

Technological Polyphony: Innovation Logic and Path of AIGC in Dialect Publishing

YanhuaQin¹, Yining Ma^{1*}

¹*School of Journalism and Communication, Beijing Normal University, Beijing, China*

**Corresponding Author. Email: 503896025@qq.com*

Abstract. Generative artificial intelligence is profoundly reshaping the boundaries and logic of the publishing industry. This research focuses on the application of AIGC in the specific field of dialect publishing, and dialectically explores its “polyphonic” role. On the one hand, AIGC provides a technical engine for the large-scale archiving and innovative product development of endangered dialect audio resources through efficient and low-cost speech synthesis and reproduction capabilities, and acts as an “enabler” to activate cultural heritage and expand publishing boundaries. On the other hand, the inherent standardization tendency, data bias and business logic of technology may also make it a “alienator” of the uniqueness of dialect culture, leading to the “decontextualization” of publishing results and the loss of cultural authenticity. By constructing a “protection-digestion” critical analysis framework, this article aims to provide theoretical reference and practical paths for how the publishing industry can responsibly apply AIGC technology and realize the unity of cultural benefits and commercial value.

Keywords: generative artificial intelligence, Audiobook, Dialect protection, Publishing Ethics

1. Introduction

The traditional publishing industry is transforming to knowledge service. In recent years, the rapid development and wide application of artificial intelligence technology have had a profound impact on many emerging fields. Among them, the digital publishing industry is a typical representative, which is undergoing subversive changes. The impact of artificial intelligence on the publishing industry is mainly reflected in process reengineering, content reshaping, format innovation and intelligent transformation of traditional publishing formats [1]. At the same time, the audiobook market is growing rapidly, seeking breakthroughs in content and form.

AIGC, especially speech synthesis and cloning technology, provides a new tool for the cost reduction and form innovation of content production. However, due to the high production cost of dialect publications and the relatively small audience, the development is limited, and its cultural value and specific market demand are real.

The application of AIGC in dialect publishing presents a distinctive “polyphonic” feature. Is the publishing industry using AIGC to develop dialect audio content on a large scale, opening a dusty “cultural treasure house”, or creating a seemingly real “cultural imitation” that has lost its soul? This

is not only a “weapon” to break through the bottleneck of traditional publishing and activate cultural heritage, but also a “risk” that dispels cultural authenticity and triggers publishing ethical issues, providing publishing organizations with forward-looking strategic thinking and risk warning.

2. The dimension of empowerment: AIGC as the “activation engine” of dialect publishing

2.1. From “unpublishable” to “publishable”: activating stock copyright content

Local publishing houses have made the collections of local documents, local chronicles and folk stories that have never been converted into audio books at low cost and high efficiency through AIGC, which has revitalized the stock of copyright assets and created a new product line. Similarly, the initial materials such as the manuscripts of the programs Sichuan Dialect and Panda Rolling to are also from Sichuan Dialect Records and Sichuan Dialect Records compiled and published by the Sichuan Local Records Office [2].

Qilu Culture Publishing House has a large number of precious, but out-of-print or low-sales paper local documents, including a collection of Jiaodong Folk Tales collected in the 1980s. This book is rich in content, including dozens of stories such as The Legend of Penglai Fairy Island, Laoshan Taoist Biography, and The Mirage of Mirage. It has high folklore value and literary value.

However, due to the pure text book and the strong localization of the content, it has not been able to enter the mass market, and the copyright assets are in a dormant state. The innovation of the publishing house has filled this gap. In 2022, Jinan Publishing House and Shandong Miaoyin Media cooperated to launch the radio drama Taishan Shigandang, which adopted a grand structure of 467 story units and more than 800 character dubbing to provide a model for the audio adaptation of Jiaodong folk stories. Qingdao Publishing House embedded two-dimensional code in the album of Qingdao Ocean Folk Tales, scanning the code can listen to the dialect version of the audio content of the stories such as Eight Immortals Crossing the Sea and Wangfu Reef, strengthening the interaction of intangible cultural heritage communication.

Facing the dilemma of concrete and scene-oriented traditional production, Kaifeng's first dialect audio book Laojie Guild Hall was carefully produced based on the famous writer Wang Shaohua's novel Laojie Guild Hall. It was recorded by Kaifeng Radio and Television Station announcer, host, and many performing artists. Once it went online in 2022, it was loved by many Kaifeng listeners. At the beginning of the New Year in 2024, reviewing the classics and re-publishing this audio book, let us feel the past and present of a city, the ups and downs of a guild hall, the past and present of an old street and the joys and sorrows of a human world in the strong taste of old Kaifeng.

