This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

Virtual Human: A Comprehensive Survey on Academic and Applications

LIPENG CUI, (Member, IEEE), JIARUI LIU
Research Institute of China Telecom Corporation Ltd., Beijing 102209, China

Corresponding author: Lipeng Cui (cuilp@chinatelecom.cn).

This work was supported in part by the Research Institute of China Telecom Corporation Ltd.

ABSTRACT As a creative method for virtual human individuals based on multiple fusion technologies such as artificial intelligence, computer graphics, and speech synthesis, virtual human technology has developed rapidly since its birth, and continuous discussions and studies have been conducted in both academia and industry. Starting from the film and television industries, the cross-disciplinary application of virtual human has been continuously recognized and applied in fields such as media, games, and finance. Although virtual human has achieved sufficient development and innovation, it faces many challenges such as emotion recognition, privacy, and security, as well as the uncanny valley effect. This article starts with the development history of virtual human and analyzes the current academic research status and application scenarios in combination with the characteristics, technical architecture, and application of virtual human technology. At the same time, this article sorts out seven mainstream application scenarios of virtual human and analyzes their main advantages and possible future challenges. This article provides a valuable reference for subsequent related research by exploring development trends, application fields, and future research trends in virtual human.

INDEX TERMS Artificial intelligence, deep learning, machine learning, human-computer interaction, virtual reality, virtual human, metaverse

I. INTRODUCTION

In 1964, Boeing developed the first digital human using an image of a human being. Since then, virtual human technology has gradually moved from the laboratory research field to the commercial sector and has become popular in the digital age. From a technical point of view, virtual humans are virtual individuals that have become a reality in recent years with the continuous development of information technologies, such as computer graphics, motion capture, machine learning, high-precision rendering, and speech synthesis. In terms of digital human image construction, early digital human models mainly used modelling software to create manual models, which is an inefficient and costly method. With the application of deep learning techniques, data-driven 3D reconstruction techniques are widely used, and new methods include deep learning-based mapping from face images to 3D Morphable Model(3DMM) parameters [1]–[3], modelling using 3D scanning techniques, etc. In the field of speech synthesis, previous speech synthesis systems have mainly used the joint of rules and language-specific knowledge. In recent years, end-to-end neural network approaches have gradually become mainstream, resulting in significant improvements in the quality of the generated speech [4]–[7]. In terms of dialogue engines, early dialogue systems could only perform common tasks. With the introduction of deep learning techniques, researchers have built a series of smooth and flexible dialogue engines, which are widely used in chatbots, intelligent customer service, and question and answer systems [8]–[10]. On the face-driven side, earlier techniques often required the use of specific markers or sensors for operation, were computationally expensive, and had to be device-specific. Existing face-driven systems perform efficient face expression recognition through convolutional neural networks by training a large amount of data and a library of facial expressions [11].

As virtual human technology continues to evolve, it has expanded its concept from the traditional film, television and media sectors to other areas such as gaming, healthcare and retail. It is constantly being integrated with other application scenarios and its cross-disciplinary

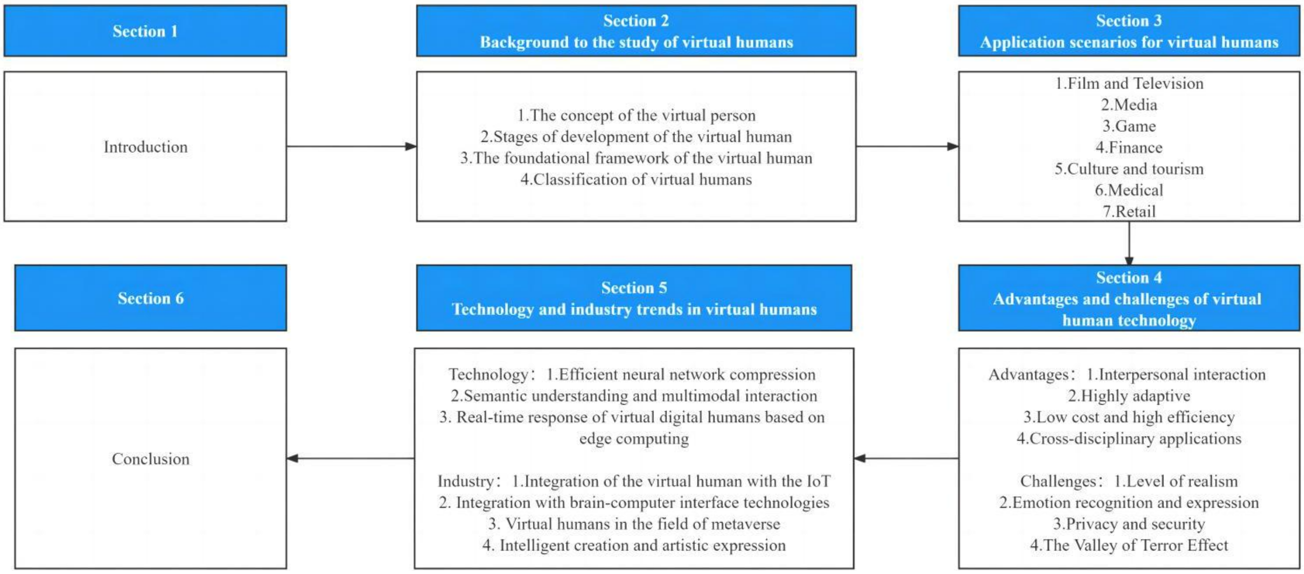

FIGURE 1. The research roadmap for this survey

applicability is being proven. At the same time, are placing great emphasis on this technology and have published a large number of academic results related to virtual human technology. There are a large number of academic studies and papers on virtual human, such as speech synthesis, face expression migration technology, face-driven technology, and some discussions have been proposed on the application and scenarios of virtual human in mature industries such as film and media.

Through the analysis of the existing papers we found that there is no relatively comprehensive and systematic review-type article on virtual human technology in the academic community. In order to highlight the differences with the above papers, we present a comprehensive introduction to the architecture, principles and applications of virtual human from a more technical and comprehensive perspective, focusing on academic data information statistics and industry applications, in order to demonstrate the state of development and the trends of virtual human. This paper is dedicated to the overall study of virtual human technology, as shown in figure 1:

1) A comprehensive description of the concept and characteristics of virtual humans from the background and development history.
2) An analysis of academic concerns regarding virtual human technology and a summary of the latest research results.
3) Analysis of application scenarios and practical applications of virtual human in various fields.
4) Explain the advantages and challenges of virtual human technology in the current situation.
5) Forecast and discuss virtual human technology trends and industry trends.
6) Draw conclusions.

II. BACKGROUND TO THE STUDY OF VIRTUAL HUMAN

A. THE CONCEPTS OF THE VIRTUAL HUMAN

Virtual humans are virtual individuals that have become a reality in recent years with the continuous development of information technologies. The existence of virtual humans is a product of the convergence of multiple technologies, giving them four distinctive characteristics. First, Unlike real people in the physical world, virtual humans exist in a virtual world and differ in some way from real human figures. Second, virtual humans are technological products created by the interplay of multiple technologies, including computer graphics, motion capture, machine learning, high-precision rendering, and speech synthesis. Third, virtual humans have anthropomorphic characteristics; their appearance, body shape, gender, and personality are very similar to those of real people, and they also have the behavioral characteristics of real people, such as language, expressions, and movements [12] - [13]. Finally, compared to real people, virtual human are more malleable in terms of image, persona, and backstory, have more room for imagination, and can be used to meet different needs and scenarios.

When it comes to virtual humans, they often appear together with concepts such as "digital human" and digital doubles," but in order to highlight the focus and key elements of the study, it is necessary to clarify the difference between the object of this paper: the "virtual human", and the "digital human" and "digital double". A virtual human is based on computer graphics, AI, and voice technology, often with realistic physical appearance and movement characteristics, capable of simulating various intelligent human interaction behaviors, and even enabling human-machine interaction through voice recognition, natural language processing, and other

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

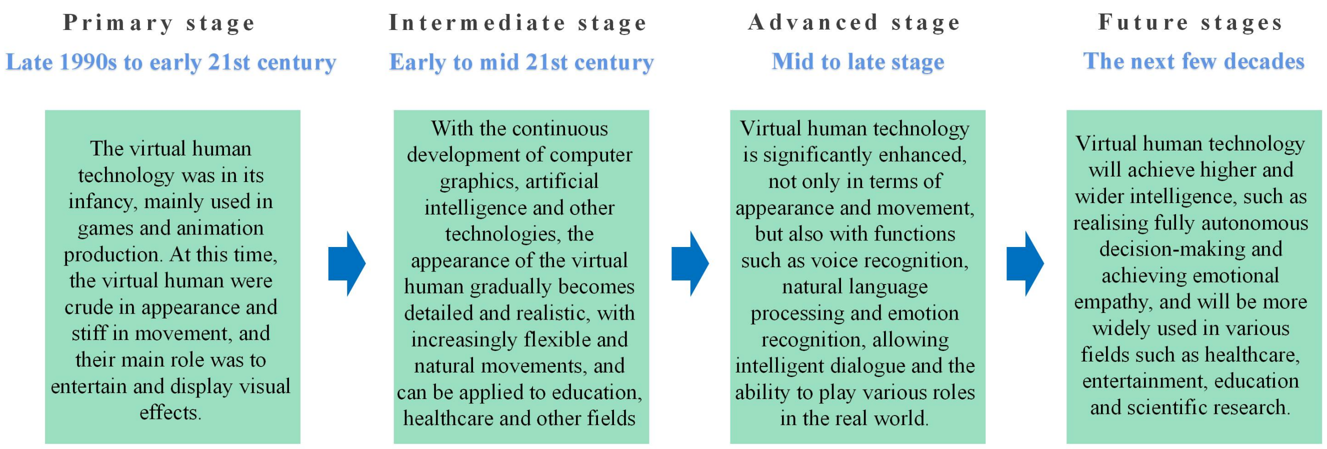

FIGURE 2. **The development of the virtual human technology**

technologies, such as customer service robots and intelligent assistants. A digital human is a persona that is generated entirely by a computer program based on 3D modelling technology. Virtual humans not only include the appearance simulation and intelligent interaction function of digital humans, but also emphasize the simulation and expression of the character's inner world and emotional experience, and are usually applied in the fields of games, film and television, literature, and art. A digital double is a digital model based on the physical appearance of a real person or an existing virtual human that is restored through 3D scanning, modelling, and mapping techniques. Unlike virtual humans, digital doubles emphasize deformation and representation while retaining physical characteristics and are used in areas such as film effects and game character design. In summary, virtual humans, digital humans, and digital doubles are all based on 3D modelling technology, but the concept and application of virtual humans are more widespread, whereas the concept of digital humans and digital doubles is more focused. Digital humans emphasize intelligent interaction and human-machine interface functions, virtual humans focus more on the expression of emotions and psychological states, while digital doubles are more often used in film and television special effects and game character design [14]-[16].

