Realistic problems and innovative path of AI dubbing in the era of intelligent media

Shen Yingmin

School of Public Administration, Hohai University, Nanjing, China

Abstract: *The rapid development of artificial intelligence technology has led human beings into the era of intelligent media, and is inextricably linked with human social production and life. In the field of media, the implementation of AI technology brings technical dividends to the development of media industry, but it is accompanied by real problems brought by immature technology and increased audience demands. This paper addresses the problems of voice homogeneity, mechanical flow of speech, and lack of emotion in AI voiceover, and makes suggestions for its innovative development by combining theories related to the art of broadcasting and hosting.*

Keywords: *Intelligent media era; Artificial intelligence; AI voiceover; Broadcasting host*

1. Introduction

Artificial Intelligence (AI), a concept born at Dartmouth College, is an intelligent technology that allows machines to perform a series of complex calculations and decision analysis instead of the human brain after deep learning, which can effectively release human brain and physical labor. With the rapid development of AI technology, its characteristics of intelligence and convenience continue to be favored by society, making its integration with human production activities continue to deepen. In the media field, the arrival of AI technology has reconstructed the information production and communication process, leading people to enter a new era of information communication with "human-computer symbiosis" and "interconnection of all things", i.e. the era of intelligent media.

2. The birth and application of AI dubbing in the era of intelligent media

2.1. Artificial intelligence technology in the field of media landing

From 1956 to the present, artificial intelligence has accompanied the great progress of computer technology and achieved certain research results in mathematics, medicine, engineering and other fields, and also realized the transformation of its application functions, and its influence on human beings has become more and more profound. Nowadays, artificial intelligence technology has penetrated into all walks of life, especially represented by the new development of the media industry.

In the field of news broadcasting, AI anchors have been seen time and again. At the World Internet Conference held on November 7, 2018, Xinhua News Agency, together with Sogou, released the world's first fully simulated intelligent AI anchor, "Xin Xiaohao."[1] On February 19, 2019, Xinhua News Agency again cooperated with Sogou to release the first AI female anchor, "Xin Xiaomeng ". Nowadays, AI anchor's high efficiency and zero error ability are recognized by more and more media organizations, the first "3D" AI synthetic anchor "Xin Xiaowei",the first sign language AI anchor "Xiao Cong", etc., like "dumplings" into the media field. In the field of text generation, in 2016, artificial intelligence experts and cartoonist Andy Hera, through the application of Google's open source machine learning toolkit TensorFlow developed a set of automatic screenwriting AI, the famous American sitcom "Old Friends" script input, automatically generated a new script, although most of the language is more confusing, but also able to isolate some of the more meaningful scenes and plot. Although most of the language was confusing, it was able to isolate some meaningful scenes and plots. Tencent also launched China's first intelligent writing robot "Dream writer", which can complete a news article in a few seconds by quickly analyzing a large amount of data and information, while Xinhua News Agency's "Quick Pen Xiao Xin" and Toutiao's "Zhang Xiaoming" can do the same.[2] Meanwhile, in the field of short video creation, researchers at Duke University in the United States established a new algorithm based on the intelligent creation of scripts in 2018, which can automatically generate a corresponding short video based on a small paragraph of text. The landing of artificial intelligence technology in the media field, although to a certain extent the living space of those media workers practicing in the traditional mode is squeezed, it is undeniable that artificial intelligence provides more possibilities for the development of the media industry.

2.2. Multi-scene application of AI dubbing in the era of intelligent media

The era of intelligent media is the era of technological dividend brought by the continuous integration of artificial intelligence technology and media industry, which accelerates the production and circulation of information and greatly meets the audience's fantasy and demand for the development of new media industry.AI dubbing, or Text-To-Speech technology (TTS), comes into being under such needs.

In terms of thematic dubbing, in 2018, CCTV broadcast the world's first documentary "Innovation China", which uses artificial intelligence to simulate voices to achieve dubbing, synthesizing the voice of Li Yi, a famous Chinese dubbing artist, with artificial intelligence, achieving a milestone innovation in AI dubbing technology.[3] In terms of news broadcasting, in addition to "New Xiaohao" and "New Xiaomeng", the AI virtual anchor "Xiao Lu" of Xiamen Daily more and more AI news broadcast voice-overs are appearing in the public eye. In terms of life applications, we are familiar with the various types of car navigation APP also appeared in the intelligent navigation voice, from the beginning of Lin Zhiling, Yue Yunpeng voice package, to the later use of speech synthesis technology KAN-TTS, through the recording of their own voice to complete the custom voice package of this personalized service, to further enrich the travel experience and life fun. Nowadays, there are also a large number of speech synthesis websites that only need to enter the text that you want to convert into audio in the text field, and then select the form of voice you want to output as well as the speed of speech and other personalized customization, which can convert the text into audio in just a few seconds, which is also the main way of audio production of audiobooks. In addition, Apple's intelligent voice interaction system "Siri" and Huawei's "Xiao Yi" are typical cases of using voice synthesis technology in daily life. Although the current AI dubbing technology has long been integrated into all aspects of our lives, and its high efficiency, high precision and high adaptation are widely recognized by audiences, there are still many shortcomings compared with the traditional dubbing industry.