If you make audiobooks in the traditional way, the cost is high, the efficiency is low, and the market risk is high. You need to hire a professional announcer and pay by the hour. If you want to reflect the local characteristics, you need to find broadcasters who can speak standard dialects. Talents are scarce and the cost is higher. After a set of recording, editing, and proofreading processes, it may take several months to produce a 300-page book. Through digitizing and organizing the text, the editor first scans the paper book with high definition and recognizes it with OCR (Optical Character Recognition) and converts it into electronic text, which is proofread by the editor to ensure the accuracy of the text. Secondly, the AIGC technology company, which provides high-quality and multi-dialect speech synthesis services, is selected to “clone” the dialect voice and find a “voice model” with good voice conditions and pure dialect.

Record hours of specified audio material in a professional studio. This material covers the unique tones, vocabulary and intonation of the dialect. The technology company used this recording to train the “dialect speech synthesis model”. This model can understand the text and read it aloud with an

extremely anthropomorphic tone and dialect characteristic voice. The generation and post-processing are generally embodied in inputting the proofread electronic text into the system, selecting the trained “dialect” model, and generating a preliminary audio book audio file.

Although AIGC has generated the main content, manual intervention is still needed in the later stage. The audio editor will adjust the speaking speed, add appropriate background music such as ocean waves and guzheng sound effects to the key plots, and make a unified opening and closing credits for each story. The finished dialect audio book will be launched on the publisher's own App, Himalaya, WeChat Reading and other audio platforms, and will be paid separately or included in the membership database. It highlights its unique selling point during marketing. At the same time, the original electronic version is attached to the book information as a comparison to satisfy some readers' curiosity about the dialect text.

Local publishing houses can start from their own unique, underdeveloped "copyright rich mine" and use AIGC as an "advanced tool" to create new cultural products with economic benefits and social values at extremely low costs and risks. This is not only an innovation in the publishing industry, but also a new paradigm of cultural inheritance in the digital age.

In the spring of 2025, the book *Traveling with Food of the Lingnan Double Line* series, carefully created by Guangdong Haiyan Electronic Audiovisual Publishing House, was published. This book brings together 60 articles by local authors from the three Lingnan dialect regions of Guangfu, Chaoshan and Hakka. From Guangfu cuisine, which is rich in ingredients and seasonal fragrance, to Chaoshan cuisine, which emphasizes original flavor and fine cooking, to Hakka cuisine, which is simple, heavy and full of mountains and wilderness, it is rich and diverse. At the same time, the audio explanation in the authentic dialect is recorded, so that the cultural essence of Lingnan can be presented in a lively, sensible and audible form.



Figure 1. Word cloud of dialect publication reports

2.2. From “niche publishing” to “precision service”: mining market segmentation needs

AIGC provides personalized and customized dialect audio content for specific groups such as overseas Chinese and local culture enthusiasts, such as home letters and local news read aloud in hometown dialect. AIGC makes “publishing on demand” possible in the field of dialect. Topic selection is the source of book publishing. Topic selection planning reflects the productivity of editors and determines the vitality of books. “Everything is difficult at the beginning” is a true portrayal of topic selection planning. The series of Visualizing Dialects classifies, explains and analyzes Lingnan dialect nursery rhymes, and carries out visual design, which has high academic value and cultural inheritance value [3].

2.3. From "single form" to "integrated product": innovative publishing form

The dialect learning platform of Ningbo Publishing House: Ningbo Publishing House has developed its best-selling dialect books, such as Ningbo Dialect Classroom, into a dialect learning platform, which is convenient for learners to use through audio reading. In terms of sales strategy, they try to make paper books and digital formats promote each other, such as buying paper books to give away digital rights, or buying digital editions to give away audio books. This shows how local publishers use the stock of copyright assets (dialect books) for audio development and integrated marketing.

Times New Media Publishing House's Huiyin Wanyun-A Collection of Anhui Nursery Rhymes contains more than 500 Anhui original nursery rhymes, and innovatively recorded two versions of audio for nursery rhymes in dialect and Mandarin. The product adopts the form of "U disk + cultural and creative notebook" and is accompanied by animation. Users can watch it by scanning the QR code, and the folk literature is digitally and multimedia-preserved and disseminated.

The Nanjing Gulou District Archives has established an audio database of Nanjing dialects in conjunction with authentic dialect speakers, which contains a large number of audio and video materials of daily words and phrases. The database is open to the public through a small program, aiming to protect and inherit dialects. This case focuses on how to systematically establish an audio file for dialects, laying the foundation for the subsequent development of audio books, developing a fusion publication of paper books + dialect AI audio guide, or embedding AI voice accompanying reading function in children's dialect picture books, which greatly enriches the experience and value of publications.

3. Alienation risk: AIGC as a potential threat to cultural authenticity

3.1. The hegemony of "standard tone" and the lack of publishing responsibility

An AI model trained on biased data will produce a "standardized" dialect, which is actually an "authoritative version" of a dialect constructed through publishing activities, virtually marginalizing other accents, which is contrary to the responsibility of publishing to promote cultural diversity.

When technology gives new life to the old sound, it may also put intangible shackles on it. When we use algorithms to "save" a dialect, we may unconsciously simplify and solidify it. In the end, we save an exquisite "specimen", but may lose its vigorous vitality.

