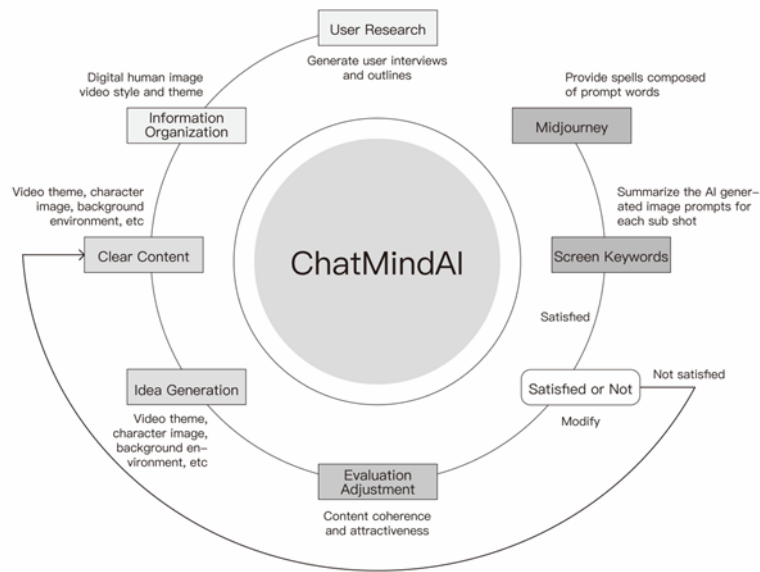


Figure 3: Flow chart for generating video script content.

5.2. Generation and Screening of Digital Human Images

In the process of generating digital human images, the designer used the Midjourney tool to assist in the creation of renderings. The entire process follows a logical sequence and is divided into four steps: keyword extraction, image generation, adjustment and optimization, and clarity enhancement.

In the early stages of generation, designers use ChatMindAi to extract keywords from video scripts and convert them into prompt text. These prompt words provide detailed descriptions of the user's desired image, including subject description, emotional state, action behavior, environmental background, clothing props, color and lighting, style and perspective parameters, and other specific aspects. Compared to directly applying existing digital models, our team can achieve more accurate personalized customization to ensure that the generated image accurately reflects the key elements in

the script. During the generation process, the extracted keywords are input into Midjourney, which generates a set of four images based on these keywords and instructions. Based on obtaining diverse images, designers manually screen and select the most suitable one in terms of effect and style from four works, and flexibly further optimize it. This process has a high degree of automation, but in order to ensure that image generation meets script requirements, manual intervention (such as image filtering and local adjustment) ensures the accuracy and artistic quality of the generated images. After selecting an image, adjustments should be made to it, including using techniques such as parameter settings, single reproduction, and local redrawing to repair unrealistic or inaccurate local content in the image. If there is still deviation in the adjusted image content, return to the first step, modify the prompt words, and regenerate the image. The introduction of this backtracking mechanism is also to further ensure that the generated results can achieve the expected effect. Once the digital human image is determined, it is subjected to clarity amplification processing to ensure that high-quality professional standards can be maintained in AI generated video effects, presenting the best visual effects and avoiding situations similar to traditional models with low clarity.

In this process, the designer efficiently utilized the Midjourney tool to assist in generating digital human images that conform to the script description. This not only ensured that the image quality and overall video effect reached a professional level, but also improved customization, creativity, and controllability, avoiding the common problems of rigidity and singularity in traditional model use, and better meeting the needs of script and video production. (Figure 4)

	Select the image used to generate the video	Eliminated images
Screen comparison		
Conclusion	By determining the color and lighting details of the image through text, the generated image content is of higher quality, more attractive, and has significant differences	The image generated after a general description lacks strong expressive power, and the character's personality is not prominent, making it difficult to form a deep impression

Figure 4: Midjourney Image Generation Case.

