Strategic Research on the Integrated Development of the Publishing Industry in the AI Era: A Case Study of AI Audiobook Business Based on the "CHESS Strategy" Postprint

Authors: Zhang Cong, Ding Yanqing

Date: 2023-03-24T00:00:00+00:00

Abstract

Artificial intelligence technology has witnessed rapid development, and the publishing industry is actively integrating new technologies under the guidance of the "Four All Media" concept. AI voice technology (TTS, ASR) represents the earliest implemented and most widely applied direction in artificial intelligence. This paper examines the strategies for publishing houses to develop audiobook businesses based on AI voice technology within the theoretical framework of the "CHESS Strategy". First, it employs the "media affordance" theory to elucidate the necessity for the publishing industry to develop AI audiobooks. Then, it analyzes the dilemmas faced by publishing houses participating in the AI audiobook industry through six dimensions of industrial driving forces: technology, policy, market, talent, management, and assets. Finally, it proposes recommendations and prospects for publishing houses to undertake AI audiobook businesses.

Full Text

Research on Strategies for Integrated Development of the Publishing Industry in the AI Era: A Case Study of AI Audiobook Business Based on the "CHESS Strategy"

Zhang Cong[1], **Ding Yanqing**[2] [1]Beijing Institute of Graphic Communication, Beijing, 102627 [2]Beijing Film Academy, Beijing, 100091

Abstract

Artificial intelligence is developing rapidly, and under the guidance of the "Four-Dimensional Media" concept, the publishing industry is actively integrating new technologies. AI speech technology (TTS, ASR) represents the earliest and

most widely applied direction of artificial intelligence. This paper examines the strategies for publishing houses to develop AI speech technology-based audiobook businesses within the theoretical framework of the "CHESS Strategy." First, we explain the necessity of developing AI audiobooks in the publishing industry through the lens of Media Affordance Theory. Next, we analyze the dilemmas faced by publishers participating in the AI audiobook industry across six dimensions: technology, policy, market, talent, management, and capital. Finally, we propose recommendations and prospects for publishers to develop AI audiobook businesses.

Keywords: Audiobooks; AI Speech Technology; Media Affordance; Industrial Convergence Theory

In recent years, China's AI speech technology has entered a period of rapid application and implementation, achieving significant breakthroughs in emotional speech synthesis and natural semantic understanding compared to before 2016. Domestic and international enterprises have opened their speech ecosystems, applying AI speech technology to various scenarios through "industrial convergence," achieving successful commercial applications in short video creation, virtual idols (anchors), intelligent customer service, smart education, smart automobiles, and other industries. The two major branches of AI speech technology—Text-to-Speech (TTS) and Automatic Speech Recognition (ASR)—are naturally suited for application in the publishing industry, which works primarily with text editing. In the rapidly developing "audio economy," publishers can apply AI speech technology to solve the problem of weak audio content production capabilities, rapidly generating audio content that is difficult for human ears to distinguish from human narration at extremely low cost, thereby gaining greater development space in the "industrial convergence" of publishing, online audio/video, and AI.

1. Explaining the Necessity of Developing AI Audiobooks in Publishing Through Media Affordance Theory

The concept of "affordance" originates from psychology, originally referring to the potential possibilities for action that organisms (or behavioral subjects) have in a physical environment, deriving from the interaction between the subject's subjective perception of utility and the objective qualities of technology [1]. Later introduced to new media research by communication scholar Pan Zhongdang, it has become a systematic media research framework for understanding new media phenomena and evaluating the development potential of new media technologies, forms, and structures [2]. Media affordance emphasizes how technology shapes communication practices, arguing that new "affordances" brought to audiences during the process of media transformation by new technologies are not predetermined goals of media innovation but rather emerge from the interaction among technology, media, and audiences in communication practice. Therefore, the media affordance theoretical framework provides new guidance for the convergence and innovation of traditional media: the goal of media

convergence is not simply to graft new functions onto old media using new technologies but to generate new "affordances" through the interaction between audiences and media. Facing the burgeoning AI technology, the strategic direction for the publishing industry to stimulate new growth space lies in exploring new "media affordances" in communication practice to provide audiences with rich and efficient media experiences.

