JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

J G T S S

Research article

https://doi.org/10.70731/8zydzz10

Exploring Key Dimensions of AI-powered Digital Human Live Streaming: A Qualitative Study Based on In-Depth Interviews with Multiple Stakeholders

Dan Hua[1,*], **Vincent Wee Eng Kima**[1]

Received 2 February 2026

Accepted 24 March 2026

Published 31 March 2026

Abstract: AI-powered digital human live streaming is emerging as a significant marketing format in the e-commerce sector. however, academic understanding of its core characteristics remains confined to conceptual transposition and theoretical deduction, lacking empirical evidence of actual perceptions from consumers and multiple stakeholders. Through in-depth interviews with 19 multiple stakeholders from China (Generation Z consumers, e-commerce practitioners, and academics), this study employed thematic analysis with manual coding throughout the process to systematically identify and define three core dimensions of AI-powered digital human live streaming: anthropomorphism, intelligent interactivity and personalized recommendation, while further revealing the multi-layered internal structures of each dimension. The findings reveal that anthropomorphism comprises three levels: visual anthropomorphism, Behavioral anthropomorphism and emotional anthropomorphism; intelligent interactivity encompasses three elements: response immediacy, response relevance and conversational coherence; and personalized recommendation consists of three dimensions: content relevance, perceived personal attention and interaction customizability. This study provides an empirically grounded, user-language-based feature dimension framework for the field of AI-powered digital human live streaming, bridging the conceptual gap between macro-theoretical concepts and micro-level user perceptions, and laying a solid conceptual foundation for subsequent scale development and quantitative model testing.

Keywords: AI-powered digital human live streaming; Feature dimensions; Anthropomorphism; Intelligent interactivity; Personalized recommendation

ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Introduction

AI-powered digital human live streaming is reshaping China's e-commerce ecosystem at a remarkable pace. According to the *Report on the High-Quality Development of Livestream E-commerce* (China International E-commerce Center Research Institute, 2025), as a technological form integrating computer vision, large language models and multimodal interaction, AI-powered digital human live streaming has been widely deployed on platforms such as Taobao, JD.com and Douyin, demonstrating significant advantages such as 24/7 service, controllable costs and consistency in information delivery. However, in stark contrast to the fervor of industry practice, academic research on the theoretical framework of AI-powered digital human live streaming has lagged significantly. A fundamental question remains unresolved: how do consumers perceive and describe AI-powered digital human live streaming? What are the key dimensions that constitute this unique experience?

This theoretical ambiguity severely hampers knowledge accumulation in this field. Existing research on the operationalization of AI-powered digital human live streaming characteristics is fragmented. Some studies directly borrow the concepts of perceived ease of use and perceived usefulness from traditional technology acceptance models, reducing AI anchors to mere technological tools (Davis, 1989; Hwang & Lim, 2025). Others transplant interaction and professionalism scales designed for human anchors, overlooking the funda-

[1]Universiti Tun Abdul Razak (UNIRAZAK), Kuala Lumpur 50400, Malaysia.
*Corresponding author. Email: hu.dan@ur2.unirazak.edu.my

mental differences in the interactive agency of AI anchors (McMillan & Hwang, 2002; Zhang et al., 2024). Yet others measure anthropomorphism as a single-dimensional construct without further differentiation, failing to capture the multi-layered connotations of appearance, behavior and emotion (Waytz et al., 2014; Kühne & Peter, 2023). This top-down theoretical approach results in a significant conceptual gap between measurement tools and consumers' actual experience. Current research often treats characteristics such as personification, intelligence and interactivity as isolated variables in a fragmented manner, and has yet to establish a classification system that is both consistent and theoretically grounded. Therefore, this study aimed to return to the authentic perceptions of consumers and multiple stakeholders to systematically identify and define the core characteristic dimensions of AI-powered digital human live streaming.

Given the fragmentation and conceptual gap in existing research on theoretical conceptualization, this study adopts an exploratory qualitative research design. Through in-depth interviews with 19 multiple stakeholders (Generation Z consumers, e-commerce practitioners and academics), it employs thematic analysis to systematically identify the core characteristic dimensions of AI-powered digital human live streaming and their underlying structure. This study aims to construct an integrative conceptual framework grounded in user language, reveal the multi-layered composition of consumer experience, and provide an empirical foundation for subsequent scale development and corporate practice.