B. STAGES OF VIRTUAL HUMAN DEVELOPMENT

The history of virtual humans dates to the 1980s [17]. In 1982, the Japanese animated film "Super Time Fortress" was produced and made the world's first virtual idol out of its heroine, Akemi Hayashi. While the virtual humans of this period were mainly synthesized through 2D animation techniques, the means of synthesizing virtual humans has continued to develop, from Computer Graphics(CG) technology to the large-scale application of motion capture and AI synthesis technology, and the level of intelligence of virtual humans has continued to increase. In addition to their use in animated films and games, the use of virtual humans has become more widespread. In addition, virtual humans are beginning to expand beyond the realm of virtual anchors and virtual idols, with more realistic realistic-like virtual humans and AI-synthesized real people, among others. The emergence of these virtual humans has not only increased the number of ways people can be entertained but has also opened up more possibilities for various industries.

Beginning in the 1980s and continuing into the next few centuries, virtual human technology is constantly evolving. The development process can be summarized in four stages, as shown in figure 2.

C. INFRASTRUCTURE OF THE VIRTUAL HUMAN

Virtual human systems typically consist of five main modules: character generation, voice generation, animation generation, audio and video synthesis display, and interaction. The details are shown in figure 3.

1) CHARACTER IMAGE GENERATION

Depending on the dimensions of the character graphic resource, character images can be divided into two categories: 2D and 3D. In terms of appearance style, they are further subdivided into cartoon, anthropomorphic, realistic, and hyperrealistic types. Model designers and modelling tools use a variety of modelling techniques to create characters in a variety of styles, including mapping, texturing, and lighting, to enhance character fidelity and visual appeal.

2) SPEECH GENERATION

This module generates the corresponding character's voice based on text, using natural language processing and speech synthesis techniques such as text-to-speech (TTS) and voice style migration to generate speech that adapts to various scenarios so that the digital person has a natural, fluent, and personalized voice expression.

3) ANIMATION GENERATION

This module generates character animations based on text content using computer animation techniques and

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

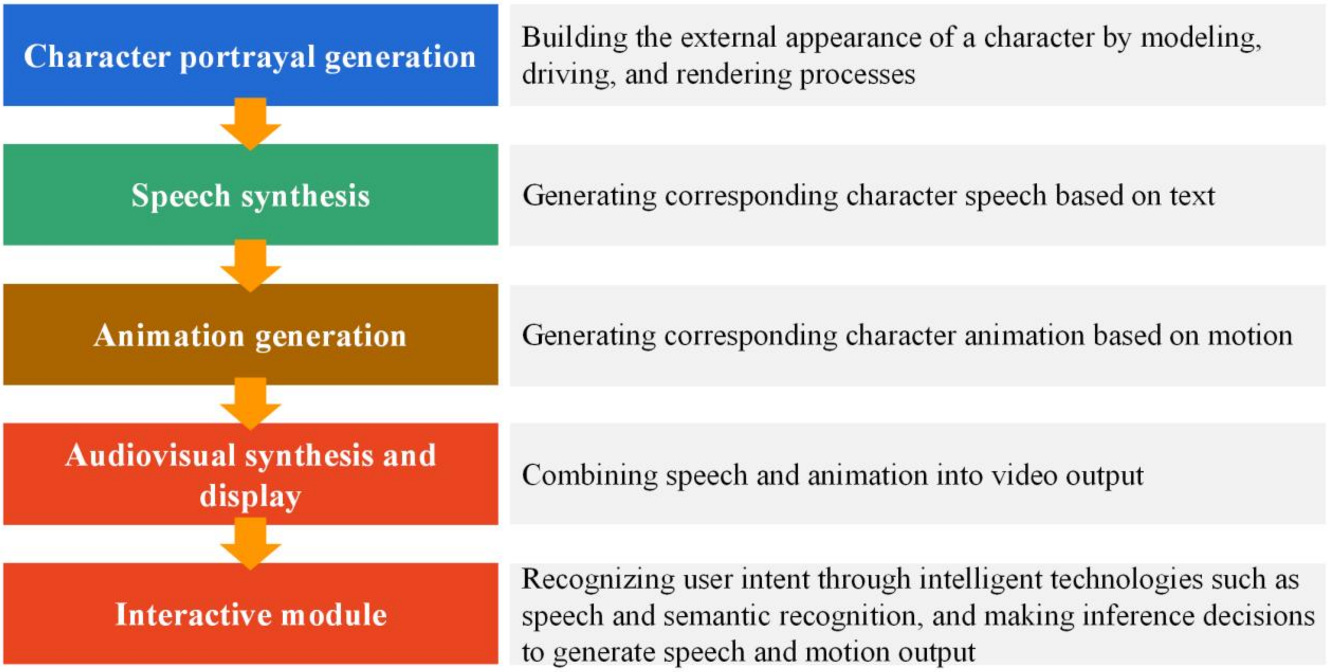

FIGURE 3. **The infrastructure of the virtual man**

character motion capture. The expressions, movements, and poses of the characters are created and driven to provide the virtual human with realistic and interesting dynamic effects in various scenes.

4) AUDIO AND VIDEO SYNTHESIS DISPLAY

In this module, the system uses video editing and rendering technologies to combine the generated voice and animation to form a complete and high-quality video. This is then displayed to the user, giving the digital person a realistic and rich audio and video effect to enhance user experience.

5) INTERACTION MODULE

The interaction module provides the digital human with the ability to interact with a user. By using intelligent technologies, such as speech recognition, semantic understanding, and emotion analysis, the system can capture the user's intent, understand it, and respond to it. Based on the current intent of the user, it determines the next voice and action of the virtual human. Responses were made accordingly, and the character was adjusted to the context for the next round of interaction, making the dialogue richer and more varied. Through the synergy and continuous optimisation of these five modules, the virtual human system provides a vivid, natural and interactive experience for a wide range of business and entertainment applications.

D. MAIN CLASSIFICATION OF VIRTUAL HUMANS

Currently, there are three main categories of virtual humans: identity-based virtual humans, service-based virtual humans, and virtual idols, as shown in TABLE 1. Real identity-based virtual humans are mainly digital images built around real people in virtual space, and are currently used in fewer cases, only in real-life celebrity brand endorsements and marketing promotions. Although real identity virtual humans are rarely used in current applications, we can foresee that they will become more widely used in digital applications in the future, such as in virtual meetings and distance learning. It can therefore be said that real-identity virtual humans are an area of great potential.

Currently, the most popular classification are service-oriented virtual humans and virtual-idol virtual humans. Virtual service humans, such as virtual customer service, virtual shopping guides, and virtual narrators, are mainly used to replace real people to provide basic customer services, such as consultation and announcements, and are relatively unexplored in creating virtual Intellectual Property(IP)s. However, virtual service humans have a wide range of applications, for example, in retail, finance, and healthcare industries, where they can help companies better serve their customers and improve customer satisfaction.

Virtual idols cater to consumer preferences through customized images and voices, build a fan economy through in-depth IP operations, and achieve commercial realization and brand promotion through brand promotion, participation in programs, live broadcasts, and peripherals. Virtual idols have become an independent cultural phenomenon and represent an important development trend in the digital entertainment industry.

TABLE 1. **Units for Classification of Virtual Human.**

Classification	Real identity virtual human	Virtual IP-based identity virtual human	Service-oriented virtual human
Positioning	A second doppelganger of a real-world character in a virtual world	Virtual characters with virtual IP, iconic identity characteristics	Virtualisation of real-world service roles such as narrators, anchors and staff
Core competencies	Meeting the needs of individuals for virtual identities in metaverse social, entertainment and service scenarios	Cashing in on the fan economy through branding, acting in shows, live streaming, peripherals, etc.	Provide consultation, demonstration and explanation to users through text and voice interaction
Interactive or not	Yes	No	Yes

III. ACADEMIC RESEARCH STATISTICS FOR VIRTUAL HUMAN

A.STATISTICAL STUDY OF THE VIRTUAL HUMAN LITERATURE

In order to provide better statistics on the status of research on virtual-human, we conducted a literature study of journal literature with the theme "v irtual-human" in the WoS (Web of Science) database, systematically reviewing the main contents and development trends of domestic and international research fields, and providing a reference basis and experience for virtual human research and practice. This paper provides an empirical reference for virtual human research and practice, and helps us to map the current status of virtual human scholarship. This article was searched for in the Web of Science (WoS) database with an advanced search conducted before the deadline of April 13, 2023, to find journal articles with "virtual human" in the subject, keyword, or title. The search yielded 615 articles, excluding book reviews, prefaces to columns, unauthored articles, and other articles irrelevant to the topic, resulting in 583 valid articles.
This section investigates the changes in the number of virtual human related scholarly results published in the WoS (Web of Science) database since 2006, as shown in Figure 4. From 2012 to 2018, there was no significant increase in the number of relevant academic results. From 2018 onwards, the number of scholarly papers on virtual humans began to steadily increase and showed a higher growth rate from 2020 onwards, and the overall number of studies is steadily increasing.

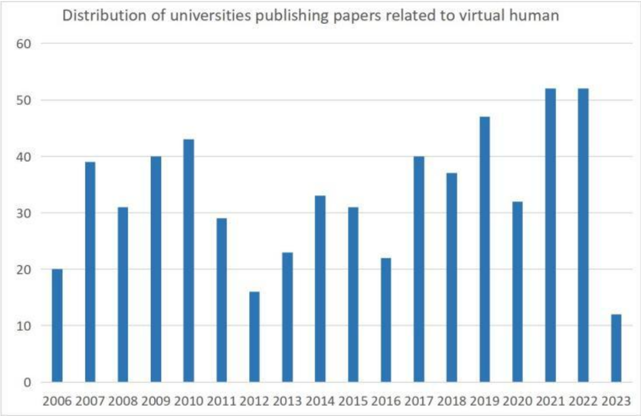

FIGURE 4. **The number of virtual human related scholarly results published in the WoS (Web of Science) database since 2006**

From the classification of virtual human related academic papers in the WoS (Web of Science) database, Computer Science Artificial Intelligence, Computer Science Software Engineering, Computer Computer Science Artificial Intelligence, Computer Science Software Engineering, and Computer Science Theory Methods are the main areas of research related to virtual human, as shown in figure 7.

