Online Journal of Communication and Media Technologies, 2026, **16**(2), e202624
e-ISSN: 1986-3497

OPEN ACCESS

Research Article

AI anchors from a uses and gratifications perspective: An exploratory study of past, present, and future trends

Yongbo Niu [1*]
0009-0001-4868-5008

Sharifah Sofiah Syed Zainudin [1]
0000-0002-4589-6176

Syafila Binti Kamarudin [1]
0000-0001-8671-8897

[1] Faculty of Modern Languages and Communication, Universiti Putra Malaysia, Selangor, MALAYSIA
[*] Corresponding author: gs75122@student.upm.edu.my

Citation: Niu, Y., Zainudin, S. S. S, & Kamarudin, S. B. (2026). AI anchors from a uses and gratifications perspective: An exploratory study of past, present, and future trends. *Online Journal of Communication and Media Technologies, 16*(2), Article e202624. https://doi.org/10.30935/ojcmt/18478

ARTICLE INFO

Received: 12 Jan 2023

Accepted: 1 Feb 2023

ABSTRACT

Artificial intelligence (AI) anchors are increasingly deployed in journalism, entertainment, and education, reshaping how audiences consume media content. Early systems such as Ananova demonstrated the feasibility of virtual presenters but lacked natural prosody and expressiveness. Recent innovations, exemplified by Xinhua's AI anchor and Microsoft Xiaoice, leverage deep learning-based speech synthesis and multimodal design to achieve greater naturalness, personalization, and interactivity. Drawing on the uses and gratifications (U&G) framework, this study investigates how AI anchors meet cognitive, affective, and trust-related needs in the Chinese context. This study combines acoustic analyses of pitch, formants and intensity with 29 semi-structured interviews, providing complementary insights at both the technical and interpretive levels. Results show that technological refinements enhance clarity and efficiency, but trust is the key mediator that enables functional gratifications to translate into emotional engagement. Participants valued efficiency and multilingual accessibility while expressing concerns about authenticity, credibility, and ethics. This research contributes theoretically by extending U&G to AI-mediated communication and empirically by combining technical and qualitative evidence to analyze user perceptions. Practically, the findings highlight opportunities for deploying AI anchors in routine news, education, and commercial contexts, while underscoring the continued role of human anchors in politically sensitive or emotionally rich communication. These insights add to recent debates on automated journalism and AI-mediated interaction (Jang et al., 2022; Wölker & Powell, 2021). Future research could expand these findings by conducting cross-cultural comparisons, testing the mediating role of trust in experimental settings, and exploring how emerging acoustic and multimodal technologies further shape audience acceptance

Keywords: AI anchors, uses and gratifications, speech synthesis, trust, audience acceptance

INTRODUCTION

The emergence of artificial intelligence (AI) anchors represents a significant innovation in the convergence of speech synthesis, natural language processing, and virtual embodiment. From the launch of Ananova in the UK in 2000 to the debut of Xinhua-Sogou's AI anchors in 2018, technological advancements have enabled 24-hour broadcasting and multilingual delivery. More recent examples, such as Microsoft's Xiaoice and Zhejiang TV's Gu Xiaoyu, employ neural models including WaveNet and Tacotron to achieve lifelike voice and visual presentation (Shen et al., 2018; van den Oord et al., 2016). Outside of the media industry, the concept of AI being embraced has become a larger trend across the world, with the aim of improving organizational

performance and efficiency in operations. Business research indicates that the implementation of AI can lead to a significant increase in productivity, decision-making, and cost-efficiency (Tuan & Tung, 2026). Placing AI anchors in the context of this broader trend assists in putting AI into perspective both as a technological innovation in journalism and a cross-sector change brought about by efficiency and scalability. These developments reflect a dual logic: innovation driven by advances in machine learning and cloud computing, and the audience's demand for immediacy, personalization, and natural interaction. The increasingly deep integration of AI in communication practices has also changed the way information is generated and transmitted, highlighting its advantages in terms of efficiency, while also revealing new challenges for communication processes centered on humans (Getchell et al., 2022).

Even if AI anchors are becoming more widely used, there are still issues with how the audience experience is affected by technology performance. Research indicates that when news is produced by algorithms, people are still dubious about its veracity and authenticity (Jang et al., 2022; Wölker & Powell, 2021), indicating a disconnect between perceived satisfaction and technological capabilities. Ethical issues related to transparency, bias, and employment displacement complicate their acceptance (Carlson, 2023; Diakopoulos, 2019). At the same time, research indicates that anthropomorphic cues—such as prosody, gestures, and emotional expression—can strengthen audience trust and engagement (Xue et al., 2022). Yet these effects vary by context, and empirical evidence on how acoustic features interact with audience gratifications remains limited.

To bridge this gap, the present study focuses specifically on audience perceptions of AI anchors as its primary unit of analysis, while treating technological features as analytical inputs and broader governance issues as contextual considerations rather than direct objects of empirical investigation.

Against this backdrop, this study addresses three interrelated research questions (RQs):

1. **RQ1.** How is the evolution of AI anchor technology understood in relation to technological progress and social demand based on selected cases and participant perspectives?
2. **RQ2.** In what ways do current AI anchors satisfy users' needs for information, entertainment, and interaction through multimodal generation, personalization, and multilingual support?
3. **RQ3.** What opportunities and challenges for the future development of AI anchors in China are identified through participant perspectives and case-based analysis, particularly in terms of technological innovation, cultural acceptance, and ethical considerations?

By engaging these questions, the study makes three contributions. Theoretically, it extends uses and gratifications (U&G) into AI-mediated contexts, highlighting a relational pattern linking functional gratifications, trust, and affective engagement. Empirically, it is among a limited number of studies that integrate acoustic analysis with qualitative interviews to explore Chinese audiences' perceptions of AI anchors. Practically, it provides insights for media organizations and policymakers on designing AI anchors that balance efficiency with authenticity, while addressing ethical and cultural concerns.

LITERATURE REVIEW

Uses & Gratifications Theory and Its Application to AI-Mediated Contexts

The U&G theory emphasizes the audience's active role in selecting media to meet cognitive, affective, and social needs (Katz et al., 1973; Rubin, 2002). Distinctions between instrumental and ritualized use provide a framework for understanding why people adopt certain media (Rubin, 1983). Modern adaptations integrate psychological motives such as self-determination with media affordances such as interactivity, customization, and anthropomorphism (Deci & Ryan, 1985; Sundar, 2020; Sundar & Limperos, 2013).

Although recent research expands U&G into AI-mediated settings, it also raises conceptual concerns about whether conventional pleasure categories adequately represent interactions with non-human communicators. Zhang et al. (2025) investigated AI-human collaboration in live-streaming e-commerce and found that "assisting" AI anchors enhanced playfulness more effectively than "supervising" collaboration, particularly when humor was included. Huang and Yu (2023) examined continuance intention toward AI news anchors, showing that anthropomorphism, attractiveness, novelty, and trust significantly predicted

satisfaction and ongoing use. In service scenarios with AI capabilities, user satisfaction is typically determined by the combined effect of perceived service quality and trust, rather than by a single factor. Nga's (2026) research indicates that system reliability, performance quality, and trust often work together to influence the overall evaluation of AI services by users. This viewpoint provides a useful theoretical perspective for understanding the way audiences assess AI anchors (rather than merely based on their functional characteristics). Recent studies have also shown that communication conducted through AI can simultaneously enhance interaction efficiency and shape users' social cognition and trust levels. This indicates that there is a complex relationship between the functional characteristics of technology and interpersonal evaluations (Hohenstein et al., 2023).

In addition to highlighting the need to reevaluate how trust and technology mediation modify satisfaction processes in automated contexts, these research show how flexible U&G is in understanding motivations, gratifications, and ongoing engagement with AI anchors. Besides reinforcing the idea that the mediation of trust and technology is necessitating a reassessment of the ways in which satisfaction processes in automated environments are modified, all these studies also indicate a general convergence in the appreciation of the significance of psychological antecedents like trust and anthropomorphism. However, existing studies show inconsistent evidence regarding how these factors translate into user satisfaction, suggesting that the relationship between technological features and gratifications remains underexplored. At the same time, building on the classic U&G framework, recent research highlights how anthropomorphism, attractiveness, and novelty of AI anchors shape audience engagement and continuance intention (Huang & Yu, 2023; Zhou & Li, 2022).