Literature Review

Conceptual and contextual characteristics

The technological evolution of AI-powered digital human live streaming has undergone a hierarchical transition from virtual humans to digital humans and finally to AI-powered digital humans. Virtual humans generally refer to digital avatars based on computer-generated technology that possesses human appearance or behavioral characteristics, emphasizing virtual existence but lacking intelligent interactive capabilities (Griffor et al., 2017). Digital humans achieve high realism through high-precision 3D modeling and motion capture, but their Behavioral logic remains dependent on pre-set scripts (Hetherington & McRae, 2017). AI-powered digital humans, on the other hand, integrate natural language processing, emotion recognition and autonomous decision-making capabilities, enabling real-time, personalized interaction with users (Machidon et al., 2018; Xu et al., 2026). This technological evolution has transformed AI-powered digital humans from objects to be viewed as entities capable of dialogue.

In the context of livestream e-commerce, AI-powered digital humans exhibit three distinctive characteristics. First, they offer extended operational hours, enabling 24/7 uninterrupted service. Second, they demonstrate intelligent interaction, engaging in open-ended dialogue and context-aware responses based on large language models. Third, they provide personalized content, dynamically adjusting recommendation strategies through real-time data analysis. These characteristics create cognitive tension for consumers, who are aware that the entity is not human yet perceive it as human-like; that is, while they clearly recognize the algorithmic nature of the interactive entity, they also perceive cues of social interaction within its anthropomorphic presentation. This experience falls precisely within the specific range described by Mori's (1970) 'Uncanny Valley' theory: When an AI anchor closely resembles a real person yet reveals non-human characteristics, it may trigger cognitive dissonance and emotional discomfort; however, when its emotional expression capabilities exceed a specific threshold, it can establish a deep connection that transcends mere appearance (Mori et al., 2012). This unique experience is the focal point of this study.

Theoretical foundations: Social presence and technological trust

AI-powered digital human live streaming blurs the boundary between humans and machines, with consumers simultaneously perceiving the algorithmic nature of the experience while experiencing social interaction cues. To systematically capture this contradictory experience, this study employs social presence theory and technological trust theory as sensitizing concepts (Blumer, 1954) to guide data interpretation. This approach aligns with the methodological tradition of grounded theory, which maintains theoretical openness whilst utilizing existing concepts to enhance sensitivity to empirical phenomena.

The theory of social presence describes the degree to which individuals perceive the actual presence of others during mediated interaction (Short et al., 1976). In the context of AI-powered digital human live streaming, this theory helps identify how consumers perceive the social presence of digital avatars through the screen, including their perception of immediate responsiveness, expectations of emotional resonance, and the cognitive tension of 'knowing they are not human yet perceiving them as human-like'. This concept provides an initial analytical framework for exploring immersive experiences and the sense of being valued.

The theory of technological trust distinguishes three dimensions: competence trust (confidence in algorithmic efficacy), benevolence trust (skepticism regarding platform intentions), and honesty trust (judgement of information authenticity) (McKnight et al., 2002). Given that the recommendation behavior of AI anchors simultaneously exhibits intelligent professionalism and algorithmic uncertainty, this theory provides a conceptual framework for understanding the ambivalent psychological state of consumers when faced with algorithmic decisions (acknowledging their efficiency while questioning their motives).

Together, these theories constitute the conceptual framework for analyzing interview data. It is important to emphasize that this study does not presuppose causal pathways or dimensional structures for these theoretical variables, but rather treats them as perspectives for identifying emic constructs. During the data analysis process, the research maintained theoretical openness to emergent dimensions (such as potential algorithmic aversion or digital ethical anxieties), allowing user language to reconfigure or transcend existing theoretical frameworks.

Research gaps and theoretical opportunities

Existing research has three significant limitations. First, the research subjects exhibited fragmented characteristics. Existing literature dissects the appearance, interaction and recommendation functions of AI anchors into isolated variables for separate examination, lacking systematic attention to them as vehicles for an integrated experience; characteristics such as anthropomorphism and intelligence are treated in a piecemeal manner, and a consistent classification system has yet to be established. Second, the underlying mechanisms are insufficiently explained. Most existing models adopt a simplified 'features → psychological variables → behavior' pathway, which disconnects the complex psychological evolution of consumers and assumes that all consumers follow a homogeneous path, thereby overlooking the differences between groups. Third, research methods are relatively limited. Existing literature relies excessively on structured questionnaires; in an emerging field where 'it is not yet known what to measure', there is a conceptual gap between pre-defined scales and consumers' actual experiences, and qualitative exploration is severely lacking.