The most cited papers under the Computer Science Artificial Intelligence category are DeVault et al. "SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support", which presents an implementation-ready virtual human interviewer designed to create an engaging face-to-face interaction where the user feels comfortable talking and sharing information. The aim is to create interactive contexts that facilitate the automatic assessment of indicators of distress, defined as verbal and non-verbal behaviors related to depression, anxiety or post-traumatic stress disorder (PTSD) [18]. The most cited papers under the Computer Science Software Engineering category are Papagiannakis et al. "Mixing virtual and real scenes in the site of ancient Pompeii", which presents a method aimed at capturing real scenes in

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

an immersive, fully mobile augmented reality (AR) environment based on a video perspective captured/real-time video sequences of real scenes in an HMD setting, with scenes enhanced by seamless and accurate real-time alignment and 3D rendering of realistic and complete simulations of virtual human and plants in a real-time storytelling-based scene [19]. The most highly cited paper under the Computer Science Theory Methods category is Kenny, P et al. "Virtual patients for clinical therapist skills training" describes an approach to building virtual human technology and applying it to the virtual patient domain, and describes the iterative design process of the approach and preliminary results to suggest that virtual patients may be a useful adjunct to psychotherapeutic education [20].

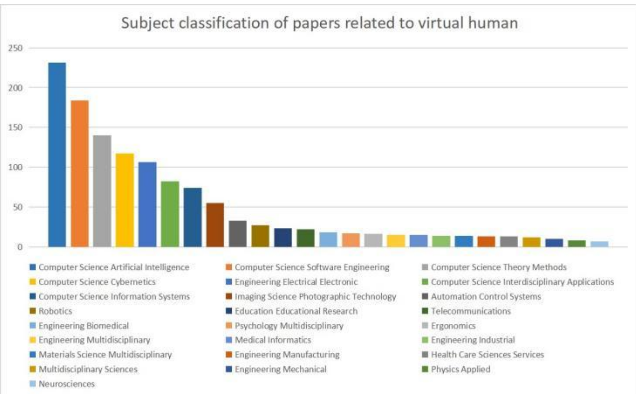

FIGURE 5. Classification of research directions in the subject of virtual human

B.LATEST RESEARCH RESULTS IN VIRTUAL HUMAN TECHNOLOGY

1) SPEECH SYNTHESIS

As an important component of virtual humans, the speech synthesis module plays the role of translator for the entire system. The speech synthesis module consists of two main parts: Automatic Speech Recognition (ASR) and text-to-speech (TTS). Mehra et and Susan, S proposed a new deep bidirectional long short-term memory model based on self-attention (SA-deep BiLSTM) [21]. By integrating the posterior probability of the SA-deep BiLSTM model for a single modality, this model achieved 80.10% accuracy for 98 word categories in Liu and proposed a new method to search for faster architectures using neural architecture search (NAS) [22]. This method effectively improves the inference speed of existing end-to-end ASR models on two commonly used Mandarin datasets with performance compared with the best human-designed baseline, the maximum acceleration achieved through CPU/GPU Zhao, W and Yang, Z proposed a new Emo-VITS system based on the high-performance speech synthesis module VETS [23]. This system extracts global and local features of reference audio through an emotion network, and then fuses emotional control of text-to-speech synthesis. This system extracts global and local features of reference audio through an emotion network, and then fuses global and local features through an attention-based emotion feature fusion module, thereby achieving more accurate and comprehensive emotional speech synthesis. Mu et al. reviewed deep learning-based end-to-end speech synthesis techniques and analysed their advantages and limitations through a systematic review [24]. Yu et al. proposed a new end-to-end neural generation TTS model based on the mechanisms of Deep-Inherited Attention (DIA) and adjustable Local Sensitivity Factors (LSF) [25]. The inheritance mechanism allows multiple iterations of DIA by sharing the same training parameters, which tightens the token-frame correlation and In addition, LSF is used to enhance context connections by expanding the concentration area of the DIA. In addition, Ding provides a systematic review of the development of speech synthesis, exploring the technological advances and challenges in recent years [26]. Kang et al. present a text-to-speech synthesis approach adapted to any speaker using diffusion modelling, with a focus on improving the naturalness and fluency of speech [27].

2) 3D FACE RECONSTRUCTION

Currently, the basic construction of virtual human images is mainly achieved through three methods: manual modelling, image capture modelling, and instrument capture modelling. Although manual modelling is still widely used, it has a long production cycle. Image capture modelling reconstructs the 3D structure of a face using multiple photos; however, the accuracy is insufficient. Most 3D facial reconstruction techniques are based on 3D Morphable Models (3DMM). The mainstream 3D facial reconstruction technology is based on deep networks, using CNN to learn the mapping from facial images to 3DMM parameters and utilizing synthetic data for data augmentation [28]-[30].

Instrument capture modelling is currently the focus of development, which can achieve a precision of up to 0.1 millimeters, but costs more. Scanning models are currently the main method in instrument capture modelling, among different scanning solutions, structured light scanning and reconstruction systems [31] are comparatively economical. These systems utilize projectors to project specific light sources and cameras to capture information, enabling the restoration of complete 3D models. The equipment requirements for these systems are relatively low. Camera array scanning and reconstruction technology [32]-[33] has replaced structured light as the mainstream. The equipment requirements were relatively low. This technology uses a camera array to take pictures, match, and calibrate the same feature points, and finally reconstructs the character model. This technology has been successfully commercialized and widely used in film and game production. Awan et al. proposed a customised convolutional neural network for facial emotional expression classification with a focus on improving classification accuracy [34].Zhao et al. proposed a deep block network for facial expression recognition and migration, especially in complex contexts [35].Shen et al.

introduced a multi-channel attentional residual network based facial expression recognition method with a focus on improving classification accuracy.The focus is on improving the accuracy of recognition [36].

3) FACE EXPRESSION MIGRATION

Facial expression is a key component of virtual humans; however, traditional methods have a limited ability to deal with facial details, resulting in poor image quality. Facial expression migration refers to the use of computer technology to transfer facial expressions from one photograph to another, in order to realistically replicate the facial expression of a digital person. The main method currently used is Generative Adversarial Networks (GANs), which automatically generates realistic digital human facial expressions by learning features from a large sample of datasets. The GANs model consists of a generator, which generates realistic digital human facial images from random noise, and a discriminator, which distinguishes between real images and those generated by the generator. Xi Chen and Hongdong Zhao proposed a novel reconstruction model based on GAN for a single-image super-resolution reconstruction task, which uses low-resolution images as input and outputs high-resolution images. They demonstrated that spectral normalization and attention mechanisms were highly effective in stabilizing the training of GANs [37].

Currently, there are two main approaches for implementing facial expression transfer: traditional algorithms and deep learning algorithms. Traditional algorithms mainly rely on image gradient-based methods and feature offset-based methods [38]-[40] by combining deformations of facial expressions from multiple people and simplifying the mesh to change facial expressions. However, these methods generate images with lower quality and inconsistent expressions. In recent years, deep learning has made great progress in the fields of computer vision and graphics, such as Groshev, A et al. who used FaceShifter architecture as a baseline method [41] and proposed several major architectural improvements, including new eye-based loss functions, face mask smoothing algorithms, new facial exchange pipelines for image-to-video facial transmission, new stabilization technology for reducing adjacent frame facial jitter, and super-resolution stages. Zhang et al. reviewed 3D facial reconstruction methods for face-related product design and analysed their applications and challenges [42]. Zhao et al. introduced a facial manipulation-based 3D facial reconstruction method focusing on improving the accuracy and usefulness of the reconstruction [43]. Wu et al. explored the methodology of 3D facial reconstruction from static images and the issues related to computer standardisation [44].

4) CONVERSATION ENGINE

The dialogue engine is one of the core technologies that enable human-computer interaction, comprises modules such as ASR (Natural Speech Recognition), NLU (Natural Language Understanding), DST (Dialogue State Tracking) and NLG (Natural Language Generation). The NLU module analyzes user questions and extracts their meanings. Finally, the NLU module generates natural language responses based on the conversation goals and the current state, and completes the interaction with the user. Deep learning technologies can be used to achieve more intelligent, flexible, and natural dialogue interactions. Wei et al. proposed an long short-term memory RNN fused social network (LSTM-SN) classification method [45] based on an extremely complex dataset, which compresses the features of the dataset, combined with social network methods derived from a specific dataset to accomplish the classification task, and then uses a complex network structure evolution method to discover dynamic social attributes. The experimental results show excellent performance for many NLU tasks based on tricky data with loose and unbalanced features (including text classification tasks). Dong et al. provided a comprehensive review of nearly two decades of natural language generation (NLG) research [46], in particular, data-to-text generation and text-to-text generation deep learning methods and new applications of NLG techniques, with the aim of providing an update on deep learning research on core NLG tasks and the architectures employed in the field. Li et al. proposed a correctable state prediction method for dialogue state tracking [47], constructing a two-stage time-gap-valued prediction process consisting of original prediction and corrected prediction, where the original prediction process jointly models the previous dialogue state and dialogue context to predict the original dialogue state for the current dialogue round.

In terms of recent research, Chakraborty et al. present a method for classifying intentions of user dialogues using BERT, with a focus on improving the response accuracy of dialogue systems [48].Mollick et al. explore the transformative potential of generative AI, particularly in dialogue systems and AI interactions [49].

5) FACE-DRIVEN

Voice-driven face technology in the digital human rendering module is one of the key technologies for achieving natural virtual interaction. The technology uses a 3D deformation model (3DMM) to decode and reconstruct the input speech signal, and then uses the Wav2Lip model to realise the lip drive [50], while converting the 3DMM mesh to a real image through a drive rendering system consisting of a neural renderer [51]. This technology enables real-time rendering with a high degree of accuracy and precision, which can significantly enhance the communication and interaction experience between virtual humans and humans, bringing a better outlook for the development of virtual human. Xie et al. explored a non-invasive 3D facial reconstruction technique that utilises millimetre wave signals for facial reconstruction, providing a new approach to face-driven technology [52]. Liu et al.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

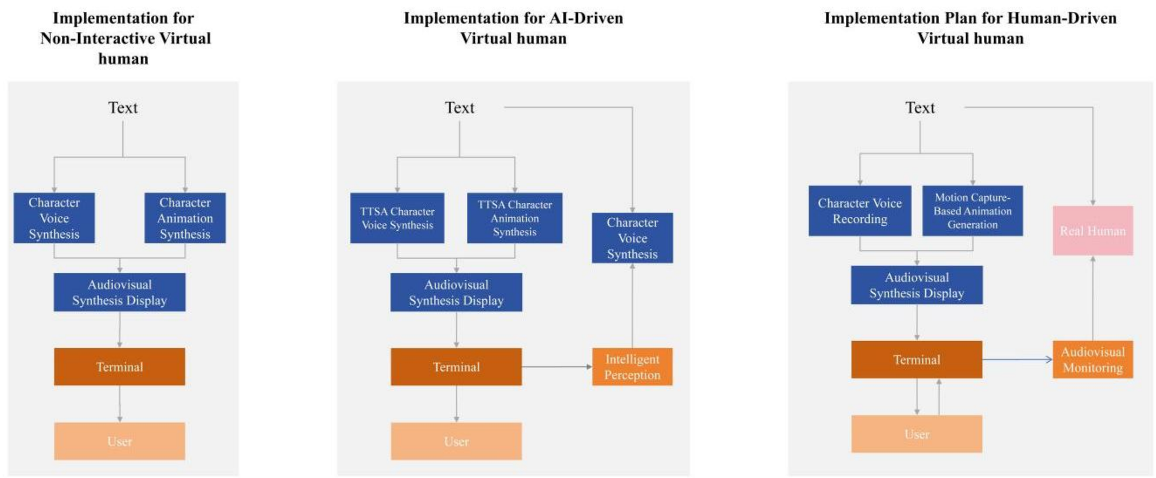

FIGURE 6. **Virtual Human Application Scenarios.**

proposed a music-driven approach to expressive singing facial synthesis, with a focus on improving the naturalness and expressiveness of facial animation [53].