Outside the media environment, research on the adoption of AI in the corporate context further indicates that through automation and cost reduction, AI can enhance organizational efficiency and performance (Tuan & Tung, 2026). These findings provide a broader foundation for understanding the functional value of AI in news production.

AI Anchors and Virtual Anchors: Technical Advances, Applications, and Audience Response

Technological progress in AI anchors—from concatenative synthesis to neural network-based systems such as WaveNet and Tacotron—has greatly improved voice naturalness and multimodal integration (Tan et al., 2021; van den Oord et al., 2016; Wang et al., 2017). Beyond technical development, audience perceptions have become a key focus. For example, Wang et al. (2025) identified interactivity, frequency, accuracy, ease of use, and emotional connection as major determinants of audience choice of AI anchors. User engagement with AI systems is not uniform across populations. Prior research suggests that factors such as gender and AI self-efficacy can influence how individuals evaluate and interact with AI technologies (Olijo, 2025). In the service scenarios enabled by AI, user satisfaction is closely related to perceived service quality and system reliability, rather than a single psychological factor. Nga (2026) indicates that trust and perceived performance often work together to influence users' evaluation of AI systems.

Experimental comparisons reveal nuanced differences between AI and human anchors. Feng and Shu (2024) found that while human anchors were perceived as more credible, immersion and liking varied by news type. Design features also matter: a study demonstrated that non-humanoid female AI anchors paired with anthropomorphic voices generated higher perceived attractiveness, shaped partly by audience impressions of traditional anchors (Zhou & Li, 2022). Beyond news contexts, AI-human collaboration in live-streaming demonstrates how virtual anchors influence customer engagement and brand loyalty (Zhang et al., 2025; Zhu et al., 2025).

Applications extend into branding and loyalty. Zhu et al. (2025) reported that perceived usefulness, enjoyment, and novelty of AI virtual anchors significantly influenced brand image, which mediated effects on brand loyalty. These results, combined with previous research, suggest some overlap when it comes to determining factors related to design and psychology that are essential in creating an audience reaction, but they also to suggest discrepancies in terms of credibility, immersion, and attractiveness ratings across the range of experimental conditions and application scenarios. This sort of ambivalent evidence indicates that the current studies have not comprehensively explained how technical design attributes lead to consistent audience perception patterns.

Integrating U&G with AI Anchor Research: Gaps and Future Directions

Integrating U&G with AI anchor research reveals both areas of convergence and unresolved tensions. First, affective and social gratifications—such as entertainment, enjoyment, and social interaction—are consistently validated, whereas cognitive gratifications (credibility, professionalism, accuracy) remain contested. Second, such psychological variables as trust, attractiveness, perceived intelligence are often addressed in the previous works (Huang & Yu, 2023), but their theoretical positioning is inconsistently described. The understanding of trust introduced in this research is an appraisal construct, expressing the judgments of credibility and reliability in communication mediated by AI, instead of a satisfaction instrument or a variable statistically measured (e.g., Mayer et al., 1995; McKnight et al., 2002). Third, cultural and contextual factors shape acceptance: cross-national evidence shows varying levels of optimism, skepticism, and mainstream adoption (Lim et al., 2023).

Despite these advances, gaps remain. Few studies directly map technical parameters (e.g., pitch, formants, intensity, and anthropomorphic design) to audience gratifications, even though AI anchors are widely found to enhance user experience and audience interaction in news broadcasting (Lyu et al., 2023). Ethical concerns—authenticity, transparency, professional identity—are often acknowledged but insufficiently theorized (Jang et al., 2022; Reuters Institute, 2024).

To bring into the picture a more concrete relationship between the technical characteristics and the user gratification, the present research pays attention to the acoustic characteristics (pitch, formants, and intensity) as these features have been already demonstrated to have a direct reference to the perceived clarity, naturalness, and emotional expression in the perception of speech (e.g., Nass & Brave, 2006; Scherer, 2003). Past studies on human-computer interaction and media psychology also state that these perceptual qualities do not only determine the cognitive appraisal of the users (e.g., credibility and intelligibility) while they also influence the affective reaction (e.g., engagement and enjoyment), thus they make fitting analytical connections between quantifiable system characteristics and the U&G typology.

Research Gap

Although existing research has generated valuable insights into the technical foundations and social implications of AI anchors, several important gaps remain. First, while studies have examined the development of speech synthesis technologies such as WaveNet and Tacotron (Shen et al., 2018; van den Oord et al., 2016) and explored user responses to anthropomorphic cues (Xue et al., 2022), there is limited work directly connecting acoustic features (e.g., pitch, formants, and intensity) to audience gratifications. This leaves a disconnect between technological design parameters and how audiences actually perceive information clarity, credibility, and emotional resonance.

Second, although trust has been repeatedly identified as a decisive factor in automated journalism (Jang et al., 2022; Wölker & Powell, 2021), few empirical studies have systematically clarified its theoretical role in relation to functional gratifications (such as efficiency and accessibility) and affective responses (such as engagement and attachment), with existing studies variously treating trust as an outcome, an antecedent, or a contextual condition. Without this relational perspective, our understanding of audience acceptance remains fragmented.

Third, while comparative studies highlight differences in how audiences across cultures evaluate AI anchors (Carlson, 2023), qualitative, in-depth research on Chinese audiences remains scarce, despite China being a global leader in the deployment of AI anchors. This lack of culturally grounded evidence limits the generalizability of existing models of AI anchor acceptance.

Addressing these gaps, the present study integrates acoustic analysis with semi-structured interviews to explore how Chinese audiences evaluate AI anchors. By linking technical data to audience perceptions within a U&G framework, the study extends theoretical debates on U&G into AI-mediated communication and provides empirical evidence for the relational pathway connecting functionality, trust, and affective outcomes.

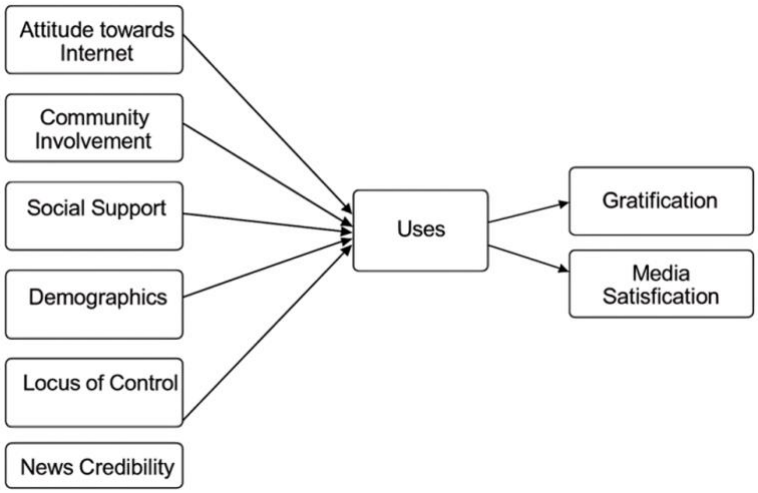

Figure 1. U&G framework (Katz et al., 1973)

METHODOLOGY

Research Design

As presented in this paper, the research takes an exploratory mixed-methodology design with qualitative-dominant orientation where qualitative enquiry forms the main analytical tool and acoustic analysis forms the auxiliary technical evidence placed in a post-positivist epistemological context to explore the formation and use of AI anchors.

According to mixed-methods typology, the research is closer to an exploratory design where various types of data are employed to form a deeper conception of the research issue instead of testing aqueous relationships (Creswell & Plano Clark, 2018). It is done at the interpretation level, where acoustic measurements are analyzed in conjunction with interview results to put into perspective and further insight on the audience perceptions as opposed to statistical aggregation (Fetters et al., 2013).

Post-positivist inquiry emphasizes the socially constructed and context-dependent nature of knowledge, highlighting the complex interplay between technology, culture, and power structures (Denzin & Lincoln, 2011). As AI anchors embody the convergence of AI, media technologies, and cultural communication, their evolution cannot be fully captured through quantitative indicators such as click rates or viewing time. Instead, qualitative approaches allow for a deeper exploration of the meanings, motivations, and socio-cultural implications underlying user engagement.

The research is guided by the U&G framework (Katz et al., 1973; Rubin, 2002), which views media users as active agents motivated by cognitive, affective, and social needs (**Figure 1**). By integrating this perspective with qualitative inquiry, the study seeks to uncover how audiences interpret and interact with AI anchors across different historical and cultural contexts.

