



Editorial:

Recent advances in artificial intelligence generated content

Junping ZHANG^{††1}, Lingyun SUN², Cong JIN³, Junbin GAO⁴, Xiaobing LI⁵,
 Jiebo LUO⁶, Zhigeng PAN⁷, Ying TANG⁸, Jingdong WANG⁹

¹School of Computer Science, Fudan University, Shanghai 200433, China

²International Design Institute, Zhejiang University, Hangzhou 310058, China

³School of Information and Communication Engineering, Communication University of China, Beijing 100024, China

⁴The University of Sydney Business School, The University of Sydney, NSW 2006, Australia

⁵Department of AI Music and Music Information Technology, Central Conservatory of Music, Beijing 100032, China

⁶Department of Computer Science, University of Rochester, NY 14627, USA

⁷School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China

⁸Electrical and Computer Engineering Department, Rowan University, Glassboro, NJ 08028, USA

⁹Baidu, Beijing 100085, China

[†]E-mail: jpzhang@fudan.edu.cn

<https://doi.org/10.1631/FITEE.2410000>

Artificial intelligence generated content (AIGC) has been a research hotspot in the field of artificial intelligence in recent years. It is expected to replace humans in performing some of the work of content generation at a low cost and a high volume, such as music, painting, multimodal content generation, news articles, summary reports, stock commentary summaries, and even content and digital people generated in the meta-universe. AIGC provides a new technical path for the development and implementation of AI in the future.

In this context, the journal *Frontiers of Information Technology & Electronic Engineering* has organized a special issue on the recent advances in AIGC. This special issue focuses on theories, algorithms, and applications of AIGC, and its related fields. By attracting high-quality papers, we hope to help researchers in academia and industry gain a deeper understanding of the fundamental theories behind AIGC and its potential applications. These high-quality works will inspire more people to join and

further advance the field of AIGC. We thus called for papers on the following topics (but not limited to): (1) AI-generated music; (2) AI-generated painting; (3) AI dialogue models; (4) AI news summaries; (5) AI and the metaverse; (6) AI and digital humans; (7) AI image editing; (8) AI-generated short videos; (9) AI-generated multi-media content; (10) Chat Generative Pre-trained Transformer (ChatGPT) related work. Twelve papers have been selected for this issue after a rigorous review process, including one comment paper, one perspective paper, three review articles, six research articles, and one correspondence paper. We organize them in three main parts, including ChatGPT, diffusion models, and prompt learning and multimodality.

1 ChatGPT

Since OpenAI released ChatGPT in December 2022, it has quickly attracted considerable attention from the industry and academia because of its impressive abilities in terms of text generation, conversation, and complex reasoning.

[†] Corresponding author

ORCID: Junping ZHANG, <https://orcid.org/0000-0002-5924-3360>

© Zhejiang University Press 2024

To foster a better understanding of ChatGPT, Junping ZHANG et al. introduced the history of ChatGPT, discussed its advantages and disadvantages, and pointed out several potential applications in detail. Then, they analyzed its impact on the development of trustworthy AI, conversational search engines, and artificial general intelligence (AGI). Note that, inspired by ChatGPT, a series of large language models (LLMs) have been proposed globally. In particular, GPT-4, GPT-4 Turbo, and Gemini introduced multimodality, giving LLMs the significant potential to become AGI, and ensuring that AIGC achieves more accurate content representation than popular prompt-based methods.

ChatGPT also has the ability to supplement regular class learning. Ying TANG et al. proposed a framework for parallel intelligent education that involves physical and virtual learning for a personalized learning experience using ChatGPT. They analyzed the pros and cons of ChatGPT for parallel intelligent education.

Text generation is another interesting direction in the AIGC domain, and it heavily depends on the development of natural language processing, machine learning, and deep learning. It enables the learning of language rules through training models to automatically generate text that meets grammatical and semantic requirements. Peng YANG et al. surveyed the main research progress in text generation and presented several typical text generation application systems. They claimed that refining the quality, quantity, interactivity, and adaptability of generated text is helpful to AIGC.

Note that LLMs depend on the quality of tokens. However, it is difficult to decompose Chinese characters into meaningful tokens according to the English convention, such as prefixes and suffixes. Li WEIGANG et al. proposed Six-Writings multimodal processing (SWMP) to integrate Chinese natural language processing with morphological and semantic elements. The framework considers Six-Writings pictophonetic coding (SWPC), providing more suitable granularity for Chinese characters and words. It has the possibility of constructing a set of Chinese-oriented tokens specifically tailored to the needs of Chinese LLMs. This can also potentially improve both computational efficiency and accuracy over the existing methods.