6) MOTION-DRIVEN

At present, capturing the action to migrate to a virtual human is the main method of a virtual human drive module. The core technology is motion capture, through the optical type [54], inertial type [55], electromagnetic type [56], and computer vision-based motion capture [57]. At this stage, optical, computer vision-based, and inertial motion captures dominate. Computer vision-based motion capture is currently less accurate, but has low environmental requirements, a large movable range, imaginative use of scenarios, and consumer-level applications. Bian et al. introduce a motion-driven human-ai collaborative music composition and performance system, with a focus on improving the naturalness and creativity of music composition and performance [58]. Wang et al. propose a method for motion-driven tracking via end-to-end coarse-to-fine verification, with a focus on improving the accuracy and stability of tracking [59].

IV. APPLICATION FOR VIRTUAL HUMAN

A. MAIN IMPLEMENTATION OF VIRTUAL HUMAN

In the digital era, the technology and implementation of 'virtual humans' have become a hot research topic. According to their implementation mechanisms and interaction modes, we can classify them into three categories, as shown in figure 6:

1) NON-INTERACTIVE VIRTUAL HUMAN IMPLEMENTATION

This type of virtual human is mainly based on pre-set target texts to generate corresponding task speech and animation. The core idea is to convert textual information into audio and video formats, and then display them to users through various terminal devices [60]. The advantage of this method is its simple process and easy implementation, but its disadvantage is also very obvious, that is, it lacks real-time interaction with users [61]. Therefore, it is more suitable for application scenarios that do not require real-time feedback, such as advertising, educational videos [62]-[63], etc.

To summarise, this type of virtual human is more suitable for application scenarios that do not require real-time feedback. As its core lies in transforming predefined textual information into audio and video forms, it is particularly suitable for areas such as advertisements, educational videos, online courses, and news broadcasts. In these scenarios, the content is fixed and does not require real-time interaction with the user. In addition, this technology is also suitable for industries that do not have high technical requirements but need to generate avatar content quickly and easily.

2) INTELLIGENT-DRIVEN VIRTUAL HUMAN IMPLEMENTATION

Unlike non-interactive virtual humans, intelligent-driven virtual humans have stronger interactivity. It relies on advanced AI technology to read, parse, and recognize external textual or audio information [64]. Based on these parsing results, the system will decide the next action of the virtual human. In addition, the models of this type of virtual human usually require pre-training with AI to enable them to generate corresponding speech and animation through text, thereby interact with users in real-time [65]. To enhance their interactivity, researchers have also configured preset Q&A knowledge bases for their dialogue systems. The advantage of this method is its strong interactivity, which can provide users with a more realistic experience [66], but its implementation difficulty is relatively high, requiring refined modelling and key point binding of the original model [67].

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

TABLE 2. Virtual Human Application Scenarios.

Scenarios	Description	Roles
Film & TV	Virtual characters in commercial blockbuster shoots	Virtual actors
Media	Virtual hosts/anchors/idols to create brand-specific IP images, enable audience interaction and optimise the viewing experience	Virtual hosts, virtual anchors, virtual icons
Games	Virtual digital human game characters give gamers a greater sense of immersion	Digital roles
Finance	Customer-centric, intelligent and efficient human services through roles such as intelligent financial advisors and intelligent customer service	Intelligent Customer Service
Cultural Tourism	The virtual digital person can be displayed as a scenic spot employee, cultural and tourism ambassador, etc. on short videos, virtual live broadcasts, interactive screens and other platforms	Virtual Guides, Virtual Docent
Medical Field	Based on a real persona, it can provide a sense of care and authenticity to patients, offering counseling, care and companionship in generic scenarios or in specific care-type scenarios	Family doctor
Retail	Pre-promotion of retail merchandise, live-streaming with merchandise, virtual try-ons, brand showcases, virtual human and live anchors interacting with viewers and introducing merchandise during e-commerce live-streaming	Virtual anchors

To summarise, due to its powerful interactivity and real-time feedback capabilities, intelligence-driven virtual human is particularly suitable for customer service, online consulting, e-commerce, gaming, entertainment, education and other fields. In these scenarios, users may have a variety of questions or needs, and the intelligence-driven avatars can provide real-time answers or services to users based on their preset knowledge base or AI technology. In addition, this technology is also suitable for industries that require in-depth interaction with users or need to provide personalised services.

3) HUMAN-DRIVEN VIRTUAL HUMAN IMPLEMENTATION

This is a brand new way of implementing virtual humans, and its core is to use real humans to drive the behavior of virtual humans [68]. Specifically, the system will capture real-time video information of users through video monitoring devices, and interact with users through voice. At the same time, through motion capture technology, the system can capture the movements and expressions of real humans, and map them to the image of virtual humans [69]. It has the advantage of highly realistic interactions, and because it is driven by real people, the interaction experience is more realistic and natural. It is also capable of in-depth dialogue and complex interactions compared to intelligent-driven virtual human [70], which can be instantly adapted to the actual situation. However, it requires real-time operation by a real person, which is relatively costly and limited by the time and energy of the real operator [71]. It also requires high-precision motion capture equipment and high-performance rendering technology [72]-[73].

In summary, this implementation is particularly suitable for scenarios that require the provision of a highly realistic, immersive experience, such as in the fields of virtual reality, augmented reality, online entertainment, cinema, television, and live broadcasting. In these scenarios, the movements and expressions of a real person can be captured and mapped onto the virtual human image, thus providing the user with an interactive experience that is virtually indistinguishable from that of a real person. In addition, this implementation is also suitable for industries that require highly personalised and emotional services, such as psychological counselling, healthcare and education.

Each virtual human implementation has its own unique advantages and application scenarios. Which technology to choose depends on specific needs, budget, technology level and other factors. But in any case, we have reason to believe that virtual human technology will play an increasingly important role in various fields in the future.

B. MAIN APPLICATION SCENARIOS FOR VIRTUAL HUMAN

Virtual humans can be widely combined with various industry sectors to replace real people for services, content production, entertainment and social activities, with the advantages of cost reduction and efficiency, and autonomy and control. Driven by technological upgrades, the form and function of virtual humans are constantly improved, developing towards refinement, intelligence, efficiency and diversification, and application scenarios continue to be broadened, as shown in table 2. Many references currently only introduce the application scenarios of virtual humans in their proprietary fields, while this paper will provide a detailed and systematic definition and classification of the existing application scenarios of virtual humans.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

1) VIRTUAL HUMAN APPLICATIONS IN FILM AND TELEVISION

While virtual humans are currently used in many industries, the most widespread use of virtual human technology is in films and televisions. The advancement and maturity of virtual human technology has led to an increasing number of films and TV shows featuring virtual humans, and the use of virtual humans have also contributed to the development of films and the television industry. The use of virtual humans has also contributed to the development of the film and television industry by stimulating the imagination of film and television creations. For example, in The Avengers, an imaginative film, the exterminator is a hallmark application of virtual human technology to films and televisions. In addition, virtual human can help directors achieve content and effects that cannot be represented in realistic filming and have now become an important technical tool and selling point in the filming of special effects commercial blockbusters. Thalmann et al. presented an integrated and functional system for digital actor animation that described, in particular, the integration of motion control techniques [74]. Al-Jawaheri et al. presented a review of efficient computerised facial animation techniques, presenting two basic techniques for generating extreme 3D facial animated expressions that are used in PC games, movies, and intuitive visual and sound applications [75].Yang et al. used an inertial sensor-based gesture capture device and accompanying software to capture people's dance movements through sensors and controllers.By creating virtual character models in 3ds Max software and virtual stage models in Cinema 4D software, highly restored virtual characters and character dance animations were produced, and the optimised character models were combined with scene models to generate virtual character dance animations in virtual scenes for film and television animation [76].

2) VIRTUAL HUMAN APPLICATIONS IN THE MEDIA SECTOR

In the era of artificial intelligence, the media industry has been completely reconfigured and disrupted, and mainstream media organizations have invested in exploring applications related to virtual humans, with digital "media person" applications such as virtual hosts, virtual anchors, virtual journalists, and virtual cultural promoters. In addition to generating videos from audio and text content with a single click, which enables the rapid and automated production of program content, real-time interaction with viewers, and optimized viewing experience, enterprises are creating brand-specific Intellectual Property(IP)s through the creation of virtual humans to enhance their brand image and visibility [77]. Yen et al. define and classifie virtual idols, explaining that virtual idols are highly profitable, malleable, and controllable compared to traditional idols, and that many companies launch virtual idols to increase efficiency and profits [78], explores the mediating and moderating role of consumers' brand image, perceptions, and word-of-mouth perceptions of virtual idols, suggesting that companies should work to enhance the brand image of virtual idols and allow consumers to perceive the highlights and characteristics of virtual idols. The production of virtual idols began to replace real idols to satisfy fans' needs for the perfect idol of their desires, and rahmi et al. aimed to describe the post-human value of virtual idols as virtual products mixed with human reality, such as Hatsune Miku, who exhibited post-human values in appearance and voice [79].

3) VIRTUAL HUMAN APPLICATIONS IN THE GAMING SECTOR

The gaming industry is one of the original application areas for virtual humans and has a much deeper deposit in virtual human technology, research, and development. Users are made to feel more immersive by increasingly realistic virtual human game characters, which also enhance the playability of the game. In action-based VR games, where players see digital virtual humans from a first-person perspective and virtual humans replicate the user's hand, head, and body movements, lugrin et al. explored the possible impact of the visibility of virtual human body parts on player experience and performance in current virtual reality gaming platforms [80]. Ferrer et al. converted this educational simulation into an effective educational game by introducing the gameplay elements of an interactive virtual human. Extending the scope of passive solar education and illustrating through user studies how the inclusion of interactive virtual humans affects the perception of user learning [81]. Arroyo-Palacios et al. developed an interactive game system in which participants interacted with a virtual human character through their psychophysiological activity, and investigated the assessment of a physiologically aware virtual human mobile game as a method of regulating participants' emotional and physiological states [82]. The results showed that by using the game with a physiologically aware and emotionally charged virtual human, participants were able to evoke their emotions in an activation condition and relax themselves in a relaxation condition.