Data Sources

Multiple sources of evidence were triangulated to capture the multi-dimensional nature of AI anchor development and use:

1. **Literature and industry reports:** A review of recent academic studies, industry white papers, and media reports was conducted to analyze the technological features and societal functions of AI anchors. For instance, Xie et al. (2023) explored the role of anthropomorphism and personalization in sustaining user satisfaction, while Huang and Yu (2023) discussed multimodal applications of AI anchors across education, entertainment, and journalism.
2. **Case studies:** Representative AI anchor systems were selected to trace the technological trajectory and user adoption patterns. These cases included various AI anchors, such as early systems like Ananova (2000, UK), and later developments like Xinhua's AI anchor (2018, China), Microsoft's Xiaoice (2020, China), and Zhejiang TV's Gu Xiaoyu (2022, China). The cases were chosen based on their

technological significance, application diversity, and socio-cultural relevance (Liu et al., 2025; Wiederhold, 2023; Tan et al., 2021).

3. **Semi-structured interviews:** Data were also collected through semi-structured interviews with 29 participants, aged 16 to 52, representing a diverse demographic engagement with AI technology. The participants were selected from a range of professional backgrounds, including media professionals, technology students, and academics. The interviews provided qualitative insights into user motivations, satisfaction, emotional responses, and expectations for future AI anchor developments.

The data collection process was done with much consideration to the ethics. The study purpose was explained to all the participants in a manner that they understood that they were participating in the study on a voluntary basis and they had a free right to practice a right to withdraw. All participants were informed of the need to participate in the interviews. In the case of the participant who was 16 years old, parental or guardian consent was provided along with the assent of the participant as per the clearly stipulated ethical principles of conducting a research with children or young people (American Psychological Association, 2017; British Educational Research Association, 2018). To ensure the privacy of the subjects, all data were anonymized and no further information about the subjects in the analysis or reporting was remembered.

The semi-structured interviews, which lasted between 10 to 31 minutes, were conducted in various locations, including cities such as Hefei, Kunming, Guangzhou, Chengdu, Nanjing, and Shandong. This diverse geographical representation ensured a broad range of perspectives. The sample size of 29 participants was deemed sufficient for thematic saturation, as research suggests that 6 to 12 interviews are typically enough to reach data saturation in qualitative studies (Guest et al., 2006; Hennink et al., 2019; Vasileiou et al., 2018). No significant new themes emerged during the interviews, indicating that the sample size was adequate for the research objectives.

Sample Description

The sample for this study was purposefully selected to ensure diversity in age, educational background, profession, and geographic location. The participants included students, media professionals, academics, and individuals with experience in various industries, such as communication, broadcasting, and technology. The sample was designed to reflect a broad spectrum of user demographics, ensuring that the findings were not limited to a single group but instead captured a wide range of perspectives on AI anchors.

1. **Age:** Participants ranged in age from 16 to 52, ensuring that both younger and older generations were represented. This range allowed the study to examine how different age groups perceive and engage with AI technology.
2. **Educational background:** The educational backgrounds of the participants varied from high school to doctoral levels. This diversity provided insights into how different levels of education might influence perceptions of AI anchors and their perceived usefulness.
3. **Professional background:** The sample included media professionals, communication students, and academic researchers, all of whom have varying levels of exposure to media technology. This range of professional experiences allowed for a comprehensive understanding of AI anchors' impact across different sectors.

This diverse sample helped ensure that the study captured a wide range of experiences and opinions, strengthening the external validity of the findings.

Data Analysis Methods

Qualitative techniques were combined in the data analysis of the information gathered by interviews. In this exploratory mixed-methods design, the thematic analysis and acoustic data would be combined during the stage of interpretation but acoustic features would only be used to provide a qualitative result in the manner in which the users felt the content was clear, natural, and engaging instead of providing a statistic relationship and causal response.

The interview guide was derived on U&G framework, with semi-structured questions suggested to extract both cognitive, affective, and social motivation of the users and their view on the performance and the constraints of AI anchors. The questionnaire consisted of structured questions and open-ended questions

that give the respondents liberty to express their experiences using their own words. Before actual data collection, piloting interview was undertaken to improve question phrasing, order and comprehensibility so that interview process could be understandable and able to generate rich qualitative information.

Thematic Analysis: Thematic analysis was conducted to identify recurring themes and patterns in the interview data (Braun & Clarke, 2006). Following the classification of instrumental and ritualized gratifications (Rubin, 1983), the interview transcripts were coded to reveal themes related to cognitive, affective, and social gratifications derived from AI anchors. This process involved reading the transcripts multiple times, generating initial codes, and then categorizing these codes into broader themes.

The process would entail reading the transcripts several times, coming up with initial codes and subsequently classifying these codes or codes as broad categories. The coding procedure was in a process which was informed by deductive and inductive approaches. A first coding scheme that is aligned to the categories of the U&G (e.g., informational, emotional, and social gratifications) was formulated and later narrowed down with the open coding techniques wherein additional themes were found on the data. Each code was described using a structured codebook to establish the inclusion criteria of each code and examples to be used across all of the coding, to make sure they are consistent throughout the coding process.

The researcher coded all the interview data, and two coding cycles were repeated to enhance internal consistency. In order to increase the level of reliability, portions of data were reread and cross-coded in cases where the decision made was related to a code. To encourage transparency and traceability between unstructured data and the claims of analysis, representative snippets were saved.

In an effort to make the qualitative analysis more credible, a number of strategies were applied based on the guidelines (Lincoln & Guba, 1985). Prolonged interaction with the data and the repetitive comparison of interviews in different interviews enhanced credibility. Reliability came with keeping a strong audit trail of the decisions on coding and theme development. The issue of confirmability was resolved with the help of the direct quotations of the participants and the reduction of subjectivity, accomplished by means of systematic coding. Lastly, the description of the sample and the context of the research allowed transferability in that the readers can also determine the relevance of the findings to other contexts.

1. **Content analysis:** Literature and case study materials were also analyzed to trace the technological evolution of AI anchors and their adaptation to user needs. Particular attention was given to advancements in speech synthesis, virtual appearance, and interaction modes.
2. **Acoustic analysis with Praat:** Praat software (Boersma & Weenink, 2023) analyzed voice data of the AIs anchors to generate suggestive acoustic variables of pitch, formants, and intensity. The acoustic data involved five AI anchor systems of various levels of technological development: Ananova (UK), Xinhua AI anchor, Microsoft Xiaoice, a Gu Xiaoyu of Zhejiang TV, and an AI Avatar (Synthesia) on sale. Six audio clips were studied, two of them of Ananova (10 seconds and 5 seconds) and one clip of each of the other systems (between 26 and 35 seconds).

All the audio material was recorded by means of iPhone 16 Pro based on the publicly available video sources such as media platforms and official demonstrations. The clips picked were chosen based on criteria that focused on clips that had clear speech with minimum or no background noises and representative news or presentation-like delivery so as to compare across systems.

The sample of Mandarin Chinese (Xiaoice, Gu Xiaoyu, and Xinhua AI anchor) and English (Ananova and AI Avatar) speech is contained in the dataset. Although the cross-linguistic differences are considered, the comparison between the languages is not placed on the aspect of a direct linguistic comparison but rather on general acoustic characteristics related to the perceived naturalness and clarity.

All audio clips had been converted to WAV format before analysis and processed to undergo basic processing, such as noise removal, and segmentation to divide them into continuous speech segments. Parameters that were extracted in Praat were fundamental frequency (pitch), first and second formants (F1 and F2), and intensity. All the samples were measured under similar settings so that they could be compared.

The researcher did the analysis and segments were chosen manually depending on continuity and clarity of speech. Alternatively, the resulting acoustic measures are taken as illustrative and comparable measures to put the audience perceptions of voice naturalness and credibility in perspective, instead of using them as

Table 1. Pitch analysis of AI anchors

Anchor/system	Mean pitch (Hz)	Minimum pitch (Hz)	Maximum pitch (Hz)
Ananova (2000, female)	175.24	121.73	239.90
Xinhua AI (2018, male)	110.05	63.63	197.97
Xiaoice v2 (2020, female)	274.43	57.69	417.43
Gu Xiaoyu (2022, female)	286.16	173.67	484.20
AI Avatar (2024, male)	105.46	46.91	220.13

Table 2. Formants analysis of AI anchors

Anchor/system	Mean F1 (Hz)	Mean F2 (Hz)	Mean F3 (Hz)
Ananova (2000)	668.87	1,877.42	3,025.36
Xinhua AI (2018)	752.07	2,052.47	2,944.11
Xiaoice v2 (2020)	883.96	2,176.61	3,034.07
Gu Xiaoyu (2022)	721.53	1,893.06	2,876.72
AI Avatar (2024)	832.36	2,036.60	3,004.83

the basis of statistical inference or causal statements. These tests are designed to provide relative technical background towards evaluating audience naturalness and believability as opposed to delivering causes and effect or normative properties.