3. The reality of AI dubbing in the era of intelligent media

3.1. Homogenization of sound forms

Nowadays, most of the voices of AI dubbing are inevitably homogenized. It is not difficult to find that the voice forms provided by AI dubbing circulating in the market are mainly "teenage voice", "middle-aged voice", "loli voice" and other highly labeled and mapped social groups. The voice forms provided by AI dubbing are mainly "teenage voice", "middle-aged voice", "loli voice", etc., which are highly labeled and mapping a certain social group. At present, the application scenario of AI dubbing is mainly focused on voice broadcasting, short video narration and other living applications, but with the increase of application scenario and the number of applications, the original several types of voice forms can no longer meet the audience, although many voice synthesis platforms have made certain innovations, but after careful differentiation, we can find that it is still essentially "a change of soup but not a change of medicine Although many voice synthesis platforms have made certain innovations, it can be found that they are still essentially "the same soup but not the same medicine" and cannot fundamentally solve the problem of homogenization of AI voice acting.

Voice form is the external expression of language expression, the use of breath, vocal cord conditions, oral opening, etc. will have an impact on the sound quality, and this is the key to the uniqueness of the voice. Different voices bring different auditory experiences to the audience, and this is the characteristic brought by the differentiation of voice. Similar or solidified sound constantly pulls down the audience's sense of freshness, and the listening experience is significantly reduced. With the abundance and development of accompanying communication media, the importance of sound form is self-evident. However, AI voice-overs that are mass-produced, algorithmically fixed and widely used in the market cannot take as long as three months to complete all the voice-over work as the production team of the documentary "Innovation China" did, during which a lot of costs were invested to achieve the desired purpose.

3.2. Mechanization of phonological changes in speech flow

"Mechanization" is a common problem of artificial intelligence in the field of humanoid simulation, which presents the problem of poor speech flow and insufficient linguistic tension in AI voiceover. The flow of speech is the process of language expression created by the combination of ideographic material of words, phrases and sentences. The sound change is a fixed and unique pronunciation rule of Mandarin Chinese, which means that the vowels, rhymes or tones in some syllables change due to the influence of adjacent syllables in the flow of speech, in order to reduce the blockage and jams caused by the adjacent sounds of individual words in daily communication. Due to the vast size of China, there are differences in the phonetic habits of different regions, resulting in a richer variation in the phonetic variation of speech flow or pronunciation patterns in different regions, thus giving birth to dialects. In Mandarin Chinese dubbing, although AI can find fixed and objective sound change patterns through algorithms, it does not have anthropomorphic oral and breath movement processes, and cannot fully simulate the natural and smooth language processing of real people, so the audience will obviously feel different and uncomfortable in listening, not to mention AI dialect dubbing.

The external skills of language not only include the shallow flow of speech and sound changes, but also the external skills in the professional internal and external skills of broadcasting and hosting, which include stopping, stress, tone and rhythm, which are deeper laws of language expression. The AI dubbing commonly used in the market can only make the so-called sentence breaks through punctuation when facing a text, which is not consistent with the normal language expression rules, i.e., the lack of stopping, and the lack of stopping makes the AI dubbing in the change of language rhythm is quite monotonous. Likewise, because AI does not have human social consciousness and social thinking mode, it is unable to select important words for proper emphasis when dealing with the script, and lacks the application of accent and tone.

3.3. Empty generalization of emotional expression

The essence of the mechanization of the external expression of language is the absence and emptiness of emotion. The famous CCTV dubbing teacher Sun Yebin once mentioned in his book "The Voice" that the meaning of language is to convey emotions and meanings, and dubbing is never the sound of words.[4] The dubbing work is the process of secondary creation of the manuscript, and the works created by the dubbing artist often contain his own understanding of the manuscript, his upbringing, life experience, and personal dealings will affect his secondary creation of the manuscript, full of personal emotions, which is the unique thinking process and thinking mode of human beings, and the fusion of emotion and expression, internal and externalization into one, making the dubbing meaningful. Since AI dubbing does not have human social thinking and cannot express emotions, it cannot be called secondary creation for dubbing scripts, but is closer to the "sounding" of words, just transforming words into sounds.

At the same time, the lack of emotional expression in AI voice-over can also be expressed as a lack of abstract thinking. The internal skills in the internal and external skills of announcing and hosting include situational presence, object sense, and internal language, which are the externalization of these abstract concepts through voice after processing. AI voice-overs do not understand abstract thinking as humans do, so they cannot effectively understand the content of the text and convey it through expression skills. This makes the effect conveyed by AI dubbing raw and flat in the sense of listening, and cannot resonate with the audience emotionally, so it cannot achieve the effect of sound into the heart.