4) VIRTUAL HUMAN APPLICATIONS IN THE FINANCIAL SECTOR

The application of virtual humans in the financial sector is the most notable through roles such as intelligent financial advisors and intelligent customer services, enabling customer-centric, intelligent, and efficient human services. Priya M et al. argued that customer service in the financial sector is crucial and that virtual customer assistants generated by virtual human technology revolutionize the way users interact with machines [83], can support customers in performing operations such as repaying money and trading stocks, and provides a toolkit that gathers best practices for designing and developing virtual customer assistants in the financial sector to ensure a high-quality user experience. Q. Shao et al. argued that intelligent financial advisors in online financial applications bring new dynamism to personal investment in the financial sector by providing users with appropriate and high-quality investment portfolios [84]. One of the main tasks when introducing a service virtual human into innovative investment projects such as financial institutions is to assess the economic benefits of the project. Bataev et al. proposed a model based on a total cost of ownership approach to assess the effectiveness of service virtual humans used together with artificial intelligence systems to ensure customer and financial institution interaction, and found that the introduction and use of legal problem areas implemented by financial institutions with artificial intelligence virtual human systems would result in 21.9% cost savings [85].

5) VIRTUAL HUMAN APPLICATIONS IN THE CULTURAL AND TOURISM SECTOR

Virtual humans have been widely used as a portal for interaction between virtual and real worlds for the interactive presentation of virtual cultural heritage environments and exhibitions, a new way of integrating culture and technology that is of great significance to the innovative development of the cultural tourism industry. The deployment of virtual humans in virtual cultural heritage environments is thought to enhance the interactivity and playability of reconstructed scenes [86]. In a museum visit experience, virtual humans can act as storytellers, curators, guards, tourists, personal guides [87], and take on the role of guiding visitors while providing a wealth of information in a multimedia format to enhance the visitor experience. In another application [88], a virtual human provides users with a context for virtual exhibitions through legends, stories, poetry, rituals, dances, and customs and can also be used to preserve and simulate cultures and teach crafts [89]. Research has shown that virtual humans can influence the virtual experience in the cultural tourism industry, stimulating the attention and engagement of the audience [90], thus making the stories presented in the virtual environment more believable and having a positive and constructive impact on the user's feelings during the tour.

6) VIRTUAL HUMAN APPLICATIONS IN THE MEDICAL FIELD

The use of virtual humans in the medical and healthcare industries is significant, with an increasing number of healthcare organizations utilizing virtual human technology in a variety of forms to provide better services to patients. Artificial intelligent virtual humans have capabilities such as voice recognition, natural speech, and visual interaction to enhance natural human-computer interaction. They can incorporate sensors to monitor environmental behaviors, collect information on a patient's healthcare, and act as a companion and family doctor. The virtual human enables family companionship, provides psychological counseling, and offers timely coping advice for the physical and mental health of family members [91], it describes the application of virtual human technology to virtual patients for mental health diagnosis and clinician training and discusses possible ways in which virtual humans can be used to assist in healthcare and integrate multimodal inputs to enhance the application of virtual human systems in the medical field. In diagnostic medicine, virtual human anatomy software allows users to simultaneously study 2D anatomical cross sections and reconstruct 3D views of actual images, facilitating undergraduate medical students' learning of 2D image skills for interpreting cross-sectional anatomy [92]. Virtual human technology and lens model methods can be used as alternatives to extend pain-related decision-making investigations [93], and simulation studies have used novel virtual human technology to investigate differences in pain-related clinical decision-making. In addition, virtual humans are often used to help medical students practice their communication skills [94]. By creating and testing virtual human patients, physicians can be taught ways to communicate in patient encounters, including empathy, among others, as well as providing simulation and practice opportunities for medical history, critical thinking, and empathic communication [95].

7) VIRTUAL HUMAN APPLICATIONS IN THE RETAIL SECTOR

Virtual humans have become a new application in brand marketing, bringing more exposure and marketing effectiveness to brands. Brands create virtual humans to present brand images and product features, providing consumers with a more visual, vivid, and three-dimensional marketing effect [96]. Virtual humans offer the promise of personalized shopping experiences and 24/7 online services that can reduce human workload and provide seamless customer experience. The objective of this study was to quantify consumer attitudes towards the propensity to interact with virtual humans to help fashion businesses seek to diversify their operations [97], and the results highlight the important potential impact of consumer responses to virtual human formats on both secular and digital solution decisions in the fashion retail business. Virtual humans can cut through scenarios such as personalized marketing and unmanned shops to build new ways of delivering retail services. Virtual anchors can join forces with live anchors to interact with viewers and introduce and showcase products. In addition, the study demonstrated the importance and role of the digital body in the accuracy of consumer body measurements in retail stores [98].

V. THE ADVANTAGES AND CHALLENGES OF VIRTUAL HUMAN

With the rapid development of artificial intelligence and other related technologies, virtual digital human technology has gradually become a hotspot for research and application. Owing to their high degree of flexibility and customizability, virtual digital humans can be widely used in many fields. At the same time, however, there are some disadvantages and limitations to virtual digital human technology. In the following section, we discuss the advantages and disadvantages of virtual digital human technology, as shown in figure 8.

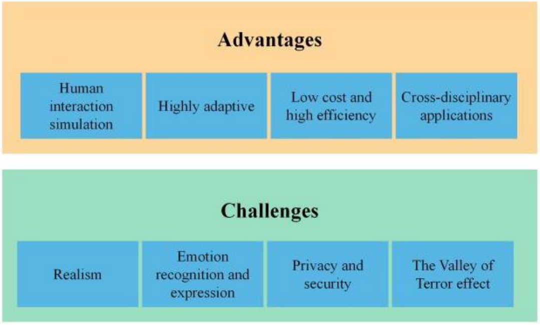

FIGURE 7. **The advantages and challenges of virtual human**

A. ADVANTAGES OF VIRTUAL HUMAN

Virtual human technology provides users with a natural, comfortable, and personalized experience by simulating human communication and interaction behaviors. Using artificial intelligence and machine-learning technologies, virtual humans can be highly adaptive and provide personalized and practical services to users, they offer significant advantages in terms of cost reduction and efficiency compared with traditional human services. In addition, technology has the potential to be used in a wide range of fields, offering opportunities for technological advancement across a wide range of industries.

1) HUMAN INTERACTION SIMULATION

Virtual human technology can simulate real human communication and interaction behavior, providing a natural, comfortable, and personalized experience for users, effectively meeting their needs in scenarios such as chatting, consulting, and training [99].

2) HIGHLY ADAPTIVE

By applying artificial intelligence and machine learning technologies, virtual humans can adapt to the user's behavior, habits, and preferences [100], providing more personalized and practical services.

3) LOW COST AND HIGH EFFICIENCY

Compared to traditional human customer service, virtual humans incur no manpower costs and can keep operating costs low while meeting the needs of a large number of users.

4) CROSS-DISCIPLINARY APPLICATIONS

Virtual human technology can be used in a wide range of fields, and technological developments and innovations in each field can build on and promote each other [101], creating an interdisciplinary dynamic of technological progress.

B. THE CHALLENGE OF VIRTUAL HUMAN

There is still room for development of virtual human technology in terms of realism and natural language processing. In the process of improving realism, issues related to the “uncanny valley” effect need to be addressed. In addition, the improvement in emotion recognition and expression remains an unresolved challenge. At the same time, virtual human technology needs to focus on and address the privacy and security issues associated with the handling of user data. By addressing these challenges, virtual human technology will continue to improve and provide more innovation and convenience to human society.

1) REALISM

Virtual humans do not yet fully simulate human appearance and behavior and may still differ from realistic interactions. For example, the realism of facial movements, emotional expressions, and body movements could be improved.

2) EMOTION RECOGNITION AND EXPRESSION

To achieve a realistic and satisfying interactive experience, virtual humans must be able to recognize and express emotions. However, current technologies still struggle to accurately capture and understand the emotional needs of users [102] and respond accordingly, and some biases still exist.

3) PRIVACY AND SECURITY

Virtual humans must handle large amounts of user data, such as voice and behavioral habits. The collection, storage, and analysis of data often involves user privacy and security issues [103]. Therefore, data protection and information security are important considerations in the development and application process.

4) THE “UNCANNY VALLEY” EFFECT

As a virtual human gradually approaches human appearance and behavior, it may cause discomfort to the observer. This phenomenon is known as the valley effect of the error. When a virtual human reaches a significant level of realism, the nuances that may still exist can be disconcerting [104]. The challenge of tackling this phenomenon requires finding a balance between maintaining realism and avoiding over-simulation to reduce the potential valley of the terror effect.

VI. TRENDS IN VIRTUAL HUMAN

A. TRENDS IN VIRTUAL HUMAN TECHNOLOGY

Virtual human technology is a comprehensive technology that involves many fields such as computer graphics, artificial intelligence. In recent years, with the rapid development of these technologies, virtual human technology has been widely used and has shown several technical development trends as follows.

1) EFFICIENT NEURAL NETWORK COMPRESSION AND OPTIMIZATION

Researchers are continuously developing new techniques such as neural network pruning and quantization to improve the computational efficiency of the underlying neural networks of virtual humans. These techniques improve training and inference speed by reducing network parameters and computational resource consumption while maintaining performance [105]. In addition, these optimizations will allow virtual humans to run more efficiently on edge devices,

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3329573

such as mobile phones and smart home devices, enabling a wider range of application scenarios.

2) SEMANTIC UNDERSTANDING AND MULTIMODAL INTERACTION

In future technological developments, virtual humans will be able to understand text, voice, expressions, and other multimodal information in greater depth, thereby enabling a more natural interaction experience. To further improve the quality of interaction, a virtual human can use emotion recognition technology to determine a user's emotional state, analyze user needs, and provide corresponding feedback. Additionally, the interaction effect can be further enhanced by introducing sensing technologies such as vision and haptics [106].

3) REAL-TIME RESPONSE OF VIRTUAL HUMAN BASED ON EDGE COMPUTING

Edge computing technologies can reduce latency and improve real-time performance by migrating computing tasks from the cloud to devices at the edge of the network closer to the user. When virtual humans adopt edge computing, they can respond faster to user needs and enable real-time interaction. This will provide more powerful support for virtual humans in areas that require immediate responses, such as telemedicine, education, and entertainment.

