

# Ethical Risk and Pathway of AIGC Cross-Modal Content Generation Technology

Ling Jiang & Yiting Zhang

College of Publishing, University of Shanghai for Science and Technology, China.



## Abstract

This study analyses the core technologies underlying AI-generated cross-modal content (AIGC), identifying data, algorithms, and computing power as the fundamental pillars supporting AIGC operation. And data are recognized as the underlying logic driving AI's continuous development and the source of ethical issues within AIGC. By integrating Gilbert Hottois' concept of technological accompaniment, this research incorporates multiple stakeholders to dissolve the binary opposition between humans and machines. This study explores pathways to scientifically and positively advance AIGC technologies at the micro, medium, and macro levels. It advocates for human-machine symbiosis, enhances the frequency and potential of users' digital interactions, improves their understanding and autonomy in applications, and promotes digital literacy in the intelligent era. Additionally, it emphasizes the importance of government-led initiatives and global dialog to establish a multi-stakeholder regulatory framework and conventions, aiming to create a more harmonious human-machine community with a shared future.

**Keywords:** Human-Machine Communication; AIGC Cross-Modal Content Generation Technology; Technology Accompaniment; Human-Machine Community with a Shared Future.

## I. Introduction

As artificial intelligence technology continues to innovate, breakthroughs in the field of content production have been achieved. AI is poised to become the leading technology for future world economies and efficient content production. The focus on artificial intelligence aims to promote the evolution of content production methods in intelligent, cross-modal, autonomous, and sustainable directions, making it a significant contemporary research topic. In a report of the 20th National Congress of the Communist Party of China, Secretary Xi Jinping emphasized the strategic importance of the digital economy: "Accelerate the development of the digital economy, promote the deep integration of the digital economy and the real economy, and build internationally competitive digital industry clusters."

In the evolution of human-machine communication, the advancement of artificial intelligence-generated content (AIGC) has shifted it from being a traditional communication intermediary to a communication collaborator, with content production moving toward cross-modality. It is crucial to examine whether AIGC has entered an autonomous phase, upheld correct values, and explored how humans and machines can coexist harmoniously. However, current cases and the use of AIGC tools reveal a host of ethical issues, including disputes over content production copyrights, the emergence of AI filter bubbles in the intelligent era, the spread of false information, and the widening of the digital divide.

In this "technological carnival," it is essential to look forward to clarifying the development process of AIGC, the core technologies of cross-modal content generation, and the ethical issues they entail. Simultaneously, we must look backward, deeply analyzing from the perspective of "technology accompaniment" to guide and drive the scientific and positive development of AIGC cross-modal content generation technology, with the goal of building a more harmonious human-machine community with a shared future.

## Evolution of AIGC in Human-Machine Communication What is AIGC?

Artificial intelligence-generated content (AIGC) refers to a new content production method that uses artificial intelligence technology (generative AI paths) to generate content. Broadly speaking, AIGC encompasses all the content generated through AI technology. From a conceptual analysis perspective, AIGC differs from generative AI and artificial intelligence synthetic media in that it autonomously completes the innovation process. This includes partial generation on the basis of key strategic clues, underlying understanding and fusion generation on the basis of multimodal content, and feature generation on the basis of comprehensive or segmented scenarios. From the

perspective of content characteristics, it is "a type of content classified from the perspective of content producers, a method of content production, and a collection of technologies for automated content generation".

Driven by intelligent technology, data, algorithms, and computing power form the cornerstone supporting the normal operation of AIGC, emerging as the primary content production method following PGC, UGC, and PUGC. Relevant documents released by the European Data Protection Board and the "Regulations on the Administration of internet Information Service Algorithm Recommendations" emphasize the importance of focusing on technology ethics and big data security issues. Thus, in the process of human-machine communication, AIGC tools can indeed achieve real-time interaction and cross-modal content generation, but they also pose potential data ethics risks.

In summary, this paper defines AIGC as a content production method based on artificial intelligence technology for cross-modal content generation, characterized by massive data, algorithm-driven processes, computing power support, real-time interactivity, and modality diversity.