5.3. Digital Human Video Material Generation

When using the generative AI video platform Runway to create digital human videos, designers can optimize the generation process through the following aspects to achieve high-quality and effective videos. Firstly, select the already created digital human image as the starting point for video generation and set it as the first frame of the video. To ensure the rationality of the video, Gen-2's video model can be used to control the content generated by the video, making the video smoother and visually better. Secondly, textual descriptions can be used to guide the changes in the video frame, such as Runway's mural painting tool, which can generate specified expressions for digital faces based on the text content, accurately reflecting the emotional changes of the characters. At the same time, text descriptions can also add new elements or character actions, making the video content richer, more interactive, and more in line with the script's needs. In addition, designers can determine the camera's motion effect and image texture by adjusting parameters. Precise control of parameters can ensure the dynamic and smooth feeling in the image, enhance the viewing quality of the video, and strengthen the depth and texture of the video, making the generated content more attractive. In the video filtering and expansion stage, a series of video content will be continuously generated, and

segments that meet the script requirements will be selected from them, avoiding the monotony of relying on fixed preset models. For the requirement of lens length, if the original video clip is too short, the 4-second video can be extended to 8 seconds to meet the duration requirements of specific scenes, ensuring the coherence and rhythm of the video content. Of course, the Generative Audio feature of Runway can also be used to create a synchronized mouth shape effect for digital humans, which is difficult to achieve in traditional preset models. In Runway, as long as the character dialogues in the script and the pre generated character videos are uploaded and appropriate sound templates are selected, the dialogue content in the digital human video can be efficiently generated, ensuring synchronous and natural flow of sound and image. (Figure 5)

Through these processes, using the generative AI video platform Runway to generate digital human videos is not only more efficient, but also has stronger customization capabilities, creative space, production efficiency, and quality control. The personalized generation process and diverse content make video creation more flexible and able to quickly adapt to different needs.

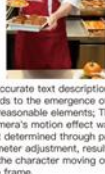
Contrast	After optimizing the process path				Before optimizing the process path	Conclusion
Project	Case 1	Case 2	Case 3	Case 4	Case 5	
Generate Results						After modularizing the description of the background, character movements, and expressions through text, the generated video content is more logically sound. Video content generated from chaotic text descriptions is prone to illogical visual content. The lack of detailed textual descriptions of the character's facial expressions and hand movements resulted in the generated digital person facing down and having illogical hand movements. Inaccurate text description leads to the emergence of new unreasonable elements; The camera's motion effect was not determined through parameter adjustment, resulting in the character moving out of the frame.
						
						

Figure 5: Runway video generation case.

5.4. Post Production Editing of Digital Human Videos

After generating all the content specified in the script, post-processing with video editing software such as Jianying can further ensure that the final output video meets the expected quality and effect.

The first step is to manually stitch the AI generated segments of the video based on the content of the storyboard, re edit them according to the story logic, and complete the preliminary editing, namely rough editing. This stage ensures the coherence of the video content and the fluency of the story, enabling the video to effectively convey the core information in the script. Then, using the intelligent subtitle function, generate subtitles for the dialogues in the script and accurately match them with the corresponding video frames. At the same time, titles, brand slogans, etc. can be added to the video for text information annotation, which will significantly improve the clarity of information communication and make the video more professional. In addition, due to the possibility that some videos in the process of generating digital human videos may not have digital population production, it is necessary to regenerate dubbing for these videos and turn off the voice content of the previously

generated videos. Choose appropriate tones, such as dialects, for reading aloud, while adding video background music and special effects to complete the dubbing and music work of the video, enhancing its infectiousness and emotional expression. Finally, add video effects, transitions, and acceleration functions as needed to optimize and improve the digital human video in the final stage. These post-processing techniques can be said to make up for the shortcomings of traditional fixed model digital humans, such as flexible narrative structure adjustment, emotional dubbing sound effects, strong visual impact, and precise control of details. They provide greater flexibility, customization, and creative space, making the final video significantly better in visual effects and viewing experience than videos using model digital humans.

Through the above steps, the designer has completed the post editing work of the digital human video, ensuring the quality of the digital human short video placement, further enhancing the attractiveness of the video, and improving the viewing experience of the audience.