B. INDUSTRY TRENDS FOR VIRTUAL HUMAN

1) INTEGRATION OF VIRTUAL HUMANS WITH THE INTERNET OF THINGS (IOT)

IoT technology is penetrating many fields, such as smart cities and smart transportation, where virtual humans can serve as human-machine interaction interfaces [107] and provide users with unified intelligent management and control. For example, in smart home scenarios, virtual humans can help monitor the operation of home appliances, automatically adjust indoor temperature and lighting, and intelligently voice control the home devices. At the same time, virtual humans will also be promising in smart city management and environmental monitoring, improving the efficiency of urban governance.

2) INTEGRATION OF VIRTUAL HUMANS WITH BRAIN-COMPUTER INTERFACE TECHNOLOGY

With advances in brain-computer interface technology, it may be possible for virtual humans to communicate with human users through brain waves. This direct and efficient method of communication removes the barriers posed by traditional means of interaction. For example, virtual humans could provide more efficient and personalized services in pragmatic scenarios for people with disabilities and also help to explore new approaches to mental health counselling [108].

3) APPLICATION OF VIRTUAL HUMANS IN THE METAVERSE DOMAIN

As the concept of metaverse gradually takes shape, virtual humans, such as personal assistants and virtual advisors, play key roles within the virtual space [109]. Users can freely customize their virtual humans in the metaverse, interact with other users, and jointly participate in various activities within the virtual world, such as shopping, gaming, entertainment, education, and socializing, thus enriching the virtual life experience.

4) INTELLIGENT CREATION AND ARTISTIC EXPRESSION

Virtual humans are expected to be used in film and television dramas, games, and other fields, bringing into play the unique value of technologies, such as artificial intelligence (AI) synthetic anchors and virtual actors. For example, virtual humans can appear in special film effects, animation production, and VR experiences, bringing a more realistic sensory experience to the audience. Additionally, in game development, virtual characters generated using AI technology can be freely and individually customized to enhance user immersion.

VII. CONCLUSION

Virtual human technology, with its features of virtualization, multi-technology integration, anthropomorphism, and plasticity, has sparked a surge in research in both academic and social applications. This article provides an introduction to the development background and fundamental concepts of virtual human technology. It examines the research conducted by numerous universities and research institutions on virtual human technology, and presents an analysis and summary of seven application scenarios, including film and television, media, games, and finance. Additionally, the article concludes by summarizing the advantages and challenges of virtual human technology, and proposes future development trends.

Based on the aforementioned analysis, the following conclusions can be drawn. In recent years, various universities have conducted extensive research on virtual human technology, resulting in an increase in the number and quality of published papers. Virtual human technology boasts a wide range of application scenarios, displaying strengths in human interaction, adaptivity, cost-effectiveness, efficiency, and cross-domain applications. However, challenges remain, such as the need to enhance the realism, accurately sense emotions, and address privacy and security issues. Therefore, a future development trend for virtual humans is proposed to tackle these challenges.

In terms of technology, the focus will shift towards efficient neural network optimization, semantic understanding, multimodal interaction, and real-time response based on edge computing, among other areas. In the industrial field, integration with Internet of Things technology, brain-computer interface technology, metaverse, virtual creation, and art will take place, fostering closer collaboration with different domains. The future development of virtual human technology relies on close cooperation between academia and industry, working together to promote the co-development of the digital world and economy.

REFERENCES

[1] M. Sela, E. Richardson, and R. Kimmel, “Unrestricted Facial Geometry Reconstruction Using Image-to-Image Translation,” in 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, pp. 1585–1594. doi: 10.1109/ICCV.2017.175.

[2] A. Tewari et al., “MoFA: Model-based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction,” in 2017 Ieee International Conference on Computer Vision (iccv), New York: Ieee, 2017, pp. 3735–3744. doi: 10.1109/ICCV.2017.401.

[3] L. Tran and X. Liu, “Nonlinear 3D Face Morphable Model,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA: IEEE, Jun. 2018, pp. 7346–7355. doi: 10.1109/CVPR.2018.00767.

[4] H. Kaushik, T. Kumar, and K. Bhalla, “iSecureHome: A deep fusion framework for surveillance of smart homes using real-time emotion recognition,” Applied Soft Computing, vol. 122, p. 108788, Jun. 2022, doi: 10.1016/j.asoc.2022.108788.

[5] D. Wang, X. Wang, and S. Lv, “An overview of end-to-end automatic speech recognition,” Symmetry, vol. 11, no. 8, p. 1018, 2019.

[6] Y. Lei, S. Yang, X. Wang, and L. Xie, “MsEmoTTS: Multi-Scale Emotion Transfer, Prediction, and Control for Emotional Speech Synthesis,” IEEE/ACM Trans. Audio Speech Lang. Process., vol. 30, pp. 853–864, 2022, doi: 10.1109/TASLP.2022.3145293.

[7] S. Li, B. Ouyang, L. Li, and Q. Hong, “Light-TTS: Lightweight Multi-Speaker Multi-Lingual Text-to-Speech,” in ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada: IEEE, Jun. 2021, pp. 8383–8387. doi: 10.1109/ICASSP39728.2021.9414400.

[8] J. Liu, X. Wang, Y. Tan, L. Huang, and Y. Wang, “An Attention-Based Multi-Representational Fusion Method for Social-Media-Based Text Classification,” Information, vol. 13, no. 4, p. 171, Mar. 2022, doi: 10.3390/info13040171.

[9] V. Raina, S. Krishnamurthy, V. Raina, and S. Krishnamurthy, “Natural language processing,” Building an Effective Data Science Practice: A Framework to Bootstrap and Manage a Successful Data Science Practice, pp. 63–73, 2022.

[10] S. Latif, A. Zaidi, H. Cuayahuitl, F. Shamshad, M. Shoukat, and J. Qadir, “Transformers in speech processing: A survey,” arXiv preprint arXiv:2303.11607, 2023.

[11] K.-C. Wang et al., “CA-Wav2Lip: Coordinate Attention-based Speech To Lip Synthesis In The Wild,” in 2023 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 2023, pp. 1–8.

[12] R. Eiris and M. Gheisari, “Research trends of virtual human applications in architecture, engineering and construction,” Journal of Information Technology in Construction (ITcon), vol. 22, no. 9, pp. 168–184, 2017.

[13] Bao D., “Exploration of virtual human running based on CCD algorithm,” Concurrency and Computation: Practice and Experience, vol. 31, no. 10, p. e4901, May 2019, doi: 10.1002/cpe.4901.

[14] T. S. Perry, “Leaving the uncanny valley behind,” IEEE Spectr., vol. 51, no. 6, pp. 48–53, Jun. 2014, doi: 10.1109/MSPEC.2014.6821621.

[15] M. Lycett and A. Reppel, “Humans in (Digital) Space: Representing Humans in Virtual Environments,” in Proceedings of the 2022 International Conference on Advanced Visual Interfaces, Frascati, Rome Italy: ACM, Jun. 2022, pp. 1–5. doi: 10.1145/3531073.3531172.

[16] P. Domingos and A. VEVE, “Our digital doubles,” Scientific American, vol. 319, no. 3, pp. 88–93, 2018.

[17] N. M. Thalmann, P. Kalra, and M. Escher, “Face to virtual face,” Proc. IEEE, vol. 86, no. 5, pp. 870–883, May 1998, doi: 10.1109/5.664277.

[18] D. DeVault et al., “SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support,” AAMAS’14: PROCEEDINGS OF THE 2014 INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS & MULTIAGENT SYSTEMS. ASSOC COMPUTING MACHINERY, 1515 BROADWAY, NEW YORK, NY 10036-9998 USA, pp. 1061–1068, 2014.

[19] G. Papagiannakis et al., “Mixing virtual and real scenes in the site of ancient Pompeii,” Comput. Anim. Virtual Worlds, vol. 16, no. 1, pp. 11–24, Feb. 2005, doi: 10.1002/cav.53.

[20] P. Kenny, T. D. Parsons, J. Gratch, A. Leuski, and A. A. Rizzo, “Virtual patients for clinical therapist skills training,” INTELLIGENT VIRTUAL AGENTS, PROCEEDINGS, vol. 4722. in Lecture Notes in Artificial Intelligence, vol. 4722. SPRINGER-VERLAG BERLIN, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY, p. 197+, 2007.

[21] S. Mehra and S. Susan, “Deep fusion framework for speech command recognition using acoustic and linguistic features,” Multimed. Tools Appl., Mar. 2023, doi: 10.1007/s11042-023-15118-1.

[22] Y. Liu, T. Li, P. Zhang, and Y. Yan, “SFA: Searching faster architectures for end-to-end automatic speech recognition models,” Comput. Speech Lang., vol. 81, p. 101500, Jun. 2023, doi: 10.1016/j.csl.2023.101500.

[23] W. Zhao and Z. Yang, “An Emotion Speech Synthesis Method Based on VITS,” Applied Sciences, vol. 13, no. 4, Art. no. 4, Feb. 2023, doi: 10.3390/app13042225.

[24] Z. Mu, X. Yang, and Y. Dong, “Review of end-to-end speech synthesis technology based on deep learning.” arXiv, Apr. 20, 2021. doi: 10.48550/arXiv.2104.09995.

[25] J. Yu et al., “DIA-TTS: Deep-Inherited Attention-Based Text-to-Speech Synthesizer,” Entropy, vol. 25, no. 1, Art. no. 1, Dec. 2022, doi: 10.3390/e25010041.

[26] M. Ding, “A Systematic Review on the Development of Speech Synthesis,” in 2023 8th International Conference on Computer and Communication Systems (ICCCS), Apr. 2023, pp. 28–33. doi: 10.1109/ICCCS57501.2023.10150729.

[27] M. Kang, D. Min, and S. J. Hwang, “Grad-StyleSpeech: Any-Speaker Adaptive Text-to-Speech Synthesis with Diffusion Models,” in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Jun. 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10095515.

[28] Y. Pang, J. Lin, T. Qin, and Z. Chen, “Image-to-Image Translation: Methods and Applications,” IEEE Trans. Multimedia, vol. 24, pp. 3859–3881, 2022, doi: 10.1109/TMM.2021.3109419.

[29] A. Tewari et al., “High-Fidelity Monocular Face Reconstruction Based on an Unsupervised Model-Based Face Autoencoder,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 357–370, Feb. 2020, doi: 10.1109/TPAMI.2018.2876842.

[30] D. Ghafourzadeh et al., “Local control editing paradigms for part-based 3D face morphable models,” Computer Animation & Virtual, vol. 32, no. 6, p. e2028, Nov. 2021, doi: 10.1002/cav.2028.