### **Evolution of AIGC in Human-Machine Communication**

The evolution of human-machine communication typically requires the upgrading and support of data, algorithms, and computing power. With the successive introduction of deep learning algorithms such as GANs, transformer models, and diffusion models, AIGC has transcended the auxiliary phase of simple text content generation, achieving cross-modal content generation and initial autonomous content output, thereby assisting humans in various tasks. In more advanced stages of human-machine communication, more sophisticated multimodal models, vast amounts of data, and robust computing power are needed to support the construction of digital spaces and content generation. Virtual humans empowered by AIGC can potentially become dynamic communication subjects. Therefore, this paper traces the evolution of AIGC in human-machine communication along the lines of changes in data, algorithms, and computing power, dividing it into three stages: the auxiliary tool stage, the communication collaboration stage, and the dynamic communication subject stage.

#### **Auxiliary Tool Stage: Assisted Content Generation**

In the first half of the 20th century, the British mathematician Alan Turing designed the "Turing Test" to verify whether a computer could exhibit intelligence. Against this backdrop, the development of early constrained dialog exemplified by the "ELIZA program" and the "Chinese Room" problem spurred deeper exploration of "intelligence" in AI research. These pioneering works laid the foundation for the later emergence of AIGC.

In human-machine communication, AI at this stage is limited by immature algorithm models, relying on "modular customization" for simple text content generation, assisting the communication subject in completing specific tasks, and serving as a tool. For example, the Stats Monkey software developed by Northwestern University in 2009 generated a baseball game report in 12 seconds, and Xinhua News' "Kuaibi Xiaoxin" completed a financial news report in 3 seconds in November 2015. At this stage, AIGC can be regarded as an auxiliary tool for information processing and presentation, mainly focusing on highly structured text generation, with generated content being more rigid and lacking autonomous learning and creation capabilities.

#### **Communication Collaboration Stage: Cross-Modal Content Generation**

In the early 21st century, the advent of deep learning algorithm models such as GANs, transformer models, diffusion models, and CLIP models, coupled with the massive growth of global internet data and advancements in hardware computing power, provided the foundation for AIGC's cross-modal content generation. AI at this stage can output high-resolution realistic images based on text input, achieve initial autonomous content output, and assist humans in more complex information transmission and creation, thereby increasing production efficiency.

Modality refers to the form in which information, such as text, images, and videos, is stored. Cross-modal generation is based on the semantic consistency between modalities, realizing mutual conversion between different modality data forms, which helps improve transferability between different modalities. Cross-modal content generation disrupts traditional single-modality interaction methods, exemplified by the AI painting tools Stable Diffusion and Midjourney, which emerged in

2022, and the text-to-video tool Sora, which was released in early 2024. However, video and animation modality conversion remain in the exploratory stage. With further technological and industrial development, cross-modal technology represents a new breakthrough direction for AI technology and industry. Future challenges include cross-modal retrieval and cross-modal human-machine interaction.

### **Dynamic Communication Subject Stage: Digital Survival**

Following ethical and moral guidelines, achieving personalized content services, high adaptability scene matching, and real-time human interaction based on cross-modal content generation is the next breakthrough point for AIGC. In the future, in digital spaces where reality and virtuality coexist, humans will interact in real time with virtual humans empowered by AIGC, making digital survival a new state of existence.

To this end, the establishment of digital spaces and content production supported by mature multimodal models, vast amounts of data, and powerful soft and hardware computing power will be realized through further technological upgrades. This not only enhances production efficiency but also transforms production relationships. In this stage of human-machine communication, machines exhibit a certain degree of quasi subjectivity and are capable of self-adjustment and optimization according to the environment and goals, demonstrating a degree of initiative and purposiveness. Virtual humans empowered by AIGC might alter production relations, serving both as tools for content generation and as subjects for content production and dissemination. Digital residents will not only need to accept digital products and services but also adjust the dominance of instrumental rationality and human value rationality, acknowledging the relationships and issues of human-machine communication. However, humans must also recognize the distinction between the quasi subjectivity limited by programming rules and data training and real subjectivity and understand the importance of human-machine coexistence. In intelligent communication activities, various human subjects and intelligent technologies form a new network of actors together.