Thematic analysis and acoustic data were integrated to form a comprehensive understanding of how AI anchors meet users' needs for information, entertainment, and interaction. Given the exploratory nature of the study, the qualitative findings are intended to provide interpretive insights rather than statistically generalizable conclusions.

DATA AND MATERIALS

Case Selection

To capture the historical trajectory and technological evolution of AI anchors, five representative systems were selected as case studies:

1. Ananova (2000, UK)–the world's first virtual anchor, based on early text-to-speech (TTS) and facial animation technologies (Wiederhold, 2019).
2. Xinhua-Sogou AI Anchor (2018, China)–the first large-scale commercial AI anchor capable of 24/7 news broadcasting in Chinese and English (Xinhua News Agency & Sogou, 2018).
3. Xiaoice v2 (2020, Microsoft China)–integrated conversational engines, emotional speech synthesis, and multimodal perception, enabling context-sensitive dialogue (Tan et al., 2021).
4. Gu Xiaoyu (2022, Zhejiang TV, China)–developed with Unreal Engine and real-time motion capture, representing China's progress toward hyper-realistic AI anchors.
5. AI Avatar (2024, global)–combining GPT, diffusion models, and multimodal generation, offering multi-language and emotion-synchronized broadcasting (Bohacek & Farid, 2024; Tan et al., 2021).

These cases were chosen because they reflect milestones in speech synthesis, visual realism, and multimodal interaction, as well as their adaptation to different socio-cultural and institutional contexts.

Technical Parameters

To assess the naturalness and quality of AI anchors' speech, three key acoustic features were compared across the selected cases: pitch, formants, and intensity. These parameters were measured using Praat software, with samples extracted from representative broadcast clips.

From **Table 1**, early systems such as Ananova had limited pitch variation, producing mechanical intonation. Later systems (Gu Xiaoyu and AI Avatar) achieved richer pitch ranges and smoother control, reflecting the impact of neural speech synthesis models like WaveNet and Tacotron (van den Oord et al., 2016; Wang et al., 2017).

Table 3. Intensity analysis of AI anchors

Anchor/system	Mean intensity (dB)	Maximum intensity (dB)
Ananova (2000)	67.50	80.00
Xinhua AI (2018)	65.66	74.36
Xiaoice v2 (2020)	65.31	76.80
Gu Xiaoyu (2022)	66.81	77.82
AI Avatar (2024)	60.05	69.91

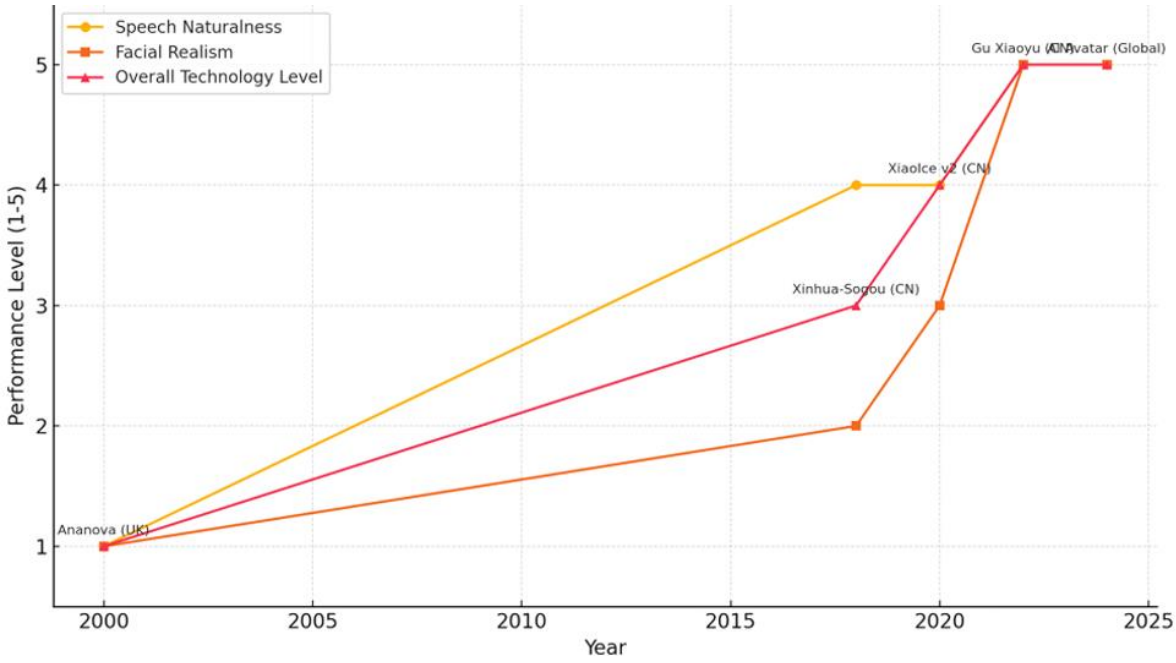

Figure 2. AI anchors: Technology evolution over time (Source: Created by the authors)

According to **Table 2**, improvements in F1-F3 distributions indicate increasing clarity and naturalness of vowel articulation. Particularly, Xiaoice v2 and AI Avatar demonstrated resonance frequencies closer to human speech norms, enhancing intelligibility (Peterson & Barney, 1952).

While average intensity values remained within standard broadcasting ranges, more recent systems (Gu Xiaoyu and Xiaoice v2) exhibited smoother loudness control, improving the naturalness of delivery. The relatively lower intensity of AI Avatar reflects its positioning as a general-purpose assistant rather than a news anchor (**Table 3**).

Based on the three parameters in **Figure 2**, it is evident that the naturalness of AI anchors' speech has significantly improved over the past 20 years. The early Ananova, although a pioneer in speech synthesis, still exhibited limitations in prosody and dynamic expression. The Xinhua AI anchor marked the practical implementation of deep learning-based TTS systems; however, constrained by early synthesis technology, its intonation and expressiveness remained somewhat mechanical. After 2020, Xiaoice and Gu Xiaoyu represented the practical application of neural network models such as WaveNet and Tacotron 2, with speech exhibiting more human-like control over pitch, clarity, and intensity. In particular, Gu Xiaoyu demonstrated highly optimized pitch control, formant structure, and loudness parameters, reflecting a more advanced end-to-end speech generation model (Zhang et al., 2025). AI Avatar, on the other hand, showed broader technological integration but may be limited by the need for differentiated speech styles—for example, the requirements for commercial versus broadcasting scenarios—and the samples were not specifically designed for broadcasting, resulting in lower loudness levels. Recent psychophysiological evidence suggests that audiences process AI-generated voices differently from human ones, affecting perceptions of credibility and emotional resonance.

These acoustic patterns to analyze the relationship between these and data defining audience perceptions are selected out of the interview material and linked to the acoustic patterns analytically. In particular, the differences in the range of pitches, the distribution of formants, and the intensity can be viewed as technical

Table 4. Frequency of themes identified (N = 29 interviews)

Theme	Frequency (f)	Percentage (%)
Functional and cognitive gratifications	21	72
Affective and emotional responses	18	62
Trust, authenticity, and anthropomorphism	15	52
Social and educational applications	12	41
Concerns and ethical considerations	10	34

Table 5. Themes, illustrative quotes, and related literature

Theme	Summary of findings	Illustrative quote	Related literature
Functional and cognitive gratifications	Efficiency, accuracy, and 24/7 availability were key gratifications	"AI anchors don't get tired and can work around the clock." (P4)	Huang & Yu (2023) & Rubin (2002)
Affective and emotional responses	Participants noted lack of empathy but acknowledged gradual improvements	"I don't feel any warmth—it's just information." (P12)	Lim et al. (2023)
Trust, authenticity, and anthropomorphism	Credibility questioned in sensitive news; anthropomorphic cues improved acceptance	"I can accept AI anchors for weather, but not for political news." (P7)	Zhang (2024)
Social and educational applications	Potential in education, e-commerce, and customer service	"An AI anchor in class could repeat explanations patiently." (P22)	Zhang et al. (2025) & Zhu et al. (2025)
Concerns and ethical considerations	Job displacement, bias, and transparency were main concerns	"If AI anchors keep improving, many human anchors might lose their jobs." (P9)	Jang et al. (2022) & Napoli (2021)

Table 6. Integrated analysis of acoustic parameters and audience perceptions

Acoustic feature	Perceptual dimension	Interview theme	U&G category
Pitch variation	*Emotional expressiveness*	*Affective response*	*Affective*
Formants clarity	*Speech intelligibility*	*Functional gratification*	*Cognitive*
Intensity control	*Natural delivery*	*Trust & authenticity*	*Social/trust*

correlations of the perceived naturalness, clarity, and expressiveness of emotions, as they were quoted numerous times by participants. This offers some foundation to merge technical measurements and qualitative in a systematic way.