2 Diffusion models

Diffusion models have attracted attention from many researchers since they can generate images by iteratively degenerating an image to its wholly noisy version and then decoding it back to its original one. In a similar manner, they can generate diverse contents by tuning the parameters and redesigning the structure along the degenerating path and the decoding path.

Lequan LIN et al. presented an inclusive survey of probabilistic diffusion models for time-series applications, covering five key aspects: time-series forecasting, imputation, generation, practical applications, and future research directions. The survey initially explored three fundamental formulations of diffusion models, followed by an exhaustive review of existing methods for time-series forecasting, imputation, and generation. Among them, special attention was paid to how these formulations were adapted for supervised time-series tasks. The comparison of methods included an assessment of their capabilities in handling various tasks and data types, along with an examination of their diffusion and sampling processes. The paper showcased the practical utility of these methods through an in-depth discussion of their application in real-world domains such as energy control and transportation. Additionally, it listed valuable resources of frequently used public datasets, to support researchers and practitioners in testing and understanding diffusion models for time series. Furthermore, it provided a discussion of future research directions, highlighting limitations and challenges drawn from the existing methods.

Recently, various AIGC algorithms have been developed for generating appealing music. A disadvantage is that the control of music style, which is one of the most salient qualities of music, is less studied. By introducing a style-conditioned linear Transformer and a style-conditioned patch discriminator, Xiaofen XING et al. successfully kept the consistency of the generated music concerning music styles for music generation.

Although AIGC can produce high-quality content, its interpretability and controllability remain a challenge. Li LIU et al. combined causal representation learning with bi-directional generative adversarial networks (GANs), allowing users to control image attributes and to generate counterfactual images. The

proposed causal controllable image generation (CCIG) can effectively learn the causal relationships between image attributes and joint optimization with the encoder, generator, and joint discriminator in the image generation module. Experiments showed that using intervention operations and learned latent representation leverages the performance of facial image generation.

While AI generation of two-dimensional (2D) images and text has progressed significantly, three-dimensional (3D) content generation remains challenging due to the need for spatial awareness and complex shape representation, and the lack of large 3D datasets. Conventional 3D modeling with computer-aided design (CAD) software poses difficulties for novices due to steep learning curves. Lingyun SUN et al. proposed a sketch-based modeling approach by leveraging intuitive human-computer interaction. Specifically, the proposed Deep3DSketch-im approach leverages implicit surface representations to generate detailed 3D models from single freehand sketches, employing a convolutional neural network (CNN) to encode sketches into signed distance field (SDF) feature vectors for shape prediction and sampling points at infinite locations for infinite-resolution implicit surfaces, demonstrating state-of-the-art performance on synthetic and real datasets in user studies. This research presents a novel solution that could elevate 3D modeling accessibility to benefit industries through quicker, easier custom 3D content creation.

Qibin ZHAO et al. presented their newly proposed diffusion model TendiffPure for robust data generation, including image generation, with high efficiency. They purified noised or adversarially perturbed data, i.e., removing the noises and adversarial attacks in the data, without any assumption of the forms of noises and adversarial attacks. They also largely reduced the number of parameters of U-Nets, which are the backbones of diffusion models, leveraging the tensor-train decomposition. Here, the tensor-train decomposition provides a potential benefit to tightening the robust error bound and hence amplifies the robustness of data generation. Experiments were conducted on CIFAR-10, Fashion-MNIST, and MNIST datasets. The results demonstrated that not only the generation quality on clean images was improved but also the purification quality on noised and

adversarially attacked images was enhanced by TendiffPure. This work broadens the horizon of future research on diffusion models for efficient image generation and adversarial purification. It also revitalizes classical methodologies by innovatively applying tensor decomposition and tensor network techniques, demonstrating their enduring effectiveness in current machine learning contexts.

To obtain the same performance as the artist-ability paintings, Taihao LI et al. collected the opinions from three groups of individuals with varying levels of art appreciation ability, and used these opinions to play an auxiliary role to shorten the distance between painters and artists. Specifically, a multi-stage text-conditioned strategy was fused into the diffusion model so that a multilevel semantic representation could be shown in a generated image.