### **Core Technologies and Ethical Risks of AIGC Cross-Modal Content Generation**

#### **Core Technologies of AIGC Cross-Modal Content Generation**

With the rise of "deep learning," subfields of machine learning, algorithm models have become capable of more precise and efficient information processing and learning. Researchers have thus begun exploring the potential of using deep neural networks for cross-modal content generation. Early attempts included the use of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for multimodal task processing. A series of advanced models subsequently emerged, such as generative adversarial networks (GANs), transformer models, diffusion models, and contrastive language-image pretraining (CLIP) (Figure 1). Deep learning has become the core technical logic of AIGC cross-modal content generation, which is characterized by data-driven approaches, large-scale parallel computing, and multimodal integration. These technologies are widely used in modern strong AI applications.

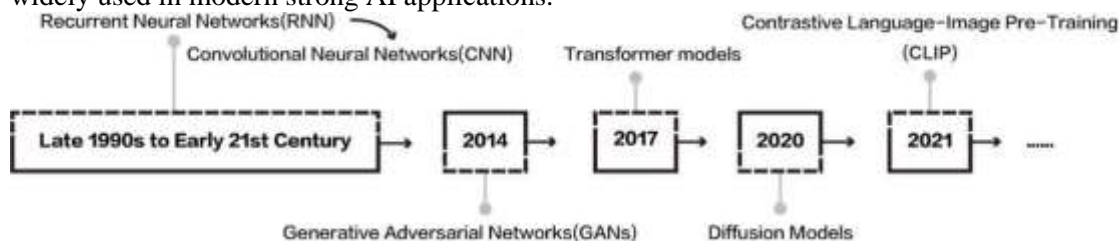


Figure 1. Timeline of Core Technologies in AIGC Cross-Modal Content Generation

#### **Generative Adversarial Networks (GANs): Traditional Approach to Cross-Modal Content Generation (2014)**

Generative adversarial networks (GANs) are a traditional approach to image generation that can directly learn the distribution of sample data and sample from it to obtain new samples such as the original data. GANs consist of three parts: generation, discrimination, and adversarial training. The generator is responsible for producing various modalities based on random vectors, including images, text, music, etc., depending on what content the user wants to "create." The discriminator judges whether the input content is from the dataset or machine generated. The alternating training between

the two constitutes the "adversarial process." For example, Midjourney combines GANs and natural language processing (NLP) to achieve cross-modal content generation by generating images that match given textual descriptions. While GANs have made significant achievements in image generation and editing in human-machine communication, they face issues such as nonconvergence and mode collapse during training. When the model finally converges during training, further training and larger datasets may not improve the model.

### **Transformer Models: Landmark Models in the Field of NLP (2017)**

In 2017, Google's team proposed the transformer architecture, revolutionizing the way machines process language. This model abandons traditional CNNs and RNNs, relying entirely on the attention mechanism, making it a landmark model in the field of natural language processing (NLP). The transformer architecture allows parallel attention, meaning that it can simultaneously focus on all parts of the input text and process it synchronously. This advantage enables it to train on larger datasets, excel at handling long texts and complex language structures, and improve information processing capabilities and efficiency. For example, GPT-4, released on March 14, 2023, uses the transformer architecture, and is optimized with 176 billion parameters to increase performance. Google's Video BERT also attempts to extend transformers to the "video-to-text" domain, demonstrating the technical feasibility of Transformer + pretraining in multimodal integration. However, for specific fields or languages with scarce or difficult-to-obtain data, the transformer model may face insufficient training data issues, impacting its effectiveness and applicability in these areas.

### **Diffusion Models: Emerging Generative Models (2020)**

In 2020, the introduction of denoising diffusion probability models (DDPMs) further advanced the technological revolution and content innovation in cross-modal content generation. Compared with GANs, diffusion models have a relatively stable training process and typically exhibit good generalization ability when trained on large-scale data, making them an emerging model in the field of cross-modal content generation. This model achieves cross-modal generation through forward diffusion and reverses diffusion processes, adding Gaussian noise to the original image and denoising to restore it during training. Notably, the final generated latent image is highly similar but not identical to the original image. Popular diffusion model applications include Dall-E 2, stable diffusion, and Sora. However, the generation process of diffusion models requires multiple iterations over the entire dataset, computing numerous probability distributions, which may hinder real-time data processing in resource-constrained environments such as mobile devices or embedded systems, imposing certain demands on user device performance.