Semi-Structured Interview Data

Thematic analysis of 29 interviews generated five major themes. **Table 4** summarizes the frequency distribution of themes, while **Table 5** presents representative quotations and links to prior literature. In order to systematically integrate the qualitative research results with the acoustic analysis, we constructed a comprehensive explanatory framework that linked specific acoustic features (pitch, resonance peaks, and intensity) with the corresponding perceptual dimensions identified in the interviews (e.g., naturalness, emotional engagement, and trust).

The reported percentages are descriptive rather than inferential. Each participant may be involved in multiple topics. These figures indicate the proportion of participants who mentioned each topic at least once, which is consistent with the practice of qualitative thematic analysis.

The findings indicate that there is an overall trend of opportunities and challenges related to AI anchors, as well as show that there is a systemic correspondence between acoustic features and audience perceptions (**Table 6**). To begin with, numerous interviewees noted functional and cognitive gratifications, especially efficacy, correctness, and 24/7 availability (72%). Such perceptions are similar to the more advanced systems (e.g., Xiaoice v2: F1 = 883.96 Hz; F2 = 2176.61 Hz) that have a better influence on speech intelligibility and aid in informational processing. This is consistent with the utilitarian aspect of the U&G model (Rubin, 2002) and previous perceived usefulness (Huang & Yu, 2023).

Conversely, there was little emotional involvement. Though 62 percent of the respondents said that they saw an improvement in voice tone and facial expression, AI anchors were often called cold or mechanical. This impression can be linked to the comparatively less varied up and down pitch change in previous systems like Ananova (average pitch: 175.24 Hz) versus broader pitch changes of Gu Xiaoyu (286.16 Hz) and Xiaoice v2

(274.43 Hz), which exhibited a perception of greater expressiveness. The results confirm the fears of emotional weaknesses of AI-mediated communication (Lim et al., 2023).

Another theme of great importance (52%) was trust and authenticity. Individuals were less convinced by the use of AI anchors in highly sensitive situations, although the acceptance also improved with anthropomorphic signals such as increased natural pitch modulation and consistent patterns of intensity. The given trend demonstrates the role of vocal naturalness in the formation of perceived credibility and aligns with previous studies on anthropomorphism and trust (Zhou & Li, 2022).

There were also public opinions about its bigger applications outside the news delivery, like education and e-commerce, where they preferred personalization and consistency. These perceived benefits are correlated with technical potential of AI systems to ensure unchanged acoustic output and constant delivery, which will be beneficial regarding their use in scalable communication scenarios (Zhang et al., 2025; Zhu et al., 2025).

Last yet not the least, ethical topics (34%), especially about job displacement, prejudice, and transparency of the algorithm, were brought up all the time. Interestingly, these issues were more relevant in those instances when the voice was perceived as highly human like, which further may indicate that the greater the acoustic realism, the more may be expectations of accountability. This observation relates to the discussion of the idea of algorithmic governance in automated journalism (Jang et al., 2022; Napoli, 2021).

DISCUSSION

Historical Evolution and Technological Progress of AI Anchors

This study examined the historical evolution, current practices, and future prospects of AI anchors in China through the lens of the U&G framework. Early systems like Ananova pioneered the automation of news delivery but faced limitations in expressiveness and naturalness, leaving audience expectations unmet. With the introduction of Xinhua's AI anchors (2018), a more realistic performance was achieved, although critiques about rigid delivery persisted. As one participant reflected, "It looks real, but I still don't feel it is alive" (P3). Over time, systems like Xiaoice v2 (2020) and Gu Xiaoyu (2022) have made strides in multimodal expression, incorporating personalization and even multilingual capabilities. These developments suggest a gradual alignment between technological capabilities and audience expectations for warmth, interactivity, and relatability, rather than a fully realized convergence.

Meeting User Needs for Information, Entertainment, and Interaction

RQ2 delved into how current AI anchors meet user needs for information, entertainment, and interaction. Interviews revealed that AI anchors provide strong functional gratifications, particularly in terms of efficiency and consistency. As one participant stated, "AI anchors don't get tired and can work around the clock" (P4), highlighting the clear functional advantage they offer. This concern for efficiency can, in turn, be placed in a broader context by referencing research on AI use beyond the media sector. Tuan and Tung (2026) demonstrate that the uptake of AI in small and medium-sized enterprises is beneficial not only in terms of operational efficiency gains but also for broader business outcomes, such as better decision-making, reducing costs, and increasing productivity. Accordingly, subjects' recognition of AI anchors' efficiency may relate not only to immediate functional benefits but also to a broader logic of the organization in which AI is seen as a valuable means of cheapening, continuity, and scale in content production.

However, this did not always translate into affective engagement. Another participant commented, "I don't feel any warmth—it's just information" (P12), reflecting a key challenge: while AI anchors are efficient, they often fall short in fostering emotional connections. This distinction resonates with Lim et al. (2023), who describe the divide between "optimizers" who value efficiency and "skeptics" who resist AI news due to a perceived lack of authenticity. Trust emerged as a critical factor: "I can accept AI anchors for weather, but not for political news" (P7). These functional benefits, along with user perceptions, also indicate the more general trends of AI use in industries. The previous studies indicate that the implementation of AI can result in the improvement of organizational efficiency and performance outcomes, especially in the form of automation and cost reduction (Tuan & Tung, 2026). Instead of it being a mediating variable, it seems that trust in this case works as a co-occurring factor with the perceived service quality in influencing user acceptance. The

subjects were more likely to report increased satisfaction with AI anchors that were perceived as credible and stable, and the degree of emotional connection was low in low-trust conditions. This trend is consistent with that of Nga (2026), who singles out perceived service quality and system reliability as some of the factors that correlate with user satisfaction in AI-enabled services. Meanwhile, not all users responded in the same way. According to previous studies, gender and AI self-efficacy are the factors that may affect the way people assess and use AI systems (Olijo, 2025), which might be used to explain the differences between the participants of the current study. This conditional trust aligns with Jang et al. (2022), who found that audience knowledge of automation moderates perceptions of credibility.

This pattern can also be understood as an absence of a causal relationship between trust and affective engagement. Other studies in similar AI service contexts have revealed similar patterns, e.g., Nga (2026) found that trust and service quality are two major factors that influence users' satisfaction in mobile money services. This finding supports the idea that the satisfaction that users obtain from AI systems may always be based on relational factors and confidence-earned (rather than just functionality).

More specifically, participants identified that multimodal design elements - like facial expressions, eye gaze, and clothing - are critical to affect. A number of interviewees highlighted that AI hosts with more realistic facial expressions and lip-syncing were rated as more engaging and less robotic.

Attire was also a critical component. In general, interviewees felt more confident when AI agents wore appropriate professional attire, whereas unrealistic and stylized looks were seen as distracting and less believable（P3 and P6). These findings suggest that the gratifications of affect are not only tied to voice quality, but also to the congruence and realism of visuals. This is consistent with previous work that shows multimodal congruence in AI agents, especially the congruence of facial expressions, gestures and voice, increases social presence and emotional engagement (Nowak & Biocca, 2003).

Among these cues, facial expressions and eye contact were most frequently associated with perceived warmth, while clothing has a greater impact on perceived professionalism than on emotional engagement.