According to iResearch's "2021 China Online Audio Industry Research Report," China's online audio industry scale is expected to reach 22.9 billion yuan in 2022, with audiobooks still in a thriving development state, and AI audiobooks rapidly rising to a position on par with human-narrated audiobooks. AI audiobooks convert text information into audio signals with human emotions and linguistic characteristics through speech synthesis technology (TTS), and then complete conversational human-computer interaction through speech recognition technology (ASR). In the production process, AI audiobooks replace the manual processes of "script marking," auditioning, recording, listening review, and post-production required for human-narrated audiobooks with fully automated long-text understanding, speech synthesis control, and automated post-production, even enabling "dialogue" with audiobooks through voice interaction.

The history of audiobooks predates printed books. Before the invention of printing machines, "minstrels" who practiced "mouth-to-ear transmission" profoundly influenced the cultural life of the common people, and China's "storytelling" art directly impacted the development of popular literature. In 1877, Edison's invention of the phonograph marked the entry of audiobooks into the "mouth-machine-ear transmission" stage. Initially, audiobooks were widely used in public welfare projects targeting the blind and children. With advances in information technology, audiobook development has gone through three stages: cassette tapes, CDs, and digital formats, and has now become an important component of the publishing industry. Between 2011 and 2013, Qingting FM, Ximalaya, and Lazy Audio successively launched mobile clients, marking China's audiobook industry's entry into the mobile internet era. At the industrial level, comprehensive online audio platforms and vertical audiobook platforms engage in differentiated competition, while e-book platform listening functions, knowledge payment platform listening functions, and reading WeChat official account listening functions coexist harmoniously, leaving traditional publishers in a weakened position.

Catalyzed by AI speech technology, audiobooks have begun a cyclical communication model of "machine production-machine 传播-user experience-human-computer interaction." Below is a brief analysis of this model from the perspective of media affordance (see Figure 1 [Figure 1: see original paper]). AI speech technology primarily shapes the production 环节 of audiobooks, with new affordances manifested as: production cycles shortened from months to real-time synthesis; preliminary script marking, mid-stage dubbing, and post-stage listening review and packaging can be fully automated by machines; marginal costs dominated by technology production factors approach zero; and AI au-

diobook refinement can be achieved through iterative technology and offline synthesis. Improvements in production affordances make content supply 端 pay greater attention to AI audiobook applications, with platforms such as digital publishing "Qidian," Q&A community "Zhihu," and news media "Caixin" comprehensively adopting AI speech technology. Against the backdrop of integrated media and the audio economy, these platforms originally skilled in graphic content can rapidly build their own audio communication capabilities at low cost by introducing technology. On the reader experience 端, AI audiobooks provide more accessible and affordable audiobooks, personalized voice selection, diverse reading scenarios, and barrier-free voice interaction. The popularization of AI speech technology enables one-click conversion from text to audio, and multimedia reading combining vision and hearing brings better experiences to readers. For example, writer, poet, and short video creator "Xuyi" (Douyin ID: xuyi59) uploads poems synthesized by AI speech technology to the Douyin platform, receiving approximately 200 times more likes than his ordinary graphic works. Currently, AI audiobooks still have problems with unnatural emotion, pauses, stress, intonation, and speed, but the transformation of content production and communication by technology that creates new media is an entropy-increasing process—developing from simple to complex and irreversible. As AI technology advances, AI audiobooks are expected to become the mainstream form of future audiobooks, and publishers should seize the opportunity of rapid AI audiobook development to explore new business growth points.

[Figure 1: see original paper]

2. Problems Faced by Publishers in Developing AI Audiobooks

2.1 Incomplete Applicability of Current Copyright Law Increases Industrial Development Uncertainty

Copyright law is a product of printing technology, and its emergence and development have always been closely linked to technological progress, manifested in the continuous expansion of copyright objects and the enrichment of work utilization methods [3]. Secondary creation of works based on AI speech technology still falls into a fuzzy category under China's current copyright law. Whether AI speech synthesis infringes on the performance rights or reproduction rights of works, what legal differences exist between real-time and non-real-time speech synthesis, whether audio works completed through AI speech technology secondary creation have copyright, and who owns the copyright of works created by AI speech technology imitating a specific real person's voice—these issues have not yet formed legal consensus. Due to the lack of clear legal definitions, publishers are more likely to encounter economic disputes in the process of producing and operating AI audiobooks, and the lagging copyright system adds enormous risks to this business.