### **Multimodal Pretrained Model CLIP: A Crucial Node in Cross-Modal Content Generation (2021)**

In 2021, OpenAI released the multimodal pretrained model CLIP (contrastive language-image pretraining). This model combines cross-modal generation, contrastive pretraining, and unsupervised learning, becoming a crucial node in cross-modal content generation. OpenAI collects a dataset of over 400 million "image-text" pairs called WIT (Web Image Text) for CLIP, which employs contrastive learning through image and text encoders, and its pretraining process is unsupervised, providing strong generalization performance for cross-modal tasks. In the field of cross-modal content generation, "CLIP + other models" has become a common practice, effectively achieving image search and retrieval tasks; solving visual question answering and image description generation; and performing artistic creation, evaluation, and analysis of the performance of other multimodal models. However, like other models, CLIP also faces challenges such as high computational resource consumption, sample bias, and data privacy issues.

### **Ethical Risks Arising from AIGC Cross-Modal Content Generation Technologies**

An exploration of the core technologies underlying cross-modal content generation reveals that all algorithm models heavily rely on big data. Data have become the foundational logic driving the continuous development of artificial intelligence. However, from existing cases and the usage of cross-modal generation tools, a series of ethical issues triggered by AIGC cross-modal content generation have emerged, with data being the root cause of these ethical concerns.

### **Lack of Transparency in Data Scraping, leading to Content Production Copyright Disputes**

Data scraping is the primary step in the operation of generative artificial intelligence. The richer and more diverse the scraped data are, the more accurate and profound the generated content. Cross-modal content generation models, which are based on billions of existing datasets, possess strong affordances. However, most AI companies often keep the sources of their training datasets confidential and lack transparency. In November 2023, OpenAI reneged on its transparency commitment, refusing to disclose internal documents to the public. Concurrently, there is information asymmetry in human-machine communication, where the scraper holds more information than the scrapped party does, placing the latter at a disadvantage in copyright issues and making it difficult to protect their rights effectively. Many datasets are collected, used, and disclosed without authorization or the knowledge of the original creators. For example, on January 15, 2023, artists Sarah Anderson and two others used Stability AI, Midjourney, Stable Diffusion, and DeviantArt to use their original images as training data without permission. This was the first lawsuit filed by copyright holders against technology companies for using their work in AI training.

Additionally, there is no consensus on the copyright ownership of intelligent publications involving AIGC. Since the current AIGC operates by processing data through large cross-modal content generation models and lacks autonomy, only a few countries, such as the UK, recognize some AIGC-generated content as "works" from a legal perspective. Most countries' copyright laws and precedents do not yet acknowledge AIGC-generated content as original "works." Prominent international journals, such as *Nature* and *The Lancet*, have declared that they do not recognize AI participation in articles. In China's judicial practice, the copyright of AIGC is still judged based on "natural person and originality," which are key considerations for determining the qualifications of a work.

In the era of human-machine communication, the opacity of the data scraping process urgently calls for professional and feasible solutions to copyright issues. Some industry experts have proposed that "technology-induced problems should be addressed with technology." In June 2021, the Ministry of Industry and Information Technology and the Cyberspace Administration of China jointly issued Guidelines on Accelerating the Application and Development of Blockchain Technology. These guidelines highlight the potential of blockchain to address trust and security issues in cyberspace. Blockchain-based digital copyright protection mechanisms provide traceability during data scraping, protecting the interests of copyright holders.

### **Algorithm Bias Deepens Data Set Bias, Exacerbating AI Filter Bubbles in the Intelligent Era**

"Behind the digits are humans; digital systems integrate closely with the culture in which they are rooted." When AI models are developed, human biases, including racial, gender, and cultural prejudices, are often embedded. The designed algorithms might "replicate" and amplify these biases to some extent. For example, in 2014, Amazon's AI screening system was found to favor men over women. Since multimodal pretraining models can learn from user feedback, human subjective biases can further intensify algorithmic biases. Although data purification techniques can delete or anonymize private or sensitive information, they might reduce data quality by removing or altering useful information, introducing double biases, and increasing the harmfulness of large language model outputs.

"AI filter bubbles" refer to the phenomenon where AI, which is based on training datasets and algorithms, understands user needs and filters out heterogeneous information, creating a "bubble" around users. With the advent of the intelligent era, cross-modal content generation will be fully improved. In human-machine communication, users enter a space where the virtual and real coexist, enjoying the convenience brought by AI. However, this can also lead to AI filter bubbles due to underlying logical biases, causing new rounds of information bias and stereotypes and weakening subjectivity. Thus, balancing shared ethical standards with encouraging open and flexible thinking remains a long-term and complex challenge.