AIGC models are not created out of thin air, they rely heavily on training data. When the data itself has deviations or deficiencies, the model will construct a "standard" or even a "fictional" dialect. The training data of AI speech technology is often concentrated in mainstream languages such as Mandarin and English, resulting in a significantly lower recognition rate of dialects and minor languages. This imbalance of data has objectively aggravated the "digital divide" in the AI voice field. With their advantages, leading companies may monopolize high-quality data resources, while small and medium-sized enterprises rely on limited open source data sets, resulting in a continuous widening of the technical performance gap.

In the face of dialects that account for a very low proportion of training data, the performance of AI may not only be "poor", but completely invalid. For example, the song Shi Ye in Chaoshan dialect, due to its unique expression and arrangement, makes the AI model misdecode "You don't scatter" (don't talk nonsense), or recognize "Pai Puzi" (playing cards) as "Pai Putin". Behind this is the cognitive dilemma of the algorithm caused by the low data of some dialects in the training library. The defeat of AI in the face of complex dialect culture is sometimes not a pure technical defect, but more like a kind of "civilization generation difference". Algorithms are difficult to

analyze those cultural antibodies that grow in the sound wave folds and carry deep local culture and life wisdom.

3.2. Lack of publishing responsibility:when the cultural guardian becomes “conspiracy”

The characteristics of body fission propagation allow dialect propagation to achieve second-level propagation, and such fission speed has become a double-edged sword for dialect propagation. At the same time, dialect propagation based on micro-platform has the characteristics of low threshold, grassroots, and large traffic. The cost of information replication and dissemination is extremely low. Within one second, information may spread all over the world, and it is even more difficult to correct errors. Even though information technology continues to innovate at this stage, a series of methods of Internet supervision have emerged, such as keyword filtering, sensitive information shielding and other supervision methods, but there is still no way to control video content, especially live content. For dialect live broadcast, technical means can not grasp the psychological dynamics of the communicator, can not control the verbal expression of the communicator, and can not control the content choice of the communicator. Only after the information is transmitted can it be quickly captured. At this stage, the capture of video content needs to be improved, which gives a few communicators the opportunity to exploit the loopholes. In order to gain profits from traffic, they constantly touch the bottom line of cultural communication, at the expense of public trust in cultural communication, and have a negative impact on dialect communication. Therefore, in the face of the difficulty of supervising the spread of dialects on the Internet, it is necessary to formulate an effective prevention mechanism.

Publishing organizations should have assumed the responsibility of promoting cultural diversity, but when introducing AIGC technology, if they lack sufficient vigilance, they may change from a guardian to a conspiracy of “standard sound” hegemony. In the era of AIGC, the focus of publishing science needs to cover copyright issues such as ownership, infringement and protection of rights and interests, as well as ethical issues such as technical ethics and academic ethics anomie, and cultural security issues such as ideological penetration and deviation of cultural values that may be caused by this. This means that when publishing organizations adopt AIGC, they must not only consider efficiency and cost, but also examine its cultural consequences.

When a publisher uses a specific AIGC dialect model to mass-produce audiobooks, it virtually establishes the dialect accent generated by the model as a “correct” and spreadable version. This marginalizes the accents of other villages and towns within the dialect, and essentially constructs an “authoritative version” of the dialect through publishing activities, which is contrary to the responsibility of publishing to promote cultural diversity.

AI may generate inaccurate or erroneous information when generating content, that is, “AI illusion”; When this phenomenon appears in the cultural field, it will give birth to “cultural illusion”. Cultural illusion has the characteristics of universality, concealment, easy diffusion and long-term nature. AI may simplify the philosophical concept of “the unity of man and nature” into environmental protection slogans, or compress the complexity of historical figures into rigid labels. When distorted content is spread through publishing channels, it will affect the public's understanding of traditional culture, and ultimately threaten the authenticity and integrity of cultural memory.

The “AI Dialect Protection Program” of IFLYTEK Suzhou Research Institute of HKUST launched the short film “Gu Su Suo Ji” dubbed in Suzhou dialect. In this process, the team found that the AI synthesized Suzhou dialect still has the problem of “a little electronic sound”. In order to avoid technology affecting the aesthetics and completion of the video, they chose to use AI dialect

as a narration, creating a subtle contrast of “future people telling ancient things”, which reflects the understanding and strategic avoidance of technical limitations in the application of technical institutions.

The project is promoted by building a fine dialect phonetic database. In order to train the TTS model of Suzhou dialect, the team has accumulated more than 1,000 hours of corpus over three years. This highlights the fundamental role of high-quality data for technical effects. The project is also committed to exploring application scenarios, through the use of the short video mode that Generation Z loves to see and the “Hanfu circle” where traditional culture fans gather, to arouse the empathy of contemporary young people with dialects.