The Role of Acoustic Features in Emotional Engagement

To further illustrate this dynamic, **Table 1** provides insights into the pitch analysis of various AI anchors. Early systems, such as Ananova (2000), exhibited limited pitch variation (mean pitch: 175.24 Hz), which contributed to mechanical intonation and a lack of expressiveness. However, more recent systems, such as Gu Xiaoyu (2022) and AI Avatar (2024), have achieved broader pitch ranges (mean pitch: 286.16 Hz and 105.46 Hz, respectively), which allow for smoother, more natural-sounding voices. For context, human speaking fundamental frequency typically ranges from 100-120 Hz for adult males and 180-220 Hz for adult females (Titze, 1994). Compared to these ranges, the pitch values in **Table 4** for AI anchors partially overlap with human ranges and partially differ. Specifically, the lower pitch range of some AI anchors correspond to typical male speech and the higher values exceed what is commonly said in conversation. Therefore, current systems are able to approximate human pitch distributions but not fully reproduce them, meaning that the lower and higher pitch ranges for some AI anchors partially align with human ranges, but the overall range does not. This lack of fully overlapping ranges suggests that while technological variation has enabled greater expressiveness and variability, the control over pitch for anchors is less context-sensitive than humans, where prosodic variation is more nuanced by communicative need and emotion (Barakat et al. , 2024; Kane et al. , 2024).

This change highlights the growing sophistication in AI speech synthesis, driven by advancements like WaveNet and Tacotron (van den Oord et al., 2016; Wang et al., 2017), which have been shown to support more natural-sounding speech production in prior research. Prior studies suggest that broader pitch variation in AI voices is associated with higher perceived emotional engagement and trust (Xue et al., 2022).

Similarly, **Table 2** shows the improvement in formant distributions (F1, F2, and F3) across different AI anchors. While Ananova (2000) had more limited resonance (mean F1: 668.87 Hz), Xiaoice v2 (2020) and AI Avatar (2024) demonstrated formants more aligned with human speech characteristics, resulting in clearer and more natural vowel articulation. This shift is associated with a more natural-sounding output and may relate to improved user experience (Peterson & Barney, 1952). As one participant noted, "The voice of AI anchors today feels more natural, almost human-like" (P11). This more human-like speech quality appears to

be associated with greater perceived emotional resonance. Previous studies indicate that AI voices closer to human speech characteristics tend to be perceived as more emotionally engaging (Huang & Yu, 2023).

Classic phonetic studies suggest that humans tend to produce vowels with F1 values broadly ranging between ~250 to 750 Hz and F2 values spanning around 850 to 2200 Hz (Hillenbrand et al., 1995; Peterson & Barney, 1952). Given these baseline comparisons, the formant values of modern AI anchors are in a reasonable range for being produced by humans (and thus represent a step in the right direction towards being physiologically realistic).

Yet, there's reason to interpret this with care. Humans don't produce speech with formant ranges that include only static boundaries—the formants themselves vary dynamically with phonetic context. Meanwhile, we've seen AI speech tends to be more accurate in overall values but still lacking in that subtle micro-variability that gives rise to our perceptions of naturalness and expressivity. This goes some way in explaining why participants tell us that AI voices sound "more natural" but still not human.

Despite these technological advancements, the challenge of establishing emotional connection remains. As **Table 3** highlights, intensity levels (mean intensity of Ananova (2000): 67.50 dB; AI Avatar (2024): 60.05 dB) reflect the degree to which AI anchors can modulate their delivery for emotional impact. Although these systems have made strides in volume control, creating a more dynamic and engaging experience, the overall lack of warmth in AI-generated speech remains a barrier to fostering emotional bonds with audiences. As Zhang et al. (2025) suggest, a balanced intensity level is critical in delivering emotionally engaging content, especially for audiences seeking both functional and emotional gratifications from AI-generated media.

Overall, the comparison of acoustic measurement results with human physiological benchmarks indicates that AI speech synthesis has partially achieved acoustic convergence with human speech. However, the remaining gaps between the two, especially in terms of variability and context adjustment, suggest that the perceived naturalness is not only determined by the matching of numerical ranges but also lies in recreating the dynamic characteristics of human communication.

These auditory refinements should not be construed as causal mechanisms; rather, they are better understood as descriptive patterns that appear to resonate with audience-reported experiences. Higher variability in the degree of pitch, more consistent patterns of formants, and even intensity control are all associated with more articulate and natural-sounding speech, which the participants often attributed to higher intelligibility and interest.

In this respect, the acoustic patterns may help contextualize how certain AI anchors are perceived as more effective in information delivery. Meanwhile, an increased degree of perceived presence and emotional involvement appears to be associated with more human-like prosody and a more controlled variation of intensity, especially in case the voice is compatible with the communicative situation (Schreibelmayr & Mara, 2022).

Nevertheless, such associations are always interpretative and one should not take them as a sign of a directional or causal pathway. The findings do not prove how the acoustic properties directly lead to an affective reaction; rather, they indicate a co-occurrence between technical features and user perceptions, which warrants further investigation using more rigorous analytical approaches.

Future Opportunities and Ethical Concerns

RQ3 addressed future opportunities and challenges. Domains such as education and e-commerce emerged as promising areas where functional benefits align closely with user needs. For example, as one participant stated, "An AI anchor in class could repeat explanations patiently" (P22), underscoring the appeal of AI anchors for their ability to provide repetitive, personalized content in educational contexts. Similarly, in live-streaming commerce, cost efficiency and consistent branding were key drivers, with respondents praising AI anchors for their ability to work around the clock, reducing operational costs. These findings echo Zhang et al. (2025), who showed that AI-human collaborative anchors in e-commerce settings increase customer engagement by enhancing playfulness and reducing perceived risk. However, these benefits were tempered by ethical concerns: "If AI anchors keep improving, many human anchors might lose their jobs" (P9). Other participants raised concerns about algorithmic transparency and potential biases, reflecting broader

skepticism regarding the automation of content, as noted in the Reuters Institute's (2024) digital news report 2024.

Beyond these concerns, the findings highlight a broader tension between efficiency and journalistic credibility. While AI anchors significantly improve the speed and scalability of news production, participants frequently questioned the authenticity and reliability of AI-generated content, particularly as automated content generation becomes more widespread. This indicates a mounting concern that the increase in efficiency may come with risks to the security and credibility of information. At the same instant, the interviews indicate that confidence about AI anchors is tightly associated with the perceived social presence and plausibility. When AI-generated voices and appearances are experienced as more human-like, participants tend to report higher levels of engagement and credibility. This pattern is consistent with prior research showing that perceived realism and social presence are key factors shaping trust in AI-mediated communication (Lombard & Ditton, 1997). Taken together, these findings suggest that the adoption of AI anchors involves a context-dependent trade-off: while efficiency is widely accepted in routine or low-stakes content, concerns about authenticity and credibility become more salient in sensitive reporting contexts. In research on automated journalism, alike tensions have been discovered. The increased swiftness can question editorial judgment and public trust (Diakopoulos, 2019).

The Relational Nature of Gratifications and the Role of Anthropomorphism

When combined, the findings further develop the U&G framework in a number of critical aspects. To begin with, they imply that gratifications are not individual but relational: Functional utility, emotional resonance and trust seem to emerge simultaneously and are regarded by the participants as interrelated dimensions of their experience, rather than as independent or sequential processes. This configuration can also be explained based on broader research in AI services, where user satisfaction is associated with the combined perception of efficiency, trust, and service quality (Nga, 2026), rather than being driven by a single dimension. Second, the data show that the anthropomorphic effects (e.g., voice naturalness and gestures) are related to an increase in perceived credibility and engagement, which is also observed in previous studies (Xue et al., 2022). This effect is particularly evident when visual and auditory information are coordinated and integrated. This is because participants generally believe that synchronized facial expressions and appropriate visual presentation will bring about a higher sense of reality and engagement. Third, the research extends gratifications past the consumption of news to education and business, which underscores the cultural uniqueness of AI anchors in China. The study puts the findings in the context of a larger historical path of technological innovation to emphasize the socio-technical character of the communication mediated by AI.

The information provided by the interviews indicates the relationship between the functionality, trust, and affective engagement, and not a formally defined mediated process. Although the efficiency and reliability of AI anchors were continually identified by the participants, the emotional involvement was found to be determined by the perceived credibility. This trend can be generally compared to previous studies that have pointed out the importance of system competence and predictability when forming the user trust (Lee & See, 2004) but cannot be viewed as causal. Overall, rather than operating as independent drivers, functionality, trust, and affective engagement appear to form a loosely coupled configuration, in which improvements in one dimension do not necessarily correspond to changes in others.