### **Lack of Gatekeeping in Data Processing Results, facilitating the Spread of False Information**

"No picture, no truth" is a thing of the past; even with pictures and videos, the truth is not guaranteed. When empowered by technology, users can generate false information in batches by inputting text. The lack of gatekeeping in data processing results can easily disrupt social order and

lead to group polarization, with false information squeezing the public space. According to the social amplification of risk theory, media processing of risks can influence the intensity and breadth of risk perception, either amplifying or attenuating public awareness of risks and forming risk ripples. When false information combines with algorithmic recommendations and spreads, ordinary users' perception of information risk increases, and trust in the media decreases. In the post-truth era, emotions often outweigh facts, and multilevel dissemination blurs the truth. This trend challenges users' digital literacy, platform self-discipline, and the credibility of official media.

Although the final latent image generated by cross-modal models is highly similar but not identical to the original image, when users generate cross-modal content, the lack of gatekeeping in data processing results can easily lead to the presentation of false information. In March 2023, Elliott Higgins used Midjourney to generate images of Trump allegedly being arrested and in prison, which, although quickly identified as fake, sparked intense discussion. In May 2022, a fraud gang used deepfake technology to impersonate Elon Musk, tricking victims into joining the Bit Vex bitcoin scam platform, causing significant harm. A report by Europol even predicts that by 2026, up to 90% of online content might be AI-generated or edited.

### **Market Proliferation of Cross-Modal Content Generation Applications, Further Widening the Intelligent Divide**

In the era of human-machine communication, the market proliferation of cross-modal content generation applications is imminent. However, from the perspective of social development, the ownership of tools determines the distribution of social resources. The technological upgrades brought by industrial revolutions have sharpened the contradictions between production and capital, leading to the gradual solidification of social strata. The paradox of technological empowerment is once again raised in the intelligent communication era; technological advancements have not narrowed the gap between digitally underprivileged and digitally privileged groups. The intelligent divide is more covert than the digital divide and exhibits the Matthew effect.

Existing cross-modal content generation applications have certain usage barriers, with some applications requiring adherence to specific commands and formats and differing in question phrasing and terminology. Moreover, some advanced applications developed by foreign companies support only a few languages.

The intelligent era delineates a survival experience gap between the societal mainstream and minority disadvantaged groups, which is deeper than intergenerational or urban-rural divides. This technological-tool-based separation is both a priori and irreversible. Consequently, the digital wealth will grow stronger in technology and efficiency, while the digital poor will be further marginalized. This not only hinders some groups from enjoying the benefits of intelligent technology but also may deprive them of their rights to social participation, making social interaction and integration difficult. Identifying and bridging the intelligent divide has become a new challenge in the human-machine communication process.

### **Development Pathways for AIGC Cross-Modal Content Generation Technologies**

Philosopher Gilbert Hotto proposed the concept of "technology accompaniment," asserting that "it is not about where we should draw the line between humans and technology, but how we should build relational boundaries between them". From the perspective of accompaniment ethics, in seeking and driving positive development pathways for AIGC cross-modal content generation technologies, humans and technology should coexist symbiotically, maintaining a closely linked state to dissolve their antagonistic relationship, aiming to construct a community of shared destiny where humans and machines can flourish together.

### **Microlevel: Human-Machine Symbiosis and Digital Interaction**

In the intelligent communication era, the evolution of human-machine relationships has become a focal point for both academia and industry. Human-machine relationships should not be confined to opposition but should shift toward the concept of human-machine symbiosis, promoting mutual learning and integration in digital interactions.

Within the scope of human-machine communication, digital interaction refers to the process of communication, interaction, and exchange between humans and machines through digital platforms and technologies supported by big data. In this process, there are similarities and differences in how

humans and machines learn. Both human learning and machine deep learning require the use of data (including text and digital data) and feedback. However, human learning is limited by time and the amount of information that can be processed, whereas machine learning relies on the quantity and quality of data and lacks perceptual ability and creativity. Therefore, the key to human-machine symbiosis lies in achieving complementary cooperation between humans and machines in digital interactions.