These limitations constitute a threefold theoretical gap in the field, creating opportunities for knowledge innovation. First, constructing an integrated conceptual framework for AI-powered digital human live streaming through qualitative induction can overcome the theoretical bottlenecks of the current fragmented research paradigm, establishing analytical dimensions that encompass both social presence and technological trust. Second, by deeply describing the heterogeneous evolutionary pathways of consumer psychological experiences, a process theory distinct from linear models can be developed to explain how different groups negotiate human-machine relationships in diverse ways. Finally, a grounded conceptual system built upon users' native language will provide an embodied theoretical foundation for subsequent scale development, bridging the theoretical gap between operationalized measurements and real-world experiences. Based on this, this study uses in-depth interviews to identify core perceptual dimensions, describing the complexity and group differences in psychological experiences, thereby providing a theoretical framework grounded in users' language for the field of AI anchors consumer behavior.

Research Design

This study adheres to the interpretivist research paradigm (Creswell & Poth, 2018) and adopts an exploratory qualitative research design. Data were collected through semi-structured in-depth interviews, and thematic analysis was employed for manual coding throughout the process (Braun & Clarke, 2022).

Data collection: Semi-structured in-depth interviews

Design of the interview guide

The interview guide was structured around four core modules: experience entry and situational recall; perception of characteristics and concrete descriptions; psychological mechanisms and behavioral decision-making; and boundary exploration and future prospects. First, situational priming was used to guide participants in recalling specific AI live-streaming experiences. Second, feature descriptions invited participants to characterize the performance of AI digital human anchors using everyday language. Third, mechanism exploration delved into psychological responses such as trust and a sense of presence. Finally, boundary exploration revealed unmet underlying needs.

This study was adapted to suit different respondent groups. For Generation Z consumers, everyday language was used, such as "What was your first visual and auditory impression of this AI anchor? In what ways did it feel 'like a real person' to you, and in what ways did it still feel 'like a machine' ?" For e-commerce practitioners, the focus was on commercial logic, technical implementation and industry insights, such as "What are the primary technical or cost bottlenecks currently limiting the ability of AI-powered digital human to achieve natural, in-depth interactions?" Academics focused on theoretical critique, such as "When applying traditional anthropomorphism theories to AI-powered digital human with interactive capabilities, do their conceptual dimensions need to be re-examined and expanded?"

Sampling strategy and sample characteristics

This study employed purposive sampling combined with the maximum diversity sampling method (Patton, 2015), aiming to achieve the widest possible conceptual coverage of the research phenomenon within a small sample by selecting cases that exhibited diversity across specific dimensions. Based on this strategy, this study established three key stakeholder groups as the sampling framework (**Table 1**).

The final sample comprised 19 participants. The sample size adhered to the principle of theoretical saturation (Saunders et al., 2018). No substantive new themes emerged following the coding of data from the first 16 respondents. Three subsequent interviews confirmed the stability of the thematic framework, leading to the conclusion that theoretical saturation had been achieved.

Data collection procedure

Data collection took place between September and December 2025, employing a combination of in-person and online video interviews, with each session lasting 40–60 minutes. All interviews were audio-recorded with the participants' informed consent, transcribed verbatim within 24 hours, and retained colloquial features and emotional nonverbal cues (e.g., [loud laughter], [pause]).

Ethical considerations

This study strictly adhered to the following academic ethical standards. First, regarding informed consent, an "Informed Consent Form" was sent to participants prior to the interview, detailing the research objectives and the methods of data anonymization. Second, regarding privacy protection, all identifiable information was thoroughly anonymized during the transcription stage (coding system: C01–C10, P01–

Table 1 | Sample Composition of Respondents

Group	Sampling Criteria	Sample Size	Key Contribution
Generation Z consumers	Aged 18–25, weekly active users on mainstream e-commerce platforms, with at least one distinct and memorable experience of watching an AI-powered digital human livestreaming within the past three months	10	Providing authenticity of experience: the ultimate source of construct validity
E-commerce Practitioners	Over 5 years of industry experience, working at leading livestream e-commerce platforms, well-known brands, or digital human technology firms, with direct involvement in decision-making or operations of AI-powered digital human live streaming projects	5	Ensuring practical applicability: guaranteeing the feasibility of the dimensions
Academics	Associate Professor or Professor, with research focusing on digital media, consumer behaviour or human-computer interaction	4	Providing theoretical rigour: ensuring alignment with academic literature

P05, A01–A04). Third, regarding data security, all raw files were stored in an encrypted form, and access was strictly restricted. Finally, regarding voluntary participation, interviewees were reminded throughout the study that they retained the right to withdraw from the interview at any time.