Practical Implications and Future Directions

Practically, the findings suggest that AI anchors function effectively as complements to human anchors in specific contexts, particularly in repetitive, multilingual, and educational scenarios, while human anchors remain important in emotionally sensitive or high-stakes communication settings. This complementary role is especially evident in contexts requiring efficiency versus those requiring emotional depth (Song, 2025). Designers should calibrate anthropomorphism carefully to strengthen trust without causing discomfort, while policymakers must ensure transparency and labor protections to safeguard the long-term legitimacy of AI anchors (OECD, 2023; Partnership on AI, 2025).

In conclusion, this study demonstrates that the acceptance of AI anchors in China is shaped by the interplay of technological progress, user gratifications, and ethical concerns. The findings highlight both the potential and limitations of AI anchors, offering theoretical refinements to U&G and practical insights for

integrating AI into evolving media landscapes. As technological advancements in speech synthesis continue, it will be crucial to address the challenge of emotional connection to fully realize the potential of AI anchors in both practical and cultural contexts.

CONCLUSION

This study investigated the historical evolution, current applications, and future development of AI anchors in China using the U&G framework. By combining technological analysis with qualitative interview data, the research sheds light on how technological progress appears to interact with social demand and how audiences interpret and evaluate AI anchors in practice.

Summary of Key Findings

The findings reveal three important patterns. First, the evolution of AI anchors has generally paralleled social expectations: from the rigid intonation of Ananova, through Xinhua's early deep learning-based anchors, to the more interactive and multimodal designs of Xiaoice and Gu Xiaoyu. Second, while audiences valued functional gratifications such as efficiency and 24/7 availability, these strengths did not automatically produce trust or emotional engagement. Instead, acceptance was conditional, often mediated by anthropomorphic features such as voice and gestures. Third, participants highlighted new opportunities in education, e-commerce, and multilingual contexts, but expressed persistent concerns about bias, transparency, and job displacement (Niu, 2025; Schwartz & Schwartz, 2022).

Contributions of the Study

1. **Theoretical contribution:** It extends U&G by suggesting that gratifications are relational rather than isolated: functional, affective, and trust-based needs appear to interact in shaping cultural acceptance. This adds nuance to existing models that treat gratifications as discrete categories (Sundar & Limperos, 2013).
2. **Empirical contribution:** It provides one of the first qualitative accounts of Chinese audiences' perceptions of AI anchors, complementing recent survey-based studies on trust and continuance intention (e.g., Huang & Yu, 2023; Jang et al., 2022).
3. **Practical contribution:** It suggests that AI anchors are best positioned as complements rather than substitutes. For education, they can provide personalized repetition; in commerce, they can strengthen brand engagement; and in news, they are most suitable for low-stakes or routine reporting. Policymakers and developers should work together to ensure transparency, safeguard labor, and calibrate anthropomorphism to balance trust and professionalism.

Limitations and Future Research Directions

Although this study provides detailed insights into the evolution and reception of AI anchors in China, several limitations should be acknowledged. First, the research was deliberately confined to Chinese audiences to provide in-depth, context-sensitive analysis; as a result, the findings are not automatically generalizable to other cultural or regulatory environments. Future comparative research across different national and regulatory settings could test whether the relational patterns identified here—especially the observed association between trust and the relationship between functional and affective gratifications—hold elsewhere (Lim et al., 2023).

Second, this study relied primarily on interviews supplemented by acoustic analysis. While this mixed evidence clarified how technological refinements (e.g., improvements in pitch, formant rendering, and intensity control) are associated with audience perceptions, it did not include objective behavioral or physiological measures in situ. Recent psychophysiological research suggests that brain and attentional responses to AI versus human voices can diverge from self-reports, implying that combining interview data with experimental or psychophysiological methods would strengthen causal inference about affective engagement.

Third, this study focused on the audience perspective and did not incorporate developer, newsroom, or policymaker viewpoints. Understanding the political, organizational, and commercial constraints on AI anchor

design—and how these shape transparency, accountability, and labor outcomes—requires multi-stakeholder inquiry. The broader governance literature argues that platform governance and public-interest considerations are central to legitimizing automated media practices; future work would benefit from integrating policy and governance scholarship into empirical designs (Napoli, 2019; Wiederhold, 2019).

In short, although this research intentionally concentrated on Chinese users to generate fine-grained insights, the next stage for the field is clear: comparative, multi-method, and multi-stakeholder studies that can test generalizability, further examine potential mechanisms, and inform practicable governance responses.

Final Reflection

The case of AI anchors illustrates the broader challenge of embedding AI in communication. Their acceptance depends not only on technical sophistication but also on their ability to resonate with human needs, values, and ethics. In political or emotionally charged contexts, human anchors remain indispensable. Yet in education, commerce, and routine news, AI anchors can complement human labor and extend media accessibility. Ultimately, the success of AI anchors will rest on whether they are designed and governed as tools that enhance, rather than replace, human communicators (Niu, 2025; Sharples, 2023).

In sum, the findings of this study respond directly to the gaps identified in the literature. By linking measurable acoustic features—such as pitch variation, formant range, and intensity—to users' evaluations of clarity and authenticity, the analysis suggests a relationship between technical accounts of speech synthesis to demonstrate their communicative significance. Moreover, the interviews highlight how functional gratifications, including efficiency and accessibility, are not ends in themselves but appear to depend on trust as an important condition shaping emotional engagement. Finally, the focus on Chinese audiences contributes a culturally grounded perspective, offering insight into a context that has been central to the deployment of AI anchors yet underrepresented in qualitative research. Together, these contributions extend the U&G framework to AI-mediated news delivery and provide an empirical basis for understanding the interrelated dynamics of functionality, trust, and affective engagement.

Author contributions: Niu Yongbo contributed to Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, and Writing – review & editing. Sharifah Sofiah Syed Zainudin and Syafila Binti Kamarudin contributed to Supervision, Project administration, Resources. All authors approved the final version of the article. All authors approved the final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Ethics declaration: The authors stated that no ethical approval was required for this study.

AI statement: This study utilized generative AI and AI-based tools to assist with grammar revision, English translation, and literature searching during the manuscript preparation process. All authors have reviewed and approved the final content, and take full responsibility for the accuracy and integrity of the work.

Declaration of interest: The authors declared no competing interests.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct. *APA*. https://www.apa.org/ethics/code

Barakat, H., Turk, O., & Demiroglu, C. (2024). Deep learning-based expressive speech synthesis: A systematic review of approaches, challenges, and resources. *EURASIP Journal on Audio, Speech, and Music Processing, 2024*(1), Article 29. https://doi.org/10.1186/s13636-024-00329-7

Boersma, P., & Weenink, D. (2023). Praat: Doing phonetics by computer. *Praat*. http://www.praat.org/

Bohacek, M., & Farid, H. (2024). Human action CLIPs: Detecting AI-generated human motion. *arXiv, Article 2412.00526.* https://doi.org/10.48550/arXiv.2412.00526

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

British Educational Research Association (2018). *Ethical guidelines for educational research* (4th ed.). BERA. https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018

Carlson, M. (2023). *Automating the news: How algorithms are rewriting the media*. Columbia University Press.