In practice, the concept of human-machine symbiosis is crucial for the development of intelligent communication. For example, intelligent recommendation systems can incorporate the idea of human-machine symbiosis, consider the interaction and learning process between users and technology, and enhance the feedback mechanisms in human-machine communication. Additionally, in the field of AI-assisted creation and innovation, the concept of human-machine symbiosis provides new ideas and methods for digital communication. For example, when a screenwriter interacts with ChatGPT, the human can gain creative inspiration and fill logical gaps, whereas the machine learns new vocabulary, grammatical structures, and creative techniques from the interaction with the human.

In the process of human-machine symbiosis, humans should not worry excessively about being replaced by AI but should approach practice and interaction with an open and proactive mindset. The interaction between humans and AI forms a dynamic knowledge-sharing and collaborative model, injecting a new impetus into social development and progress.

### **Medium Level: Diverse Participation and Progressive Norms**

The stakeholders of AIGC cross-modal content generation technology are numerous, requiring diverse forms of participation and the collection of opinions from various parties to establish robust technical review and regulatory mechanisms, collect high-quality data, and build open-source databases, thereby promoting the integration of technology and society.

### **Soliciting Diverse Opinions to Establish Robust Technical Reviews and Regulatory Mechanisms**

Under the concept of human-machine symbiosis, a collaborative governance mechanism should be established promptly to ensure that decision-making is open, fair, and scientific. Governments need to establish specialized agencies or committees responsible for coordinating AI technology governance affairs, which play a leading role. By soliciting opinions from various sectors, involving users, AI technology developers/companies, scholars, and researchers in the AI field (especially experts focusing on regulatory policies and ethical issues), these stakeholders should actively participate in the decision-making process for technical review and regulation. The establishment of an open and transparent decision-making mechanism ensures the openness, transparency, and scientific nature of the decision-making process. In April 2023, the National Telecommunications and Information Administration of the United States solicited public opinion. In May of the same year, Sam Altman, the father of ChatGPT, proposed the need to introduce licenses and form expert groups to independently audit AI models. Thus, the decision-making mechanism should realize diversified participation in AI technology governance, with clear procedures and standards, regulating the decision-making process and content, balancing technological development with social responsibility, and promoting the healthy development of AI technology.

### **Collecting and Using High-Quality Data and Building Open-Source Databases**

To ensure a diverse range of data sources, it is crucial to collect data from various cultural, geographic, gender, age, and socioeconomic backgrounds to capture a wide array of perspectives and experiences. Additionally, methods such as resampling and data synthesis should be employed to ensure that the dataset accurately represents different groups. Future valuable data must ensure high quality, diversity, and accuracy, but this entails high costs and low efficiency in data collection and processing. Therefore, cooperating with diverse organizations or institutions to establish a shared document-style data collection model and open-source databases is an effective strategy for collecting and using high-quality data. Open data collection documents or platforms invite various parties to submit and edit data jointly, realizing collaborative data collection and updating. This model can attract more participants to contribute data, promoting data diversity and representativeness. Moreover, government organizations should take the lead in developing open-source database platforms for data collectors and users to share and access data. These databases can be constructed

based on open-source technology, providing efficient data storage, management, and query functions while ensuring data security and privacy.

### **Macro Level: Global Dialog and Ethical Shaping**

In the era of intelligent communication, there is no unified global consensus on human-machine ethics, and the digital divide between nations continues to widen. Therefore, recognizing the challenges in global ethical dialog can lead to a more scientific approach to guiding AI technology. One of the challenges in jointly guiding AIGC cross-modal content generation technology toward positive development is that humanity's value discrepancies and lack of digital literacy exacerbate AI system biases. While this may seem to be an objective result of AI development, to some extent, AI bias is a "mirror" of real-world biases.

Properly guiding the development of AI technology and humanistic values can be a solution. This requires encouraging multilateral dialog and cooperation among major countries to promote understanding of and respect for cultural differences. First, establishing transnational exchange platforms and cooperation mechanisms can foster mutual understanding and negotiation between different countries, providing a broader base for consensus in global AI governance. In this process, international cooperation mechanisms and multilateral negotiations play crucial roles, collaboratively formulating a Global Data Ethics Convention that outlines the fundamental principles and ethical guidelines for AI application, safeguarding the global public interest and humanity's common future. Such a global ethical framework can provide guidance in transnational cooperation, directing national behaviors to ensure the achievement of human-machine symbiosis.