The “sound wave guerrilla warfare” of Shi Ye confirms from the negative side that the vitality of dialects challenges standardized algorithms. The song Shi Ye deliberately used a large number of Chaoshan dialects and unique singing skills, which successfully interfered with the AI speech recognition system and even caused its “defeat”. This is not a technological failure, but a “quantum entanglement strangulation of algorithmic hegemony by cultural diversity”. It proves that the living and mutated oral culture itself is resistance to any force trying to standardize and centralize it.

In the face of these risks, publishing organizations and technology developers must take responsibility to ensure that technology becomes a bridge to cultural diversity, not an executioner.

3.2.1. Promote the democratization of data

When publishing institutions cooperate or develop dialect AIGC models, they should consciously collect voice samples from different regions, different ages, and different social classes, and strive to build a corpus that can better reflect the true diversity of dialects.

3.2.2. Indicate technical limitations and provide humanistic annotations

In the audio dialect content generated by AIGC, publishing organizations should frankly explain possible technical deficiencies like iFlytek. A better approach is to hire dialect scholars or local cultural consultants to review and correct the content generated by AI, and even explain the variation of certain vocabulary or pronunciation in a specific area in the form of annotations.

3.2.3. To build a legal barrier to cultural security

Some experts suggest that developers of generative artificial intelligence should be encouraged to fulfill their information disclosure obligations, such as disclosing the source and proportion of the core content of Chinese civilization in training data, and explaining the model's processing logic for cultural symbols. At the same time, establish a traceability and labeling mechanism for AI cultural information, add unique identification to the cultural information generated by AI through blockchain and other technologies, make its source, editing history and other information traceable, and help the public make more informed judgments.

The essence of technology is a tool, and the soul of culture lies in its freshness and complexity. AIGC provides us with an unprecedented tool for preserving and disseminating dialects, but we must clearly realize that while any tool is in use, it is also shaping the goals and cognition of users. As the gatekeeper of culture, publishing's ultimate responsibility is not to produce the most and cheapest cultural products, but to guard the cultural ecology full of wildness, differences and unlimited possibilities.

3.2.4. Decontextualized narrative and loss of cultural depth

Separating the dialect from its specific life and ritual scenes and simplifying it into pure “audio products” for publication may lead to the flattening of cultural connotations, and readers consume only a “local accent wonder” rather than a complete cultural experience.

The current dialect publishing, especially the core issues that may be brought about by the intervention of AIGC technology. When the dialect is extracted from the soil on which it depends and simplified into pure audio symbols, its rich cultural connotation is in danger of being hollowed out and flattened, and eventually there may be only an empty shell of “local accent” for people to hunt for.

Dialect is far more than just a set of pronunciation and vocabulary. It is a living cultural organism whose vitality is deeply rooted in specific life scenes, social relations and ritual situations.

A Southwest Mandarin “Xiong Qi!”, The thunder on the football field is different from the meaning when cheering for your peers on the construction site, depending on the specific tone, expression and occasion. The ever-changing meaning of “rectification” in the northeastern dialect must also be understood in the specific wine table atmosphere and labor scenes.

When technology only treats dialects as “voice packs” for recording or synthesis, what it does is strip away its context. It may retain the tone and vocabulary, but lose the emotional proportions, interpersonal relationships and scene spirit behind the words. This is like making a specimen of a butterfly. It retains its lifelike form, but loses its vitality flying among the flowers.

The “dissolution” of cultural depth, from emotional resonance to local accent wonders. The direct consequence of this decontextualization is the dissolution of cultural depth, which is manifested in the fading of the emotional dimension: the charm of many dialect nursery rhymes and folk songs lies not only in the words and sentences, but also in the collective memory and emotional interaction in specific situations. Ningbo's audio book *Ningbo Nursery Rhymes in Search of Sound* has preserved 200 Ningbo nursery rhymes through “audio”, but if it is only appreciated as an independent audio clip, it contains the specific background of the times. The atmosphere of family affection and teaching conveyed by nursery rhymes when elders and children are cooling on summer nights and interacting with games, and the “bloodline inheritance and nostalgic memory of Ningbo people in the past dynasties” carried by it can hardly be fully perceived. These nursery rhymes are “vivid manifestations of Ningbo's local culture and traditional scenery”, and are closely connected with local life scenes and folk customs.

Dialect words and sentences without context support can easily be interpreted by the audience with their familiar cultural framework, resulting in misunderstandings. For example, the Sichuan dialect “rake ears”, if only read from the literal, it is easy to simply understand as “fear inside”. However, the book *New Sichuan Dialect* traces its origin and points out that “raking ears” first refers to a bicycle with a “partial bucket” next to it. It is a design made by Sichuan men to make their wives more comfortable when they go out. Its original intention reflects the man's deep love for his wife. Without such cultural annotations, the wisdom of life and delicate emotions contained in dialects may be ignored.