4. The innovation path of AI dubbing in the era of intelligent media

4.1. Rich sound material

Widening the sound collection channels and enriching the sound material library can, to a certain extent, meet the audience's demand for personalized AI dubbing. In order to improve the level of anthropomorphism, AI voice-overs should first have a rich sound database, not only to discover the differences between people's sound elements, but also to incorporate the functional vocalizations that people have in their daily lives. Through a large reserve of sound materials, the algorithm can discover the rules and transform them to provide more humanized, lifelike and personalized AI voice-overs when facing the audience.

4.2. Strengthen the language tension

The strengthening of linguistic tension is reflected in the flexible use of external skills of voice. Rich language tension is the prerequisite for AI voice acting to be able to perform all kinds of voice acting work. At present, the "power point" of AI voice acting still remains in the more basic work areas such as voice navigation and movie narration on short video platforms, and it is still not capable of performing voice acting work that requires distinctive language color such as documentary narration and advertising voice acting. Therefore, AI dubbing should strengthen the level of anthropomorphism, imitate the human spitting and vocal dynamic process, in addition to the mastery of the existing speech flow and sound change rules, but also to understand the application of external skills enough to break the mechanical sense, which requires AI dubbing in the abstract thinking ability to make greater progress.

4.3. Cultivate language emotion

Meaning is the voice of the heart, and the purpose of language art is to convey the voice of the heart, i.e., to convey emotion. The cultivation of emotional and social thinking patterns is the biggest obstacle to the development of artificial intelligence. As long as enough language samples are collected, AI will be able to figure out human emotions from what people say, as well as from expressions and actions and other paralanguage. Kai-Fu Lee describes such functions as weak AI, where the AI can understand emotions, but cannot express them.[5] The dubbing work is the secondary creation of the dubbed manuscript, which requires the dubbers to understand the connotation of the text while using the combination of broadcast hosting expression skills and emotion and voice to fully convey the content of the manuscript to the audience, so that the audience can understand the text through the voice. AI dubbing needs not only to improve the level of anthropomorphism, but also to express the temperature of the work and deliver a work with heart.

4.4. Accelerate technological innovation

Science and technology is the first productive force. Looking back at the development of AI, behind every leap is the innovation of science and technology. To improve the working ability of AI voiceover, we must first break through the technical bottleneck of AI's emotional expression and skill application. By optimizing AI's deep learning model, we can improve AI dubbing's perception of words and solve its various problems such as homogenization, mechanization and emotional vacuity. Science and technology, as an important foundation for AI voiceover, is an important prerequisite for the aforementioned anthropomorphic requirements, so it is necessary to increase R&D efforts to strengthen AI's emotional expression and promote a deeper and more comprehensive integration of AI voiceover with human life. As one AI synthetic anchor said, "As an AI news anchor under development, I know I need to improve a lot more."[6]

5. Conclusions

The innovation of artificial intelligence technology is not coming straight at us, but is happening all the time, imperceptibly, and integrating into our life bit by bit. However, we only pay attention to the nodes of leap one by one. The advent of AI voice acting and its impact on our lives is a great example. People's demand has led to corresponding intelligent services. At present, although AI dubbing has been integrated into our lives and become an indispensable part of the era of intelligent media, some people have raised concerns about it. However, we believe that in the future, the traditional voice actors will win in the parts that are not easy to quantify, such as abstract thinking, such as emotional value, such as personalized expression created based on profound skills. Everything is very reasonable. Efficient and convenient information production mode is the inherent advantage of AI dubbing, but for more detailed information processing ability, the development of AI dubbing still has a long way to go.

References

[1] Du Xiaokang. Exploration and practice of Xinhua News Agency in the era of intelligent media [J]. Journalism Research Guide, 2020, 11(11):189-190.

[2] Lei X. Opportunities and dilemmas: new media industry and ethical thinking under the perspective of artificial intelligence [J]. Southeast Communication, 2020(06):32-34.

[3] Guo Yuhui, Zhang Yuying. Functional analysis of AI dubbing in documentary Creation -- A Case

study of Innovation in China [J]. World of Sound Screen, 2021(11):73-74.

[4] Voices by Sun Yebin: voiceover theory and practical skills / by Sun Yebin. Beijing: Communication University of China Press, 2016.6.

[5] Xie Xiaomin, Lin Xiaojue. New thinking brought by artificial intelligence for broadcasting and hosting art majors [J]. Contemporary Television, 2018(11):32-33.

[6] Zhang Biao, Xu Chunjuan. Limitations and prospects of AI synthetic anchors in the era of intelligent media [J]. Young Journalists, 2020(14):23-24.