Data analysis: Thematic analysis

Integrative analysis strategy

This study adopted a cross-group integrative analysis strategy, merging data from the three groups of respondents for unified coding. This strategy is based on three theoretical considerations. First, the overarching nature of the research objectives requires a focus on constructing a system of characteristic dimensions rather than comparing group differences; the value of the three groups lies in their mutual complementarity and cross-validation (Chen, 2000). Second, the diverse pathways to theoretical saturation indicate that integrating the perspectives of consumers, practitioners and academics enables broader theoretical boundaries to be explored within a shorter timeframe (Wang, 2007). Finally, the prior assurance of construct validity ensured that the extracted dimensions possessed the triple attributes of experiential perceptibility, technical feasibility and theoretical dialogability.

Throughout the coding process, the researchers maintained sensitivity to the perspectives of different groups and conducted preliminary comparative analyses. The results revealed that, although the descriptions provided by the three groups differed in wording, they were highly consistent at the level of abstract constructs, pointing to three core themes: anthropomorphism, intelligent interactivity, and personalized recommendation. This provides robust cross-group evidence for the dimensional framework. By integrating these three groups, this study achieved qualitative assurance of construct validity: consumers provide experiential authenticity, practitioners ensure practical operability, and scholars contribute theoretical logic, thereby establishing a reproducible, multi-stakeholder triangulation paradigm for qualitative research in fields characterized by rapid technological iteration.

Thematic analysis procedure

This study employed the six-step thematic analysis method proposed by Braun and Clarke (2022) for manual coding throughout the process, emphasizing deep immersion in the data and iterative analysis.

Stage 1: Immersion in the data and preliminary interpretation. The researcher conducted active, repeated and critical readings of the interview transcripts, recording analysis notes to capture initial patterns and underlying concepts.

Stage 2: Generating open codes. A line-by-line analysis was employed, prioritizing the use of respondents' native concepts as code labels. For example, regarding appearance and emotional aspects: "dimples when smiling" (C03), "slight fine lines at the corners of the eyes" (C04), "micro-expressions can't keep up, so users still find it a bit artificial" (P03), "60 marks for appearance is enough, but 80 marks for emotion is essential" (A02); interaction quality-related comments such as: "a three-second reply but the information is incomplete" (C01), "picks up on the joke and stays on topic through multiple rounds" (C07), and "conversation retention is more important than responding to the first sentence" (P04); and at the recommendation service level: "I'd just mentioned wanting a lipstick shade to make my skin look fairer, and she immediately recommended that Rotten Tomatoes shade–it was exactly what I wanted" (C02), "It felt like it was prepared just for me" (C08), "Real-time recommendations based on behavior within the livestream" (P01). A total of 127 open codes were generated at this stage, covering the diverse perspectives of the three groups.

Stage 3: Formulation of focused codes. Logical connections were identified through repeated comparison and clustering of open codes. For example, physical details such as dimples and fine lines, together with vocal tone, point to the personification of appearance; the timing of nods and the rhythm of pauses point to the personification of behavior; and "feeling"-related vocabulary and expressions of empathy point to the personification of emotion. Similarly, codes relating to interactive efficacy (*instant replies*, *picking up on jokes*, *staying on topic*) were clustered into responsiveness, relevance of responses, and conversational coherence. Codes relating to recommendation services (*tailored preparation*, *real-time adjustment*) were clustered into content relevance, sense of personal attention, and interactive customization. This stage yielded nine candidate sub-dimensions, falling under the three core themes of anthropomorphism, intelligent interactivity, and personalized recommendation. Taking the anthro-

Table 2 | Example of thematic analysis coding evolution: Taking anthropomorphism as an example