2.2 IP Fever Destroys Original Ecology, Full Copyright Library Construction Lacks Sustained Momentum

The core of the "content industry" is IP, around which various cultural industry operations such as film and television adaptation, game development, music creation, secondary creation, and derivative development can be conducted, generating greater economic benefits [4]. As a content form expressed through sound, audio works have been incorporated as an important part of the IP ecological industry. However, under this background, publishers are finding it increasingly difficult to obtain full copyright authorization from authors. The main reasons are that publishers lack the capability for full copyright operation, or high-quality IP has already had other rights granted before publication. For publishers, high-quality IP is both a core resource and a scarce resource, around which multiple economic benefits can be developed. However, when signing new books, publishers can often only obtain book publishing authorization, and the lack of capability in building a full copyright library is like "cooking without rice" for publishers to develop AI audiobook businesses.

2.3 New Business Urgently Needs New Technical Talent, Backward Management Exacerbates Vicious Cycle

Developing new businesses requires continuous investment in large amounts of professional human resources. Although technological progress improves work efficiency, the application of new technologies also places higher demands on human quality. The traditional publishing industry is a knowledge-intensive industry that gathers a large number of excellent editorial talents, but in the digital information era, the importance of technology and operations is increasingly prominent. Issues such as lack of innovation in corporate culture, rising human resource costs, and lack of incentives in human resource management make it difficult for traditional publishers to attract new types of talent while causing the loss of existing excellent talent, creating a vicious cycle over time. According to the "2021 Fourth Quarter China Enterprise Recruitment Salary Report" released by Zhaopin, the average salary in the publishing industry is 9,073 yuan per month, ranking 35th among 48 industry categories, placing it in the middle-to-lower range overall [4]. In the context of a highly marketized human resource environment, salary is becoming a decisive factor in career choice, and publishers' attractiveness to talent is gradually declining. The production and operation of AI audiobooks require talent with technical foundations and operational experience, and the recruitment and training of such talent require publishers to continuously invest substantial costs.

2.4 Integrated Publishing Increases Complexity of Content Production and Dissemination, Making Full-Platform Operation Difficult to Manage

Integrated publishing requires publishers to have the capability of "one-time production, multiple processing, multi-functional services, and multi-carrier (chan-

nel) dissemination," with each corresponding 环节 requiring investment in professional human resources with technical or experiential expertise. Among these, multi-functional services and multi-carrier (channel) dissemination primarily refer to "full-platform operation capability." After AI audiobooks are completed, they enter the operation stage. Unlike traditional book distribution, virtually existing AI reading materials are content service products, where content quality and service experience jointly determine readers' reading experience. Moreover, their dissemination capability is unrelated to replication (printing) quantity but related to the platforms and media of dissemination, with full-platform operation often achieving better dissemination effects. Corresponding to full-platform operation capability is the need for larger operation teams, with each additional dissemination platform or medium requiring multiplicative growth in operational human resources. AI audiobooks are just one of many content forms, and small-to-medium-sized publishers lack the capacity for full-platform operation.

In October 2021, the "Anti-Monopoly Law of the People's Republic of China (Amendment Draft)" underwent its first review, marking the first amendment to the Anti-Monopoly Law in its 13 years of implementation and releasing a strong regulatory signal to combat platform monopoly [5]. Benefiting from China's tolerant and prudent regulatory attitude toward new business forms and models, the internet and artificial intelligence industries have developed rapidly. In the AI audiobook industry, several platforms with monopolistic advantages have already formed in 产业链环节 such as content distribution and technical support. Super platforms' market power is too strong, seriously endangering fair market competition and technological innovation. Strengthening supervision of digital platforms from legislation (amendment) to law enforcement has become a global consensus [5]. Numerous small-to-medium-sized publishers and book companies are being crushed by powerful platforms in the industry chain. In the upstream of the industry chain, platforms such as "Yuewen" and "Jinjiang" control IP output, "iFLYTEK" basically holds a dominant position in AI speech technology services, and digital audiobook distribution platforms also fall within the sphere of influence of tech giants like "BAT." Small-to-medium-sized publishers and book companies basically have no discourse power before super platforms.

Integrated publishing has changed content production, accelerated technology integration, enriched dissemination channels, and increased service types, leading to increasingly complex collaboration and division of labor throughout the entire industry chain. Publishers in the middle of the industry will face more difficult industrial collaboration issues. AI audiobook industry collaboration requires 协同 and complementary advantages between upstream and downstream of the industry chain. The traditional cooperation model between publishers and technology platforms basically involves publishers providing content, technology companies providing technical support, and platforms providing traffic. However, as technology platforms implement content ecosystem strategies, their businesses begin to expand upstream and downstream, attempting to control the entire industry chain process to obtain greater economic profits. For exam-

ple, through its layout in the content ecosystem, “Tencent” has already obtained full industry chain capabilities for AI audiobooks from IP to production to distribution. The AI audiobook industry has already shown a Matthew effect of resources under the industry chain in its early development stages, with publishers being marginalized in the process of industrial collaboration.