In the end, this decontextualized dialect product may only satisfy many curious listeners of dialects with a shallow consumption of “foreign customs”. What everyone consumes is no longer a complete cultural experience, but a spectacle “local accent”. This is like a souvenir that tourists buy at a scenic spot. It originates from the local area, but it is often out of touch with the real and complicated life of the local area. At the press conference of “Shaanxi Dialect Integration”, although the dialect speaker showed the original pronunciation, most young readers shook their heads and

said “cannot understand”. If there is a lack of guidance and understanding, this unintelligible “native accent” may only remain at the level of novelty.

Sichuan's New Sichuan Dialect strives to go beyond pure text or pure audio. It integrates pictures, text, sound and images. It selects 400 Sichuan dialects, corresponding to 400 audios, 400 image posters and 40 micro-videos. For example, while providing dialect pronunciation for the entry “rake ears”, it also reveals the cultural connotation of “man's deep love for his wife” by tracing its origin (referring to a bicycle equipped with a “partial bucket”) and cultural “tips”. This way of combining pictures and texts and tracing back to the source is trying to supplement the context of dialect vocabulary and restore the life scene and cultural meaning behind it.

The book *Telling Stories in Anqiu Local Dialect* contains 191 short essays, including jokes, stories, documentaries, children's songs, etc. The purpose is to “naturally embed more colloquialisms and local dialects in them, so as to retain the essence of Anqiu dialect”. This method provides a native language environment and usage scenarios for dialect vocabulary and expression by telling stories. At the same time, the author recorded the audio and generated a QR code, so that readers can hear the “authentic Anqiu dialect”. The combination of sound and story text simulates the use of dialect in real life to a certain extent, and helps readers feel the vitality of dialect.

Qian Cheng's audiobook *Qian Cheng Brings Nong Baixiang to New Shanghai* uses 106 Shanghai dialect audio to connect 18 scenic spots along the Shanghai subway. It tries to combine “Shanghai gossip” with specific geographic space, historical changes, and celebrity anecdotes. For example, the book introduces some scenic spots that even the old Shanghai people may not know, such as the Red Garden and Guteng Garden around Metro Line 5, and the “No. 1 archway in the south of the Yangtze River” in the old street of Xinchang--Sanshi Erpinfang. It also introduces in depth from the perspectives of architectural style and historical changes. In this way, the dialect is no longer an isolated sound sample, but as a tool to tell and experience a specific regional culture, injecting more local knowledge into the “local accent”.

4. Combatting decontextualization as cultural gatekeepers

4.1. From "voice reproduction" to "cultural translation"

the goal of the publishing project should not stop at “using AIGC to generate dialect audio”, but should “build a three-dimensional dialect cultural experience”. In the production of dialect audio content, dialect scholars can be hired to supplement the life scenes, usage norms and stories behind the dialect vocabulary or expression in the form of voice notes, background introductions, graphic displays and even short videos. For example, Chinese Language and Culture Collection·Kaifeng strives to create a five-in-one integrated media reading experience of “sound, image, picture, text, and video”, and show the charm of dialect culture from multiple angles.

4.2. Build a contextualized corpus

When collecting and building a dialect corpus for training AIGC, we should consciously include voice samples from different scenes (such as daily conversations, folk stories, rituals, etc.), and record the contextual information of each voice in as much detail as possible. For example, Ulanqab's *Ulanqab Chinese Dialect Vocabulary* not only includes more than 27,000 dialect words, but also includes proverbs, allegorical sayings, and colloquialisms. This kind of systematic collation can provide an important reference for understanding the context of dialect use.

4.3. Technical empowerment, not technical substitution

AIGC can be used as a powerful auxiliary tool, but it cannot completely replace human work. After AIGC generates content, it must be strictly reviewed and corrected by native speakers and cultural consultants to ensure its cultural accuracy and richness. The ideal state is that AIGC is responsible for “similar in form”, and human experts ensure “similar in spirit”.

AIGC provides us with an unprecedented tool for preserving and disseminating endangered dialects, but the convenience of technology should not be at the expense of cultural depth. When we worry about the hegemony of standard pronunciation, we are defending the diversity of dialects; When we are wary of “decontextualization”, we are defending the “depth” of dialects.

4.4. Combatting decontextualization as cultural gatekeepers

4.4.1. Ethics first, clear the responsibility of publishing subject

Publishing organizations should establish a cultural advisory system for the AIGC dialect project, and invite linguists and local cultural holders to participate in the entire process from data collection to product review to ensure cultural authenticity.

While AIGC technology brings innovation to dialect publishing, we must also face up to its potential cultural risks. It is very important to construct a responsible framework, and its primary principle is ethics first, and to clarify the main responsibility of publishing organizations that cannot be shirked.

Dialect is not only a communication tool, but also a living carrier of regional culture, carrying specific lifestyles, values and collective memories. When AIGC technology is applied to dialect publishing, publishing organizations must take the lead in considering ethical issues, which are related to the authenticity and integrity of cultural heritage.