Moreover, strengthening international cooperation should focus on technology transfer and sharing. Developed countries should share advanced AI technologies and experiences with developing countries, enhancing global technological capabilities with an open and inclusive attitude to address global challenges jointly. A mechanism for technology transfer should be established to ensure that the dissemination and application of AI technology benefits all regions and groups worldwide, aiming to create a more harmonious and inclusive community of shared human-machine destiny.

## **II. Conclusion**

With the continuous evolution of deep learning algorithm models, the massive input of data, and the enhancement of computer chips, AIGC has transitioned from text to realistic text, from text to image, and from text to video, laying the foundation for new forms of human-machine symbiosis. In the future, digital space where virtual and real coexist, digital living will become a new state of existence, allowing humans and AI-empowered virtual beings to achieve seamless cross-modal interaction.

In the intelligent communication era, balancing relationships and issues in human-machine communication requires embracing the concept of technology accompaniment and guiding the high-quality and innovative development of cross-modal content generation technologies. Increasing the frequency and possibilities of digital interaction, enhancing users' comprehension and autonomy in applications, and improving digital literacy in the intelligent era are essential. Simultaneously, it is crucial to ensure that government-led initiatives, guided by global dialog, establish diverse participatory regulatory mechanisms and conventions, promoting the construction of a more harmonious human-machine community with a shared destiny.

### **References**

- Zhan, X., Li, B., & Sun, J. (2023). Scenario-based application and development opportunities of AIGC in the context of digital intelligence integration. *Journal of Library and Information Knowledge* 01: 75-85+55. DOI: 10.13366/j.dik.2023.01.075.
- China Academy of Information and Communications Technology (2022) Artificial Intelligence White Paper. Available at: <http://www.caict.ac.cn/kxyj/qwfb/bps/202204/P020220412613255124271.pdf> (accessed 10 July 2023).
- Liu, H., Chen, J., Li, L., Bao, B., Li, Z., Liu, J., & Nie, L. (2023). Cross-modal representation and generation technology. *Chinese Journal of Image and Graphics* 06: 1608-1629. DOI: 10.11834/jig.230035.

- Peng, L. (2023a). AIGC and the new survival characteristics of the intelligent era. *Nanjing Social Sciences* 05: 104-111. DOI: 10.15937/j.cnki.issn1001-8263.2023.05.011.
- Peng, L. (2024). Human actors in intelligent communication. *Journal of Northwest Normal University (Social Science Edition)* 61(04): 25-35. DOI: 10.16783/j.cnki.nwnus.2024.04.003.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems* 27: 2672-2680. DOI: 10.48550/arXiv.1406.2661.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems* 30. DOI: 10.48550/arXiv.1706.03762.
- Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33: 6840-6851. DOI: 10.48550/arXiv.2006.11239.
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In: *International Conference on Machine Learning*, 8748-8763. PMLR. DOI: 10.48550/arXiv.2103.00020.
- Chen, Y. (2023). AIGC infringement risks and digital copyright protection strategies in the context of the intelligent era. *Communication and Copyright* 17: 113-116. DOI: 10.16852/j.cnki.45-1390/g2.2023.17.030.
- Thomas, K., & Zheng, Y. (2007). *Digital Anthropology*. Central Compilation and Translation Press.
- Zhang, X. (2023). Algorithmic governance challenges and governance-based supervision of generative artificial intelligence. *Modern Jurisprudence* 03: 108-123.
- Pi, J. (2010). *Social amplification of risk*. China Labor and Social Security Press.
- Jiang, H. (2023). Coexistence and symbiosis between humans and ChatGPT: From the “digital divide” to the “digital intelligence divide” — Taking Japan’s “skill reshaping” plan as an example. *Journal of Yuejiang* 03: 74-83+174. DOI: 10.13878/j.cnki.yjxk.20230426.007.
- Verbeek, P.P., & Yang, Q. (2013). Accompanying technology: Philosophy of technology after the ethical turn. *Journal of Luoyang Normal University* 04: 18-21. DOI: 10.16594/j.cnki.41-1302/g4.2013.04.001.
- Yu, Guoming., Lin, Yutong., Li, Yunyue. (2024). Generative AI as a new content productivity: Development limitations and future directions. *Publishing Horizon* 14: 22-30. DOI: 10.16491/j.cnki.cn45-1216/g2.2024.14.004.