5.5. Post Optimization and Synthesis Video Processing

The biggest advantage of AIGC tools is their amazing generation efficiency, which allows designers to easily create complex digital human images. In the past, the production of digital human models was a time-consuming and labor-intensive process that required designers to invest a significant amount of time and effort. However, as designers begin to utilize AIGC technology to accelerate the speed of digital human design, it not only improves the efficiency of design teams in producing design projects, but also unleashes the creativity of designers. [5]. By transforming the AI design process, the problem of homogenization and low authenticity in the generation of digital human images has been effectively solved. This efficient and meticulous generation method not only improves the quality of design, but also subverts the traditional creative process of designers, bringing revolutionary changes to the creative generation of works. (Figure 6)

Designers can now devote more energy to direction thinking and detail polishing of design proposals, rather than getting stuck in tedious production work. The emergence of AIGC tools enables designers to better balance creativity and execution, creating more innovative and personalized digital human works.

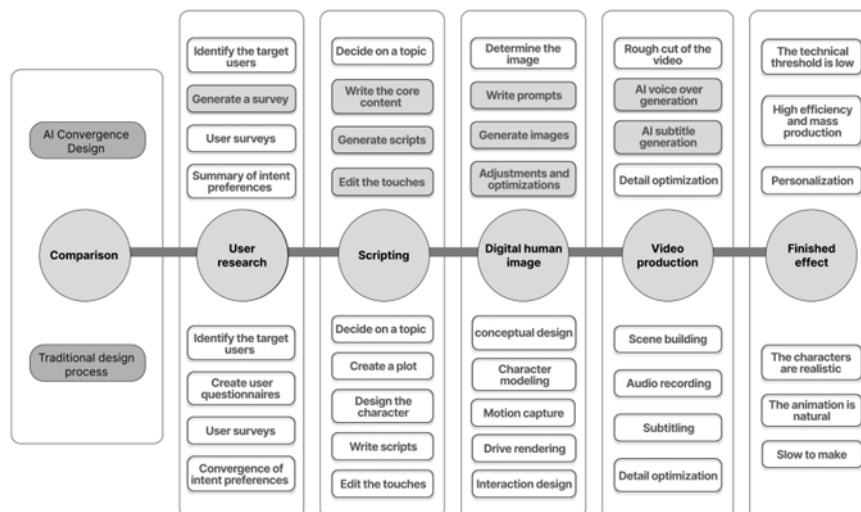


Figure 6: Comparison between Traditional Design Process and AI Fusion Design Process.

6. Conclusion

In this study, the article constructs an innovative design path that integrates digital humans and AIGC technology. This path explores how decentralization enables users to control their own data in the Web3.0 era, with an increasing emphasis on reducing dependence on intermediaries. AI, With the development of technologies such as the Internet of Things and virtual reality combined with Web3.0, it can promote richer user experiences and application scenarios. AIGC will provide a revolutionary approach for digital human design, which not only accelerates the design process but also helps designers explore new design concepts and styles. In the future, the application of AIGC technology in the field of digital human design will make the image of digital humans more diverse, personalized, and realistic, providing users with richer and deeper interactive experiences.

References

- [1] Zhaolin L, Xinheng S, Yucheng J. (2023) *Current Status and Development of Intelligent Design under AIGC Technology Trends*. *Packaging Engineering*, 44(24), 18-33+13.
- [2] Siqi C, Guoming Y, Jiayi Y, et al. (2022) *Virtual Digital Humans: An Experiential Medium—An Analysis of the Connection Mechanism and Media Attributes of Virtual Digital Humans*. *Journalism*, 7, 12-23.
- [3] Yu S, Xiaosong Q, Lingyu P, et al. (2023) *Research on AIGC Chinese Aesthetic Cultural and Creative Design*. *Packaging Engineering*, 44(24), 1-8+33.
- [4] Jian Z, Yuxin W, Zhe Y. (2023) *AIGC Empowering Traditional Culture Inheritance Design Methods and Practices: A Case Study of the Digital Exhibition Center Design of Yongle Palace in Shanxi Province*. *Design*, 36(17), 30-33.
- [5] Yuquan F. (2024) *Applications and Reflections of AIGC in Industrial Design*. *Packaging Engineering*, 45(08), 337-345.