To ensure cultural authenticity, if the training data of the AIGC model is insufficient or biased, it may produce content that does not conform to the cultural background, grammatical habits or emotional expression of a specific dialect, or even generate a “standardized dialect”, which weakens its diversity and vitality. For example, in literary creation, how to use dialects to maintain a balance between regional cultural characteristics and acceptance by a wider audience is a “degree” that needs to be grasped.

To fulfill the responsibility of intangible cultural heritage protection, from a legal point of view, dialects are often positioned as “intangible cultural heritage”. This means that publishing organizations have special protection and inheritance responsibilities when using AIGC technology to process dialects. Any commercial application that may lead to “weakening of characteristics” or “endangered crisis” requires extra caution.

To deal with the limitations of technology, technology itself cannot understand the subtleties in the cultural context. For example, AI may treat a slang full of specific emotions in a dialect as a literal meaning, leading to “decontextualization” and loss of depth of cultural connotation.

Publishing organizations should transform it into specific and operable systems and actions. One of the core measures is to establish a cultural advisory system for the AIGC dialect project.

An effective cultural consultant team should include linguists who provide theoretical support for dialect linguistics and ensure the structural accuracy of the content generated by AIGC, including local cultural holders such as elders who are familiar with local culture, non-genetic inheritors, and folk artists. They are the carriers of the “living culture” of dialects. They can judge whether the

content conforms to the local cultural logic, customs and emotional expressions, and provide local cultural researchers with deep cultural contexts from the perspectives of history and folklore.

To clarify the responsibilities of cultural consultants in the whole process, cultural consultants should be deeply involved in the data collection stage, model training and optimization stage, product review and correction stage, etc., to guide the collection of dialect corpus, to ensure its comprehensiveness(covering different ages, regions, social classes) and the authenticity of the scene(including daily dialogue, folk stories, ritual congratulations and other contexts).The State Key Laboratory of Pattern Recognition of the Institute of Automation of the Chinese Academy of Sciences will extensively recruit dialect speakers of different ages and regions for corpus collection through small programs in dialect research, assist in evaluating and labeling the cultural attributes of the corpus, prompt certain words or expressions of specific use situations, emotional colors and cultural taboos, and conduct cultural authenticity audits on the dialect content generated by AIGC, including the accuracy of the language itself, and whether it appropriately presents the cultural customs and values behind it.Assist in the design of user guides or cultural annotations to explain the regional background and cultural characteristics of dialects used in the project.Continuously collect user feedback for iterative optimization of AIGC model.

4.4.2. Establish the main responsibility of the publishing organization

The publishing organization is the final producer of the AIGC dialect product and must assume the primary responsibility, including system design and resource guarantee and transparency commitment, proactively establish and fund the cultural consultant team, incorporate it into the project budget and process, and clearly inform users of the scope of technology use and the participation of cultural consultants in AIGC dialect products, such as listing the consultant team in publications.Respect and acknowledge the intellectual contributions of cultural consultants, especially local cultural holders, and give them reasonable economic returns and respect for intellectual property rights.This is both an ethical requirement and an incentive for more sustained engagement.

When recruiting the position of "Dialect Resource Project Management", iFlytek is required to be responsible for "the whole process management of dialect resource delivery project" and "explore the dialect data resource reuse system and construction rules".This reflects the industry's professional trend in the construction and management of dialect data resources.The "Ma Xin Team" of East China Normal University supports real-time analysis of 28 dialects in its AI intelligent policy navigation system to break the communication barrier.This shows that considering language diversity (including dialects) can effectively improve the coverage and accuracy of services in specific application scenarios."The Wenju team created a "voice-text integrated dialect display platform", obtained dialect pronunciation corpus through "offline collection + online user contribution", and trained AI voice models.This method combines offline in-depth grassroots collection and online community contributions, which is conducive to the continuous accumulation of corpus and the cultivation of the community.

Implementing the principle of "ethics first" and the specific strategy of "cultural advisor system" is a key step for publishing organizations to fulfill their cultural responsibilities and let AIGC technology truly empower the protection and activation of dialects.This requires publishing organizations to always maintain awe of culture while embracing technical efficiency, and to ensure that every local accent resurrected by AIGC does not distort the soul behind it.

4.4.3. Man-machine collaboration, the pursuit of "thermal publishing"

A hybrid production model of "real human emotion template + AIGC batch generation" is adopted. That is, the local speaker records the core emotional paragraphs, and the narrative content is completed after AI learning, and a balance is achieved between efficiency and "human touch".

The core concept of the "real human emotion template + AIGC batch generation" model is to combine the unique human emotion expression with the efficient reproduction ability of AI. It is not a simple division of labor, but a deep integration. Local speakers, especially those "cultural holders" who are well versed in local culture and language, record core emotional passages, characteristic vocabulary or key narrative chapters. These recordings are not only the voice presentation of the text, but also carry the rhythm of the dialect, personal sense of language, improvisation and fullness of emotion, and inject the soul into the whole work.