[31] E. Lilienblum and A. Al-Hamadi, “A Structured Light Approach for 3-D Surface Reconstruction With a Stereo Line-Scan System,” IEEE Trans. Instrum. Meas., vol. 64, no. 5, pp. 1258–1266, May 2015, doi: 10.1109/TIM.2014.2364105.

[32] Z. Liu, J. Huang, S. Bu, J. Han, X. Tang, and X. Li, “Template Deformation-Based 3-D Reconstruction of Full Human Body Scans From Low-Cost Depth Cameras,” IEEE Trans. Cybern., vol. 47, no. 3, pp. 695–708, Mar. 2017, doi: 10.1109/TCYB.2016.2524406.

[33] H. Xu, J. Hou, L. Yu, and S. Fei, “3D Reconstruction system for collaborative scanning based on multiple RGB-D cameras,” Pattern Recognit. Lett., vol. 128, pp. 505–512, Dec. 2019, doi: 10.1016/j.patrec.2019.10.020.

[34] M. J. Awan, A. Raza, A. Yasin, H. M. F. Shehzad, and I. Butt, “The customized convolutional neural network of face emotion expression classification,” Annals of the Romanian Society for Cell Biology, vol. 25, no. 6, pp. 5296–5304, 2021.

[35] M. Zhao et al., “Deep-block network for AU recognition and expression migration,” Multimedia Tools and Applications, pp. 1–14, 2023.

[36] T. Shen and H. Xu, “Facial Expression Recognition Based on Multi-Channel Attention Residual Network,” Computer Modeling in

Engineering & Sciences, vol. 135, no. 1, pp. 539–560, 2023, doi: 10.32604/cmes.2022.022312.

[37] X. Chen and H. Zhao, "A Novel Fast Reconstruction Method for Single Image Super Resolution Task," Neural Process. Lett., Mar. 2023, doi: 10.1007/s11063-023-11235-y.

[38] S.-J. Kim, C.-H. Kim, and D. Levin, "Surface simplification using a discrete curvature norm," Comput. Graph., vol. 26, no. 5, pp. 657–663, Oct. 2002, doi: 10.1016/S0097-8493(02)00121-8.

[39] C. H. Lee, A. Varshney, and D. W. Jacobs, "Mesh saliency," ACM Transactions on Graphics, vol. 24, no. 3, pp. 659–666, Jul. 2005, doi: 10.1145/1073204.1073244.

[40] M. Hussain, "Efficient Simplification Methods for Generating High Quality LODs of 3D Meshes," J. Comput. Sci. Tech., vol. 24, no. 3, pp. 604–604, May 2009, doi: 10.1007/s11390-009-9249-9.

[41] A. Groshev, A. Maltseva, D. Chesakov, A. Kuznetsov, and D. Dimitrov, "GHOST—A New Face Swap Approach for Image and Video Domains," IEEE Access, vol. 10, pp. 83452–83462, 2022, doi: 10.1109/ACCESS.2022.3196668.

[42] J. Zhang, K. Zhou, and Y. Luximon, "A brief review of 3D face reconstruction methods for face-related product design," Convergence of Ergonomics and Design: Proceedings of ACED SEANES 2020, pp. 357–366, 2021.

[43] S. Zhao, X. Wang, D. Zhang, G. Zhang, Z. Wang, and H. Liu, "FM-3DFR: Facial Manipulation-Based 3-D Face Reconstruction," IEEE Transactions on Cybernetics, 2023.

[44] H. Zhao and B. Wu, "Three-dimensional face reconstruction of static images and computer standardization issues," Soft Computing, vol. 27, no. 2, pp. 1143–1152, 2023.

[45] W. Wei, X. Li, B. Zhang, L. Li, R. Damaševičius, and R. Scherer, "LSTM-SN: complex text classifying with LSTM fusion social network," J. Supercomput., vol. 79, no. 9, pp. 9558–9583, Jun. 2023, doi: 10.1007/s11227-022-05034-w.

[46] C. Dong et al., "A Survey of Natural Language Generation," ACM Computing Surveys, vol. 55, no. 8, p. 173:1-173:38, Dec. 2022, doi: 10.1145/3554727.

[47] Q. Li, W. Zhang, M. Huang, S. Feng, and Y. Wu, "RSP-DST: Revisable State Prediction for Dialogue State Tracking," Electronics, vol. 12, no. 6, Art. no. 6, Jan. 2023, doi: 10.3390/electronics12061494.

[48] S.Chakraborty, K. Yul Ohm, H. Jeon, D. Hyun Kim, and H. Jae Jin, "Intent Classification of Users Conversation using BERT for Conversational Dialogue System," in 2023 25th International Conference on Advanced Communication Technology (ICACT), Pyeongchang, Korea, Republic of: IEEE, Feb. 2023, pp. 65–69. doi: 10.23919/ICACT56868.2023.10079261.

[49] E. Mollick and J. Euchner, "The Transformative Potential of Generative AI: A Conversation with Ethan Mollick," Research-Technology Management, vol. 66, no. 4, pp. 11–16, Jul. 2023, doi: 10.1080/08956308.2023.2213102.

[50] K. R. Prajwal, R. Mukhopadhyay, V. P. Namboodiri, and C. V. Jawahar, "A Lip Sync Expert Is All You Need for Speech to Lip Generation In the Wild," in Proceedings of the 28th ACM International Conference on Multimedia, Seattle WA USA: ACM, Oct. 2020, pp. 484–492. doi: 10.1145/3394171.3413532.

[51] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," Advances in neural information processing systems, vol. 32, 2019.

[52] J. Xie et al., "mm3DFace: Nonintrusive 3D Facial Reconstruction Leveraging mmWave Signals," in Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services, Helsinki Finland: ACM, Jun. 2023, pp. 462–474. doi: 10.1145/3581791.3596839.

[53] P. Liu et al., "MusicFace: Music-driven Expressive Singing Face Synthesis," arXiv preprint arXiv:2303.14044, 2023.

[54] H. Hu, Z. Cao, X. Yang, H. Xiong, and Y. Lou, "Performance Evaluation of Optical Motion Capture Sensors for Assembly Motion Capturing," IEEE Access, vol. 9, pp. 61444–61454, 2021, doi: 10.1109/ACCESS.2021.3074260.

[55] G. Venture, K. Ayusawa, and Y. Nakamura, "Motion capture based identification of the human body inertial parameters," in 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC: IEEE, Aug. 2008, pp. 4575–4578. doi: 10.1109/IEMBS.2008.4650231.

[56] S. Hashi et al., "Wireless motion capture system using magnetically coupled LC resonant marker," J. Magn. Magn. Mater., vol. 290–291, pp. 1330–1333, Apr. 2005, doi: 10.1016/j.jmmm.2004.11.430.

[57] T. B. Moeslund and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," Comput. Vision Image Understanding, vol. 81, no. 3, pp. 231–268, Mar. 2001, doi: 10.1006/cviu.2000.0897.

[58] W. Bian et al., "MoMusic: A Motion-Driven Human-AI Collaborative Music Composition and Performing System," AAAI, vol. 37, no. 13, pp. 16057–16062, Jun. 2023, doi: 10.1609/aaai.v37i13.26907.

[59] R. Wang, B. Zhong, and Y. Chen, "Motion-Driven Tracking via End-to-End Coarse-to-Fine Verifying," IEEE Trans. Circuits Syst. Video Technol., pp. 1–1, 2023, doi: 10.1109/TCSVT.2023.3289620.

[60] N. L. Schroeder and S. D. Craig, "Learning with virtual humans: Introduction to the special issue," Journal of Research on Technology in Education, vol. 53, no. 1, pp. 1–7, Jan. 2021, doi: 10.1080/15391523.2020.1863114.

[61] F. D. A. Martínez Manchón and A. Šimunić, "Effects of Short-Term Exposure to Interactive vs. Non-Interactive Virtual Nature on Cognitive Performance and Mental Health in College Students," International Journal of Human–Computer Interaction, pp. 1–11, Jul. 2023, doi: 10.1080/10447318.2023.2235122.

[62] N. L. Schroeder, E. K. Chiou, R. F. Siegle, and S. D. Craig, "Trusting and Learning From Virtual Humans that Correct Common Misconceptions," Journal of Educational Computing Research, vol. 61, no. 4, pp. 790–816, Jul. 2023, doi: 10.1177/07356331221139859.

[63] T.-C. Huang, M. Limniou, and W.-C. V. Wu, "Editorial: Virtual learning environments in educational institutions," Front. Virtual Real., vol. 4, p. 1138660, Jan. 2023, doi: 10.3389/frvir.2023.1138660.

[64] Z. Huang, Y. Shen, J. Li, M. Fey, and C. Brecher, "A Survey on AI-Driven Digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics," Sensors, vol. 21, no. 19, p. 6340, Sep. 2021, doi: 10.3390/s21196340.

[65] N. W. Roslan, N. Abd Rahim, N. M. Roslan, and S. N. A. Roslan, "Students' presupposition towards incooperating AI (Artifical Intelligence) technology in virtual and face-to-face classes," International Journal of Advanced Research in Future Ready Learning and Education, vol. 27, no. 1, pp. 16–19, 2022.

[66] L. De Angelis et al., "ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health," Front. Public Health, vol. 11, p. 1166120, Apr. 2023, doi: 10.3389/fpubh.2023.1166120.

[67] S. Bharati et al., 12 Applications and challenges of AI-driven IoHT for combating pandemics: a review. De Gruyter Berlin, Boston, 2021.

[68] K. Nah, S. Oh, B. Han, H. Kim, and A. Lee, "A Study on the User Experience to Improve Immersion as a Digital Human in Lifestyle Content," Applied Sciences, vol. 12, no. 23, p. 12467, Dec. 2022, doi: 10.3390/app122312467.

[69] M. Ebnali, N. Ahmadi, E. Nabiyouni, and H. Karimi, "AI-powered Human Digital Twins in Virtual Therapeutic Sessions," Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, vol. 12, no. 1, pp. 1–4, Mar. 2023, doi: 10.1177/2327857923121000.

[70] J. O. Egede et al., "Design and Evaluation of Virtual Human Mediated Tasks for Assessment of Depression and Anxiety," in Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, in IVA '21. New York, NY, USA: Association for Computing Machinery, Sep. 2021, pp. 52–59. doi: 10.1145/3472306.3478361.

[71] T. Olugbade, L. He, P. Maiolino, D. Heylen, and N. Bianchi-Berthouze, "Touch Technology in Affective Human–, Robot–, and Virtual–Human Interactions: A Survey," Proceedings of the IEEE, pp. 1–22, 2023, doi: 10.1109/JPROC.2023.3272780.