3 Prompt learning and multimodality

Besides ChatGPT and diffusion models, prompt learning is broadly used for AIGC since prompts provide a compliment to semantic information for improving the generated quality of LLMs. Hongming SHAN et al. presented a comprehensive review of prompt learning in the computer vision community, including four popular research fields: vision-language pre-training, vision prompt learning, prompt-guided generative models, and prompt tuning. They provided a progressive and comprehensive review of visual prompt learning as related to AIGC. It first introduces vision-language models (VLMs), such as contrastive language-image pre-training (CLIP), the foundation of visual prompt learning. Then, classical vision prompt learning methods are reviewed, showing their abilities to adapt pre-trained models to downstream tasks. Based on these preliminaries, the popular prompt-guided generative models are discussed through widely studied topics such as image generation, inpainting, and editing. The authors also focused on how to improve the efficiency of fine-tuning large-scale transformer based models. Finally, they provided some promising research directions concerning prompt learning.

As for multimodal AIGC, the end-to-end cross-lingual summarization (CLS) model has achieved

impressive results using large-scale, high-quality datasets, typically constructed by translating monolingual summary corpora into CLS corpora. Due to the limited performance of low-resource language translation models, however, translation noise can significantly degrade the performance of these models. Zhengtao YU et al. proposed a fine-grained reinforcement learning approach. This method introduces the source language summary as a gold signal, and calculates word correlation and word missing degree between the source language summary and the generated target language summary to design a reinforcement reward. They combined this reward with cross-entropy loss to optimize the CLS model, addressing the low-resource CLS problem based on noisy data.

Overall, a broad spectrum of current research topics relevant to the development and applications of AIGC is covered in this special issue. These included AIGC applications in image/text generation, 3D content creation, user-centric layout in graphic design, and style-specified music generation. Some works related to causal representation learning and high-order diffusion models have been proposed to leverage the performance of AIGC. Moreover, probabilistic diffusion models, prompt learning, and ChatGPT are surveyed in detail.

Finally, we thank all the authors for their support to this special issue. We are especially grateful to all the reviewers for their insightful comments and helpful suggestions for all the submissions.



Junping ZHANG is a professor with the School of Computer Science, Fudan University. He received the BS degree in automation from Xiangtan University, China, in 1992, the MS degree in control theory and control engineering from Hunan University, China, in 2000, and the PhD degree in intelligent systems and pattern recognition from the Institution of Automation, Chinese Academy of Sciences, in 2003. He has more than 100 publications in major journals and international conferences, including *IEEE TPAMI*, *IEEE TNNLS*, *IEEE TCYB*, *NeurIPS*, *ICCV*, *CVPR*, *ICML*, and *IJCAI*. He has been an associate editor of *IEEE Intell Syst* since 2009. His research interests include machine learning, image processing, biometric authentication, and intelligent transportation systems.



Lingyun SUN is a professor at Zhejiang University. He is the director of the International Design Institute, Ng Teng Fong Chaired Professor, and the director of the ZJU–SUTD Innovation, Design and Entrepreneurship Alliance. He has over 100 publications in leading academic journals and international conferences, including *Des Sci*, *AAAI*, and *ACM CHI*. His projects have won many awards, including the SAIL award in 2021. His research interests include design intelligence, generative AI, and information and interaction design.



Cong JIN is an associate professor with the School of Information and Communication Engineering, Communication University of China. She received the BE, MAS, and PhD degrees in communication and information systems from Communication University of China in 2010, 2013, and 2021, respectively. She has been presiding over an undertaking of the youth, general, and key projects of the National Natural Science Foundation, the Xiaomi Joint Fund of the Beijing Natural Science Foundation, and National Key Research and Development Projects. She has published more than 40 papers in *IEEE Trans*, *ACM MM*, and other international journals and conferences. She has served as a session chair or a program committee member for several major international conferences and as an associate editor or reviewer for several leading journals. Her research interests include hybrid man–machine performance, reinforcement learning, and music AI.



Junbin GAO is a professor of big data analytics at the University of Sydney Business School, University of Sydney, and was a professor in computer science at the School of Computing and Mathematics, Charles Sturt University, Australia. He graduated from Huazhong University of Science and Technology (HUST), China, in 1982 with a BS degree in computational mathematics and obtained his PhD degree from Dalian University of Technology, China, in 1991. He was a lecturer in computer science from 2001 to 2005 at University of New England, Australia. From 1982 to 2001 he was an associate lecturer, lecturer, associate professor, and professor in the Department of Mathematics, HUST. He has published more than 160 papers in major journals and international conferences, including *IEEE TPAMI*, *IEEE TNNLS*, *IEEE TCYB*, *NeurIPS*, *CVPR*, *ICML*, *AAAI*, and *IJCAI*. His main research interests include machine learning, data analytics, Bayesian learning and inference, and image analysis.