"AIGC learning and completion" realizes scale, using high-quality voice cloning technology, allowing the AI system to deeply learn the timbre, intonation, rhythm and even breath of the real human model. Subsequently, AI can generate narrative, descriptive and other content parts with relatively low emotional requirements in batches based on the learned features, thereby greatly improving production efficiency. This model cleverly balances "efficiency" and "human touch". Real people guarantee the authenticity of culture and the depth of emotion, while AI solves the problems of long production cycle and high cost of pure human production, making it possible to quickly and soundly record a large amount of local documents.

People's Education Publishing House uses AI re-engraving technology to synthesize text reading audio in the audio production of textbooks. Its president Huang Qiang said that the use of this technology has a high degree of reduction and a high degree of anthropomorphism in synthetic audio. This method not only guarantees the unity and efficiency of the tone, but also strives to preserve the smoothness and naturalness of the real voice.

The collaborative creation of the AIGC picture book "Waiting for the Rabbit": Some developers used the Tencent hybrid model to create the picture book Waiting for the Rabbit, and adopted the collaborative mode of "human conception and guidance + AI generation and execution". The developer is responsible for project conception, technical selection, and writes detailed and graphic text descriptions (hints) for each scene, and then uses the AI text-generated graphic model to draw illustrations. The whole process requires human beings to continuously make aesthetic judgments and fine-tuning instructions to ensure the final quality and style orientation of the work.

The application of technology also needs to pay attention to the ethical and copyright aspects. For example, although the Human Pioneers News Agency opens its content for free use, its organizational form and mode of operation itself do not carry out copyright protection as a special case. Usually, in the "human-machine collaboration" project, it is the key to clearly define the contribution of human creators and AI, respect and protect the rights of real-person speakers.

4.4.4. Beyond audio, create a fusion scene experience

Promote the linkage of dialect audio publishing with local cultural tourism, education, cultural museums and other industries, and develop cultural experience projects that combine online and offline, so that the published dialect content can return to its cultural soil and achieve a true "activation" inheritance.

From sound archiving to cultural activation, pure sound recording can retain the "shape" of the dialect, but it is difficult to carry the "soul" on which it depends for survival. The vitality of dialects comes from specific life scenes, social relations and ritual situations. Once out of these contexts,

dialects are easily reduced to "local accent wonders" that are hunted for, and their profound cultural connotations will inevitably be flattened.

Therefore, the higher responsibility of publishing is to act as a bridge to guide the audience from "hearing" dialect to "understanding" and "feeling" the whole cultural ecosystem behind it, and to build a "dialect +" fusion ecology.

The deep linkage of dialect audio content with local cultural tourism, education, cultural exposition and other industries can effectively break the isolated island state of audio products, return them to the cultural matrix, reshape the local experience with local accents, and introduce dialect audio into local tourism scenes, so that tourists can deeply perceive the charm of local culture through local accents during on-the-spot tours.

The "Sound Tour Seven Doors and Nine Continents" dialect voice guide system created by the team of Southeast University in Nanchang is an excellent example. They set up QR codes in famous scenic spots such as Tengwang Pavilion and Bayi Yi Memorial Hall, and visitors can scan the codes to listen to the explanations recorded in pure Nanchang dialect. These audios not only tell history, but also incorporate folk proverbs and opera elements, making the tour a "voiced local cultural card". This move makes dialects no longer specimens in museums, but a living medium for tourists to explore the city and understand history.

Shanxi Province's "Sound Motion Intangible Heritage Reading Program" explores lighter scene integration. In Taiyuan Subway Station, passengers only need to touch the NFC tag with their mobile phones to listen to intangible cultural heritage items such as Lingchuan steel plate books and Hequ folk songs sung or performed in dialects. This design transforms the daily commuting space into a flowing cultural exhibition hall, allowing the original local folk songs to be cleverly integrated into modern urban life and return to its "fireworks".

In the museum scene, the use of dialects can add intimacy and dramatic tension to the heavy history, and make cultural relics "live". Attempt of "Cultural Relics Dialect Show": Sichuan Guan News Dolphin Radio planned the "Cultural Relics Dialect Show" in the special edition of Audible Cultural Relics-Hearing the Museum in cooperation with many museums in Sichuan. For example, let the Han Dynasty storytellers use Sichuan dialect to "swing the dragon gate array", or let Shi Xi speak "Trump" (Sichuan-style Mandarin). This bold creativity uses familiar local accents to narrow the psychological distance between the audience and cultural relics, and creates a unique cultural experience between the solemnity of history and the humor of dialects.