3. Paths for Publishers to Optimize AI Audiobook Products

The “CHESS Strategy” is a classic model of “industrial convergence theory,” explaining the measures enterprises need to take to achieve integrated development. In “CHESS,” “C” represents creative integration, “H” represents horizontal organizational structure, “E” represents establishment of industry standards, “S” represents economies of scale and scope, and “S” represents systematic focus on processes. Constructing a publisher integration development model for the AI audiobook industry based on the “CHESS Strategy” (see Figure 2 [Figure 2: see original paper]) and elaborating on specific path strategies can help the publishing industry and high-tech industries intersect, penetrate, and reorganize on the basis of technological and institutional innovation, forming new content industry forms.

3.1 Development Iteration: Phased Product Optimization to Improve Effectiveness and Efficiency

Publishers’ content production rhythm is slower compared to emerging media, mainly producing in-depth reading 精品 content. A book takes months or even years from topic selection to distribution, and version updates occur annually or may not occur at all. However, in the information age, things change rapidly, content has strong timeliness, and readers’ preferences and demands force continuous optimization and iteration of content. Technological updates also require continuous improvement and iteration of content forms and dissemination methods. For AI audiobooks, under conditions where publishers lack content production experience and AI speech technology is not yet fully mature, publishers need to accumulate production experience and adapt to technological upgrades through content product iteration to produce content products that continuously meet readers’ new demands. Compared with the traditional waterfall development model that aims to complete a complete system project, the iterative approach divides the entire project goal into small, easily executable tasks according to logical structure. Through iterative development, AI audiobooks can be quickly brought to market, and then the system can be continuously iterated based on user feedback, adding new functional modules to achieve high-quality, high-efficiency AI audiobooks. For example, “CITIC Academy” built by CITIC Publishing Group since 2017 initially focused on digital reading. After multiple iterations and introducing iFLYTEK’s AI speech technology, it has now developed into a full-form, systematic multimedia knowledge service platform including text, audio, and video, with a large number

of readers choosing to pay for audiobooks generated by AI speech synthesis technology.

3.2 Operational Differentiation: Leveraging Long Tail Effect and Differentiated Competition with Head 精品

Currently, human-narrated audiobooks still dominate the market. Taking Ximalaya, the platform with the largest market share of audiobooks, as an example, although it has launched a large number of audiobooks generated by AI speech technology, the head works at the top of the rankings are all recorded by well-known anchors, with voice actors highlighted as selling points in titles. Moreover, in knowledge payment and vertical content fields, the role of key opinion leaders is difficult to replace, and the emotionally delicate auditory experience and more free secondary creation of human-narrated audiobooks are difficult for AI speech technology to achieve in the short term. Therefore, the commercial path for AI audiobooks requires a differentiated strategy, leveraging their advantages of low cost, short cycle, and rapid mass synthesis to focus on 腰部 and tail works—a strategy that 恰好符合 the Long Tail Effect.

Tomato Changting, a new entrant in the online audio track focusing on free audio, has gained competitive advantages by actively introducing AI speech technology. In its content classification, "Human Narration" and "AI Narration" are 并列排布 as important classification tags, and "AI Narration" is 接近 "Human Narration" in three important indicators: audiobook quantity, listener numbers, and ratings. Publishers should divide their reserved IP resources, recording head IP independently or authorizing third parties to produce 精品 human-narrated audiobooks, while generating AI audiobooks from 腰部 IP at low cost and in large batches to achieve Pareto Optimality.

3.3 Business Platformization: Building Content Distribution Platforms to Promote Comprehensive Operations

In the Web3.0 era that emphasizes information integration and value distribution, publishers urgently need to build autonomous content distribution platforms to 掌握主动权 and reduce dependence on super platforms. Currently, publishers have two main paths to build platforms: first, developing mini-programs through social media traffic portals; second, building website (App) platforms through content, services, and brands. The first path has advantages such as low promotion costs, low development thresholds, no need for user downloads, good operation experience, and strong webpage display compatibility [6], but while leveraging social media traffic, it also deepens dependence on them, with disadvantages such as deep entry points, simple functions, instability, and poor content dissemination effects. According to the "2021 Mini-Program Internet Development White Paper" released by the Aldzs Research Institute, the number of mini-programs across the entire network has exceeded 7 million, with WeChat mini-program developers surpassing 3 million and mini-program DAU exceeding 450 million; daily average usage frequency increased by 32%