Xiaobing LI is a professor with the Central Conservatory of Music, the director of the Department of AI Music and Music Information Technology, the chair of the China Computer Federation (CCF) Computational Art Branch, the chair of the Chinese Association for Artificial Intelligence (CAAI) Art and the Artificial Intelligence Commission, a leading talent in philosophy

and social sciences under the Ten Thousand Talent Program, a member of the Four-Batch Talent Project, and the Chief Expert of Major Projects of the National Social Science Foundation of China. Graduating from the Composition Department of the Central Conservatory of Music, he studied under the guide of Prof. Zuqiang WU. He has won domestic and international awards such as the Chinese Golden Bell Award for Music, the Wenhua Award of the Ministry of Culture, the Wenhua Composition Award, the First Prize of the National Opera and Dance Drama, and the National Five One Project Award. In 2008, he held the large-scale real-life concert at the Great Wall. In 2019, he held the large-scale real-life Concert of the Future at the Yan'an Luxun Academy of Fine Art.



Jiebo LUO is currently the Albert Arndt Hopeman Professor of Engineering and professor of computer science with University of Rochester, USA, which he joined in 2011 after a prolific career of 15 years with Kodak Research Laboratories. He has authored over 600 technical papers and holds more than 90 US patents. He has been involved in many

technical conferences, including serving as a program co-chair of ACM Multimedia 2010, IEEE CVPR 2012, ACM ICMR 2016, and IEEE ICIP 2017, and as a general co-chair of ACM Multimedia 2018 and IEEE ICME 2024. He was on the editorial boards of *IEEE TPAMI*, *IEEE TMM*, *IEEE TCSVT*, *IEEE Trans Big Data*, *ACM TIST*, *Patt Recogn*, and *Intell Med*. He was the Editor-in-Chief of *IEEE TMM* during 2020–2022. He is a fellow of NAI, ACM, AAAI, IEEE, SPIE, and IAPR. His research interests include computer vision, natural language processing, machine learning, data mining, computational social science, and digital health.



Zhigeng PAN is a professor at Nanjing University of Information Science and Technology, China. He received the PhD degree in 1993. He has been a full professor with Zhejiang University since 1996. He has published more than 100 technical papers in important journals and conferences, including *IEEE TPAMI*, *TVCG*,

IEEE MM, *ACM MM*, and *IEEE VR*. He is a member of ACM SIGGRAPH. He is also a program co-chair of CASA 2011, SIGGRAPH Asia 2011 (Sketches and Posters), IEEE VR 2013, and SIGGRAPH Asia 2016 (Symposium on Education), and a conference co-chair of VRCAI 2012, VRCAI 2013, VRCAI 2015, and CW 2016. He is the Editor-in-Chief of *Metaverse*. His research interests include virtual reality, computer graphics, and human–computer interaction.



Ying TANG is a full professor and the Undergraduate Program Chair of Electrical and Computer Engineering at Rowan University, USA. She received the BS and MS degrees from Northeastern University, China, in 1996 and 1998, respectively, and the PhD degree from the New Jersey Institute of Technology, USA in 2001. Her work has been continuously supported by NSF, EPA, US Army, DOT, private foundations, and industry. She has three US patents, and over 230 peer-reviewed publications, including 77 journal articles, two edited books, and six book/encyclopedia chapters. She is presently an associate editor of *IEEE TSMCS*, *IEEE TIV*, *IEEE TCSS*, and *Disc Artif Intell*. Her research interests include cyber-physical social systems, extended reality, intelligent learning environments, modeling and adaptive control for computer-integrated systems, and sustainable production automation.



Jingdong WANG is chief scientist for computer vision with Baidu. Before joining Baidu, he was a senior principal research manager with Microsoft Research Asia. His representative works include deep high-resolution network (HRNet), object-contextual representations for semantic segmentation (OCRNet), and neighborhood graph search (SPTAG) for large-scale vector search. He has been serving or served as an associate editor of *IEEE TPAMI*, *IJCV*, *ACM TOMM*, *IEEE TMM*, and *IEEE TCSVT*, and as an area chair of leading conferences in vision, multimedia, and AI, such as CVPR, ICCV, NeurIPS, ECCV, ACM MM, IJCAI, and AAAI. He was elected as an ACM distinguished member, a fellow of IAPR, and a fellow of IEEE, for his contributions to visual content understanding and retrieval. His areas of interest are computer vision, deep learning, and multimedia search.