Carrying out integrated education in schools and communities is a fundamental plan to ensure the intergenerational inheritance of dialects. In the face of the severe reality that the use rate of young people's dialects has dropped significantly, the practice of Yunxiao County, Fujian Province provides a systematic sample. They offer dialect classes in pilot schools, resolve language conflicts through "bilingual teaching", and organize "children's voice chanting local accents" activities to allow children to recite classics in dialects. In the community, they not only opened the "Cuobian School", but also launched the "Dialect Family Pack" to build a language bridge between grandparents and grandchildren, and even introduced the "AI Dialect Doll" to help the elderly inquire about information. This "five entry" project of "entering campuses, communities, rural areas, government agencies, and enterprises" aims to create a full-coverage dialect inheritance ecological network.

When promoting Hakka culture, Taiwan's Yuanlin City specially produced a two-voice audiobook teaching material. The uniqueness of this audiobook is that it uses "the Minnan dialect of Yuanlin accent plus Raoping guest language" for two-tone recitation, and is recorded by the local guest language teacher and the residents of Yuanlin accent, which truly retains the diversity of local

languages. It is used in communities and workshops to allow residents to re-examine their cultural roots in participation.

The ultimate goal of the strategy of “transcending audio” is to realize the sustainable development of dialects by creating social value and economic value. AI translation is the only cost-effective way if a publisher wants to publish audio versions of 12 different regional dialects of India for a book in a short time.” [4]

When the dialect can be combined with the industry, it can generate internal motivation. Gao City's “Gao People Say Good Hometown” cultural brand invites township representatives and sages to serve as “hometown spokespersons”, telling local stories through short videos and audios, and successfully building an emotional bridge connecting urban and rural areas. In this way, wanderers who are working hard outside can “travel to their hometown” and relive the feelings of the hometown, which greatly enhances the sense of identity and cohesion of the regional culture. Creating a new paradigm of dialect communication and providing the power of language development, language, as an important non-material cultural element, is an important force to promote historical development and social progress, and is the fundamental guarantee for the occurrence, development and transmission of culture. Therefore, the development and application of language resources should be regarded as the top priority in the current language work. In the era of new media, the deep structure of language resources is being reorganized. How to maintain the vitality and vitality of dialects depends on the deep excavation, reshaping and re-dissemination of new media, paying attention to the accuracy and individualization of dialect cultural communication, and attaching importance to the innovation of dialect publishing [5].

“Beyond Audio” means a fundamental change in the role of publishing, from a content provider to a cross-media and cross-scene cultural experience planner. By re-embedding dialects into the network of social life, we are not only guarding a language, but also the diverse regional cultural ecology and identity behind it. Editors and reporters should not only master the operation skills of AI technology, but also understand the scope of application of AI technology, such as clarifying the mechanism of speech recognition accuracy affected by dialect accent and background noise intensity, and mastering the parameter adjustment method of style matching in image recognition. Although AI has penetrated into the whole process of news production, news is essentially the deep processing and value judgment of human information [6]. Only in this way can we flexibly formulate strategies in actual operation and avoid the homogenization of creativity due to technical dependence.

5. Conclusion

AIGC technology has brought unprecedented scale and democratization potential for dialect publishing, but the mission of the publishing industry is far more than “copying the voice”, but also “inheriting the context”. In the “polyphony” of technology, publishers must play the role of “tuner”, control the power of technology with cultural awe, and find a balance between “activation” and “alienation”. Only in this way can the publishing industry use AIGC to truly take on the responsibility of the times to protect and spread cultural diversity, and to play the harmonious resonance of tradition and modernity.

References

- [1] Wang Jianfei, Feng Liting, Chen Zihan. Intelligent publishing turn: industry applications, trends and challenges in the era of big model algorithm [J]. China Media Technology, 2025(07): 7-13. DOI: 10.19483/j.cnki.11-4653/n.2025.07.001.

- [2] Wang Yi, Jin Xiaopei. Integrated Linkage and Cultural Writing: Market-oriented Operation of Cultural Radio Programs—Taking Chengdu Story Radio Series as an Example [J]. Western Radio and Television, 2020, (11): 39-41. DOI: CNKI: SUN: XBGS.0.2020-11-019.
- [3] Zhong Yunhua. An Analysis of the Excellent Strategy and Media Convergence of Local Theme Publishing—Taking the Lingnan Cultural Audio Book Series "Painting the Sound and Coloring the Dialect" as an Example [J]. Publishing and Printing, 2020(02): 86-89. DOI: 10.19619/j.issn.1007-1938.2020.00.041.
- [4] Xu Weiyi. New Technology Enables Accelerated Change in Overseas Book Industry [N]. International Publishing Weekly, 2025-02-19(001). DOI: 10.28263/n.cnki.ngjcb.2025.000030.
- [5] Zhou Yifan. Research on Dialect Communication from the Perspective of Omnimedia [D]. Shanxi University, 2020, DOI: 10.27284/d.cnki.gsxiu.2020.000558.
- [6] Yang Chunhui. Discussion on the new mode of TV news production in the collaboration of editors and reporters and AI [J]. The Cradle of Reporters, 2025, (10): 144-146. DOI: CNKI: SUN: ZZYL.0.2025-10-048.