Couldry, N., & Mejias, U. A. (2019). Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media, 20*(4), 336-349. https://doi.org/10.1177/1527476418796632

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer. https://doi.org/10.1007/978-1-4899-2271-7

Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE handbook of qualitative research* (4th ed.). SAGE.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press. https://doi.org/10.4159/9780674239302

Feng, Z., & Shu, Y. (2024). AI anchor vs. real anchor: An experimental study based on audience perception. *China News Review, 5*(3), 16-25. https://doi.org/10.35534/cnr.0503002

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. *Health Services Research, 48*(6), 2134-2156. https://doi.org/10.1111/1475-6773.12117

Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., & Stapp, J. (2022). AI in business communication: The changing landscape of research and teaching. *Business and Professional Communication Quarterly, 85*(1), 7-33. https://doi.org/10.1177/23294906221074311

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59-82. https://doi.org/10.1177/1525822X05279903

Hennink, M. M., Kaiser, B. N., & Weber, M. (2019). What influences saturation? Estimating sample sizes in focus group research. *Qualitative Health Research, 29*(1), 148-161. https://doi.org/10.1177/1049732318821692

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5), 3099-3111. https://doi.org/10.1121/1.411872

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports, 13*, Article 5487. https://doi.org/10.1038/s41598-023-30938-9

Huang, Y., & Yu, Z. (2023). Understanding the continuance intention for artificial intelligence news anchor: Based on the expectation confirmation theory. *Systems, 11*(9), Article 438. https://doi.org/10.3390/systems11090438

Jang, W., Kwak, D. H., & Bucy, E. P. (2022). Knowledge of automated journalism moderates evaluations of algorithmically generated news. *New Media & Society, 24*(12), 2852-2870. https://doi.org/10.1177/14614448221142534

Kane, J., Johnstone, M. N., & Szewczyk, P. (2024). Voice synthesis improvement by machine learning of natural prosody. *Sensors, 24*(5), Article 1624. https://doi.org/10.3390/s24051624

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. *Public Opinion Quarterly, 37*(4), 509-523. https://doi.org/10.1086/268109

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Lim, J. S., Shin, D., Zhang, J., Masiclat, S., Luttrell, R., & Kinsey, D. (2023). News audiences in the age of artificial intelligence: Perceptions and behaviors of optimizers, mainstreamers, and skeptics. *Journal of Broadcasting & Electronic Media, 67*(3), 353-375. https://doi.org/10.1080/08838151.2022.2162901

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE. https://us2.sagepub.com/en-us/nam/book/naturalistic-inquiry

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2), Article JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lyu, X., Ramasamy, S. S., & Ying, F. (2023). The role of AI digital anchors in enhancing the news broadcasting user experience: An analysis of the interaction of AI anchors with the audience in live news programs. *EAI Endorsed Transactions on Ambient Systems, 23*(11), 2343240. https://doi.org/10.4108/eai.23-11-2023.2343240

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734. https://doi.org/10.2307/258792

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce. *Information Systems Research, 13*(3), 334-359. https://doi.org/10.1287/isre.13.3.334.81

Napoli, P. M. (2019). *Social media and the public interest: Media regulation in the disinformation age*. Columbia University Press. https://doi.org/10.7312/napo18454

Nass, C., & Brave, S. (2006). *Wired for speech: How voice activates and advances the human-computer relationship*. *Computational Linguistics, 32*(3), 451–452. https://doi.org/10.1162/coli.2006.32.3.451

Nga, L. P. (2026). Digital transformation, artificial intelligence application, and service quality as determinants of user satisfaction in mobile money services. *Ianna Journal of Interdisciplinary Studies, 8*(1), 1032-1045. https://iannajournalofinterdisciplinarystudies.com/index.php/1/article/view/1425

Niu, Y. (2025). From human anchors to AI anchors: A review of technology, ethics, and audience response in audiovisual media transformation. *International Theory and Practice in Humanities and Social Sciences, 2*(2), 361-372. https://doi.org/10.70693/itphss.v2i2.166

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 12*(5), 481-494. https://doi.org/10.1162/105474603322761289

OECD. (2023). *Ensuring trustworthy artificial intelligence in the workplace*. OECD Publishing. https://doi.org/10.1787/08785bba-en

Olijo, I. I. (2025). Gender disparities in research return rates: The moderating influence of AI self-efficacy and methodological design. *Verlumun Journal of AI, Gender and Cultural Studies, 1*(1), 90-100. https://doi.org/10.5281/zenodo.18242833

Partnership on AI. (2025). Improving labor transparency in AI through worker inclusion. *Partnership on AI*. https://partnershiponai.org/improving-labor-transparency-in-ai-through-worker-inclusion/

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184. https://doi.org/10.1121/1.1906875

Reuters Institute. (2024). Digital news report 2024. *Reuters Institute*. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024

Rubin, A. M. (1983). Television uses and gratifications: The interactions of viewing patterns and motivations. *Journal of Broadcasting, 27*(1), 37-51. https://doi.org/10.1080/08838158309386471

Rubin, A. M. (2002). The uses-and-gratifications perspective of media effects. In J. Bryant, & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 525-548). Routledge.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1-2), 227-256. https://doi.org/10.1016/S0167-6393(02)00084-5

Schreibelmayr, S., & Mara, M. (2022). Robot voices in daily life: Vocal human-likeness and application context as determinants of user acceptance. *Frontiers in Psychology, 13,* Article 787499. https://doi.org/10.3389/fpsyg.2022.787499

Schwartz, R., & Schwartz, R. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1270

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice, 9*(2), 159-167. https://doi.org/10.1080/23735082.2023.2261131

Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Wu, Y. (2018). Natural TTS synthesis by conditioning Wavenet on MEL spectrogram predictions. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 4779-4783, https://doi.org/10.1109/ICASSP.2018.8461368.

Song, P. (2025). Integrating AI virtual humans into broadcasting and hosting education: Pedagogical, ethical, and industry implications. *UKR Journal of Arts, Humanities and Social Sciences, 1*(6), 300-307. https://doi.org/10.5281/zenodo.16881288

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction. *Journal of Computer-Mediated Communication, 25*(1), 74-88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., & Limperos, A. M. (2013). Uses and grats 2.0: New gratifications for new media. *Journal of Broadcasting & Electronic Media, 57*(4), 504-525. https://doi.org/10.1080/08838151.2013.845827

Tan, L., Zheng, H., & Liu, X. (2021). Advances in emotional speech synthesis for AI anchors. *Journal of Speech Technology, 19*(4), 512-523.

Titze, I. R. (1994). *Principles of voice production*. Prentice Hall.

Tuan , H. C., & Tung, N. T. (2026). Determinants of AI adoption and how the adoption affects the business performance of small and medium-scale enterprises. *Ianna Journal of Interdisciplinary Studies, 8*(1), 25-37. https://doi.org/10.5281/zenodo.17689249

Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A, & Kavukcuoglu, K (2016). WaveNet: A generative model for raw audio. *arXiv.* https://doi.org/10.48550/arXiv.1609.03499

Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: Systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology, 18*, Article 148. https://doi.org/10.1186/s12874-018-0594-7

Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., Le, Q., Agiomyrgiannakis, Y., Clark, R., & Saurous, R. A. (2017). Tacotron: Towards end-to-end speech synthesis. In *Proceedings of the INTERSPEECH 2017* (pp. 4006-4010). https://doi.org/10.21437/Interspeech.2017-1452

Wang, S., Bao, T., & Huang, Y. (2025). A study on the factors influencing audience choices of AI anchors in the era of digital intelligence. In *Proceedings of the 2025 2nd International Conference on Innovation Management and Information System* (pp. 439-444). ACM. https://doi.org/10.1145/3745676.3745743

Wiederhold, B. K. (2019). Animated news anchors: Where to next? *Cyberpsychology, Behavior, and Social Networking, 22*(11), 675-676. https://doi.org/10.1089/cyber.2019.29167.bkw

Wölker, A., & Powell, T. E. (2021). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism, 22*(1), 86-103. https://doi.org/10.1177/1464884918757072

Xie, Y., Zhu, K., Zhou, P., & Liang, L. (2023) How does anthropomorphism improve human-AI interaction satisfaction: A dual-path model. *Computers in Human Behavior. 148*, Article 107878. https://doi.org/10.1016/j.chb.2023.107878

Xinhua News Agency, & Sogou. (2018). Xinhua-Sogou AI news anchor. https://news.cgtn.com/news/3d3d514d3055444e30457a6333566d54/index.html

Xue, K., Li, Y., & Jin, H. (2022). What do you think of AI? Research on the influence of AI news anchor image on watching intention. *Behavioral Sciences, 12*(11), Article 465. https://doi.org/10.3390/bs12110465

Zhang, Y., Wang, X., & Zhao, X. (2025). Supervising or assisting? The influence of virtual anchors driven by AI-human collaboration on customer engagement in live-streaming e-commerce. *Electronic Commerce Research, 25*(4), 3047-3070. https://doi.org/10.1007/s10660-023-09783-5

Zhang,Y. (2024). Analysis of the application of artificial intelligence in information courses in primary and secondary schools. *Lecture Notes in Education Psychology and Public Media, 59*, 241-250. https://doi.org/10.54254/2753-7048/59/20241783

Zhu, Y.-P., Xin, L., Wang, H., & Park, H.-W. (2025). Effects of AI virtual anchors on brand image and loyalty: Insights from perceived value theory and SEM-ANN analysis. *Systems, 13*(2), Article 79. https://doi.org/10.3390/systems13020079