year-over-year, while active mini-programs increased by 41% [7]. Among them, People's Literature Publishing House, Zhonghua Book Company, Higher Education Press, and other publishers have launched mini-programs. Overall, mini-programs are more suitable for publishers to optimize services and promote content payment. The second path is more difficult for small-to-medium-sized publishers and is not suitable for all publishers, requiring them to have the ability to provide irreplaceable services or products. However, its advantages are also obvious: the establishment of website (App) platforms will strengthen their moats. For example, "China University MOOC" under Higher Education Press is a successful case, applying AI speech recognition technology to quickly generate subtitles for audio and video content. However, building autonomous content distribution platforms does not mean abandoning platforms controlled by internet giants. On the contrary, publishers should strengthen their full-platform operation capabilities for AI audiobooks, which is 既有利于 enhancing dissemination effectiveness and 也有助于 curbing super platform monopolies.

3.4 IP Productization: Conducting Marketing with Product Thinking and Harmonious Coexistence with Distribution Platforms

The term "product manager" has frequently appeared in the publishing industry in recent years. The introduction of product managers into the publishing industry is a product of "industrial convergence" development and an inevitable requirement of internal operational mechanisms [8]. Although book marketing specialists and book product managers have different division of labor, the success of bestsellers requires book product managers to fully consider marketing impact throughout the entire closed loop from topic development to after-sales service. The product creation process of AI audiobooks must also fully consider marketing links, both to maximize the commercial value of IP and to enhance the sustained influence of IP. Large platforms such as WeChat, Douyin, and Ximalaya provide more effective channels for the dissemination of AI audiobooks, and the relationship between publishers providing IP content and platforms providing traffic is one of harmonious coexistence. According to the "2021 China Online Audio-Visual Development Research Report," Ximalaya's user penetration rate reaches 67.1%, firmly occupying the first-tier position in the online audio industry, with an average of 268 million monthly active users across all platforms. Therefore, publishers developing AI audiobooks need to increase content distribution on online audio platforms such as Ximalaya FM, just as they do for human-narrated audiobooks, and this does not conflict with building autonomous content platforms. "Industrial convergence" not only changes the market structure and industrial performance at the micro level but also changes a country's industrial structure and economic growth mode at the macro level [9]. The "industrial convergence" of content publishing and AI technology can not only reduce enterprise costs but also serve as an important method and means for traditional industry innovation, 有利于 the transformation and upgrading of the publishing industry structure and enhancing national cultural competitiveness.

3.5 Technology Servitization: Win-Win Cooperation with Technology Enterprises and Supporting the Servitization of Technology Products

The foundation of "industrial convergence" is technological progress and deregulation. Li Jin, General Manager of Global Technology Services at Alibaba Cloud, proposed that "the inevitable path for all technology enterprises is from technology to products, and then from products to services." A new economic form of industrial internet is taking shape, reshaping and transforming the industry chains of various vertical sectors. The publishing industry should actively utilize technology service products provided by information technology and internet platforms, increasing the proportion of technology production factors in its content production and enhancing publisher productivity through technological innovation. Currently, technology product servitization is showing characteristics of technology platformization, cloudification, standardization, and foundation, as well as service integration, diversification, personalization, collaboration, and cross-industry capabilities. "Industrial convergence" has changed the competitive and cooperative relationships between enterprises. Open platforms such as iFLYTEK, which focus on AI speech technology, achieve win-win cooperation with various industries including publishing through providing technology service solutions. The technology applied to AI audiobooks needs to possess capabilities in audio sampling and encoding, speech recognition database matching, speech-to-text, long-text understanding, emotional speech synthesis, and automated post-production, each of which requires advanced technology reserves. For example, emotional speech synthesis needs to match text emotions with voice emotions and add pauses, stress, intonation, and speed effects that conform to human language habits. Excellent synthetic speech can exceed human ears' ability to distinguish voice emotions. Currently, emotional speech synthesis remains an industry challenge, with commercial AI speech synthesis technology's emotional differentiation basically at eight types or fewer. This demonstrates that AI speech technology has extremely high technical thresholds, making win-win cooperation between publishers and technology enterprises and supporting the servitization of technology products an inevitable choice.

[Figure 2: see original paper]

Note: Figure translations are in progress. See original paper for figures.