[72] "iVisit-Communicate for AEC Education: Using Virtual Humans to Practice Communication Skills in 360-Degree Virtual Field Trips | Journal of Computing in Civil Engineering | Vol 37, No 3." https://ascelibrary.org/doi/abs/10.1061/JCCEE5.CPENG-5165 (accessed Sep. 18, 2023).

[73] M. Seymour et al., " Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments," JAIS, vol. 22, no. 3, pp. 591–617, 2021, doi: 10.17705/1jais.00674.

[74] N. Magnenat Thalmann and D. Thalmann, " Digital actors for interactive television," Proc. IEEE, vol. 83, no. 7, pp. 1022–1031, Jul. 1995, doi: 10.1109/5.390120.

[75] K. Al-Jawaheri, M. M. Adnan, K. A. Kadhim, S. M. Saadi, M. H. Ali, and A. M. Alkhfaji, "Realistic Crying Simulation of 3D virtual human: A survey, " in 2022 5th International Conference on Engineering Technology and its Applications (IICETA), IEEE, 2022, pp. 571–573.

[76] Y. Yang, X. Luo, and X. Hu, " Research on Virtual Human Development Based on Motion Capture, " in 2022 International Conference on Culture-Oriented Science and Technology (CoST), IEEE, 2022, pp. 216–220.

[77] R. Kong, Z. Qi, and S. Zhao, "Difference Between Virtual Idols and Traditional Entertainment from Technical Perspectives:," presented at the 2021 3rd International Conference on Economic Management and Cultural Industry (ICEMCI 2021), Guangzhou, China, 2021. doi: 10.2991/assehr.k.211209.058.

[78] T.-F. T. Yen, L. Han, and L. Yuan, "Modelling the Impacts of Brand Image on WOM for Virtual Idols," Asian Journal of Education and Social Studies, vol. 38, no. 3, pp. 10–24, 2023.

[79] M. S. Rahmi, N. Rahmat, and A. Saleha, "Posthuman in japanese popular culture: Virtual idol Hatsune Miku, " in AICLL: Annual International Conference on Language and Literature, 2018, pp. 81–86.

[80] J.-L. Lugrin et al., "Any 'body' there? avatar visibility effects in a virtual reality game," in 2018 IEEE conference on virtual reality and 3D user interfaces (VR), IEEE, 2018, pp. 17–24.

[81] V. Ferrer, A. Perdomo, H. R. Ali, C. Fies, and J. Quarles, "Virtual humans for temperature visualization in a tangible augmented reality educational game," in 2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR), IEEE, 2017, pp. 1–6.

[82] J. Arroyo-Palacios and M. Slater, "Dancing with physio: a mobile game with physiologically aware virtual humans, " IEEE Transactions on Affective Computing, vol. 7, no. 4, pp. 326–336, 2015.

[83] J. F. Daniel and others, " First Order Motion Model for Image Animation and Deep Fake Detection: Using Deep Learning, " in 2022 International Conference on Computer Communication and Informatics (ICCCI), IEEE, 2022, pp. 1–7.

[84] Q. Shao et al., "Toward intelligent financial advisors for identifying potential clients: a multitask perspective, " Big Data Mining and Analytics, vol. 5, no. 1, pp. 64–78, 2021.

[85] A. V. Bataev, N. Dedyukhina, and M. N. Nasrutdinov, "Innovations in the financial sphere: performance evaluation of introducing service robots with artificial intelligence," in 2020 9th International Conference on Industrial Technology and Management (ICITM), IEEE, 2020, pp. 256–260.

[86] W. Swartout, "Lessons learned from virtual humans," AI Magazine, vol. 31, no. 1, pp. 9–20, 2010.

[87] S. Sylaiou, V. Kasapakis, D. Gavalas, and E. Dzardanova, "Avatars as storytellers: affective narratives in virtual museums," Pers Ubiquit Comput, vol. 24, no. 6, pp. 829–841, Dec. 2020.J. Decker, A. Doherty, J. Geigel, and G. Jacobs, " Blending disciplines for a blended reality: Virtual guides for a living history museum, " J. Interact. Technol. Pedagogy, vol. 17, 2020.

[88] M. Huseinovic, R. Turcinhodzic, and S. Rizvic, " Interactive animated storytelling in presenting intangible cultural heritage, " Proceedings of the CESCG, 2013.S. Sylaiou, V. Kasapakis, D. Gavalas, and E. Dzardanova, " Avatars as storytellers: Affective narratives in virtual museums," Personal and Ubiquitous Computing, vol. 24, pp. 829–841, 2020.

[89] E. Dzardanova, V. Kasapakis, D. Gavalas, and S. Sylaiou, "Exploring aspects of obedience in VR-mediated communication," in 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2019, pp. 1–3.

[90] A. V. Bataev, N. Dedyukhina, and M. N. Nasrutdinov, "Innovations in the financial sphere: performance evaluation of introducing service robots with artificial intelligence," in 2020 9th International Conference on Industrial Technology and Management (ICITM), IEEE, 2020, pp. 256–260.

[91] P. Kenny, T. Parsons, J. Gratch, and A. Rizzo, "Virtual humans for assisted health care, " in Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments, 2008, pp. 1–4.

[92] L. Donnelly, D. Patten, P. White, and G. Finn, " Virtual human dissector as a learning tool for studying cross-sectional anatomy," Medical teacher, vol. 31, no. 6, pp. 553–555, 2009.

[93] A. T. Hirsh, S. Z. George, and M. E. Robinson, "Pain assessment and treatment disparities: a virtual human technology investigation," Pain, vol. 143, no. 1–2, pp. 106–113, 2009.

[94] S. Carnell, B. Lok, M. T. James, and J. K. Su, "Predicting student success in communication skills learning scenarios with virtual humans, " in Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 2019, pp. 436–440.

[95] B. Lok and A. E. Foster, "Can virtual humans teach empathy?, " Teaching Empathy in Healthcare: Building a New Core Competency, pp. 143–163, 2019.

[96] S. Z. Kamoonpuri and A. Sengar, " Hi, May AI help you? An analysis of the barriers impeding the implementation and use of artificial intelligence-enabled virtual assistants in retail," Journal of Retailing and Consumer Services, vol. 72, p. 103258, 2023.

[97] E. S. Silva and F. Bonetti, " Digital humans in fashion: Will consumers interact?," Journal of Retailing and Consumer Services, vol. 60, p. 102430, 2021.

[98] V. Cimolin, I. T. Paraskevopoulos, M. Sala, M. Tarabini, and M. Galli, " The smart body concept as a demonstration of the overarching utility and benefits of 3D avatars in retail, health and wellbeing: an accuracy study of body measures from 3D reconstruction," Multimedia Tools and Applications, vol. 82, no. 7, pp. 11079–11098, 2023.

[99] J. Chakareski, M. Khan, and M. Yuksel, "Toward Enabling Next-Generation Societal Virtual Reality Applications for Virtual Human Teleportation: A novel future system concept and computation-communication-signal representation trade-offs, " IEEE Signal Process. Mag., vol. 39, no. 5, pp. 22 – 41, Sep. 2022, doi: 10.1109/MSP.2022.3151612.

[100] R. Stephens, A. Awasthi, K. Crowley, F. Boyle, and J. Walsh, "A Literature Review of Virtual Reality Interpersonal Training for Salespeople," in 2021 32nd Irish Signals and Systems Conference (ISSC), Athlone, Ireland: IEEE, Jun. 2021, pp. 1 – 6. doi: 10.1109/ISSC52156.2021.9467845.

[101] J. Chakareski, M. Khan, and M. Yuksel, "Toward Enabling Next-Generation Societal Virtual Reality Applications for Virtual Human Teleportation: A novel future system concept and computation-communication-signal representation trade-offs, " IEEE Signal Process. Mag., vol. 39, no. 5, pp. 22 – 41, Sep. 2022, doi: 10.1109/MSP.2022.3151612.

[102] M. Kurosu, Ed., Human-Computer Interaction. Technological Innovation: Thematic Area, HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings, Part II, vol. 13303. in Lecture Notes in Computer Science, vol. 13303. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-031-05409-9.

[103] I. Vladimirov, M. Nenova, D. Nikolova, and Z. Terneva, "Security and Privacy Protection Obstacles with 3D Reconstructed Models of People in Applications and the Metaverse: A Survey," in 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Ohrid, North Macedonia: IEEE, Jun. 2022, pp. 1 – 4. doi: 10.1109/ICEST55168.2022.9828791.

[104] M. Volonte et al., "Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations," IEEE Trans. Vis. Comput. Graphics, vol. 22, no. 4, pp. 1326–1335, Apr. 2016, doi: 10.1109/TVCG.2016.2518158.

[105] S. Cong and Y. Zhou, "A review of convolutional neural network architectures and their optimizations," Artif. Intell. Rev., vol. 56, no. 3, pp. 1905–1969, Mar. 2023, doi: 10.1007/s10462-022-10213-5.

[106] T. Olugbade, L. He, P. Maiolino, D. Heylen, and N. Bianchi-Berthouze, "Touch Technology in Affective Human–, Robot–, and

Virtual–Human Interactions: A Survey," Proc. IEEE, pp. 1–22, 2023, doi: 10.1109/JPROC.2023.3272780.

[107] H. Kim, J. Ben-Othman, L. Mokdad, and P. Bellavista, "A Virtual Emotion Detection Architecture With Two-Way Enabled Delay Bound toward Evolutional Emotion-Based IoT Services," IEEE Trans. Mobile Comput., vol. 21, no. 4, pp. 1172–1181, Apr. 2022, doi: 10.1109/TMC.2020.3024059.

[108] S. Long and J. Yang, "Application of Brain-Computer Interface for Disabled People on Human Factors Engineering," presented at the 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022), 2022. doi: 10.54941/ahfe1001903.

[109] K. Wang et al., "Multimodal Activity Detection for Natural Interaction with Virtual Human," in 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Shanghai, China: IEEE, Mar. 2023, pp. 671–672. doi: 10.1109/VRW58643.2023.00178.

Lipeng Cui (Member, IEEE) Born in July 1990 in Shenyang, Liaoning Province, he received his master's degree from Beijing University of Posts and Telecommunications in 2019, and currently works as a researcher at the Strategic Development Center, Research Institute of China Telecom Corporation Ltd., Beijing, China. His current main research areas include metaverse application architecture, digital human industry applications, and knowledge management area in new technologies.

Jiarui Liu Born in Tianjin in December 1994, obtained a master's degree from Beijing University of Posts and Telecommunications in 2020, and currently works as an industry analyst at the Strategic Development Center, Research Institute of China Telecom Corporation Ltd., Beijing, China. At present, the main research areas include the application of metaverse industry and related technology research, digital society governance, etc.