Coding Level	Coding Type	Representative Content	Source Examples
Open Coding (Initial code)	Local Concept Labels	"She has dimples when she smiles, which I find very soothing"	C03
		"There are tiny fine lines at the corners of her eyes—the kind of natural creases you see when a real person smiles"	C04
		"Large language models can make digital human say 'I understand how you feel', but the micro-expressions don't keep up, so users still find it a bit artificial"	P03
		"A score of 60 for appearance is enough; emotions must score 80."	A02
		"Pause briefly before answering; it gives the impression they're thinking"	C05
		"The timing of the nods and pauses makes it feel like they're really listening to me, rather than following a pre-written script"	C06
		(Other codes relating to appearance, behaviour and emotional details omitted)	
Focus coding (Candidate themes)	Mid-level concept aggregation	Visual Anthropomorphism: dimples, wrinkles, tone of voice and other physical details	C03, CO4
		Behavioral Anthropomorphism: Timing of nods, rhythm of pauses, naturalness of body language	C05, C06
		Emotional anthropomorphism: "feeling" vocabulary, expressions of empathy, prioritisation of emotional capabilities	C03, P03, A02
Theoretical coding	Refining the Concept	Anthropomorphism: The extent to which AI-powered digital humans simulate human external characteristics and emotional expressions in terms of vision, speech, facial expressions and movements, comprising three levels: visual anthropomorphism, Behavioral anthropomorphism and emotional anthropomorphism.	Integration of all the above focused codings (consensus among consumers, practitioners and academics)

pomorphism dimension as an example, its coding evolution is presented in Table 2.

Stage 4: Review and refinement of themes. Based on the candidate themes identified above, this phase involved a systematic review of the nine candidate sub-dimensions. The internal consistency of the themes and the representativeness of the overall dataset were assessed; sub-dimensions with overlapping semantics were merged, such as combining *favorable physical appearance* and *human-like voice* into visual anthropomorphism. Candidate themes with low consensus were excluded, such as *entertainment value* and *curiosity*, which were provisionally classified as marginal themes due to their low frequency of occurrence and weak cross-group consensus. After several iterations, the independence and stability of the nine sub-dimensions were confirmed, with clear boundaries and distinct characteristics for each dimension.

Stage 5: Defining and naming themes. The connotations of the core themes and their respective sub-dimensions were precisely defined. Anthropomorphism comprises three levels: visual anthropomorphism (the degree of human resemblance in visual appearance and voice), Behavioral anthropomorphism (the naturalness of body language and interaction rhythm), and emotional anthropomorphism (emotional recognition and the ability to express empathy). Intelligent interactivity encompasses three elements: response immediacy (reaction speed), response relevance (content matching), and conversational coherence (ability to maintain context). Personalized recommendation consist of three levels: content relevance (accuracy of information matching), sense of personal attention (experience of being valued and understood), and interactivity customization (flexibility in real-time adjustment of recommendation strategies). The boundaries between each dimension are clear; while distinct from one another, they collectively form an integrated experiential framework.

Stage 6: Report writing. The findings from the preceding analyses were systematically integrated to construct a comprehensive narrative around each core theme, presenting empirical evidence and theoretical implications in detail to form the main body of Part 4 (Research Findings) of this study.

Through the analysis of the above six stages, this study distilled three core themes and nine sub-dimensions from 127 open-coded entries, systematically presenting the structural dimensions of AI-powered digital human live streaming.

Reflective strategies

This study ensured the rigour of the analysis through the following strategies. First, regarding reflective documentation, the researcher maintained analytical memos throughout the data collection and analysis process, meticulously recording theoretical assumptions, emotional responses, and conceptual uncertainties, thereby mitigating researcher bias through continuous self-reflection. Second, through counter-example analysis, the researcher proactively examined data fragments that contradicted the dominant themes (such as consumers' positive evaluations of the non-anthropomorphic characteristics of AI anchors), ensuring that contradictory evidence was not selectively ignored. Third, a comprehensive audit trail was established, systematically retaining raw transcripts, coding tables, analysis memos, iterative versions of themes, and decision logs to ensure the research process was auditable and reproducible. Finally, three participants were invited to review the preliminary findings through peer review to confirm that the extracted themes closely aligned with their original intentions.

Research Findings

Through systematic thematic analysis, this study distilled three core feature dimensions with a high degree of cross-group consensus from in-depth interview data with multiple stakeholders: anthropomorphism, intelligent interactivity, and personalized recommendation. The connotations and multi-layered internal structures of each dimension are elaborated below.

Anthropomorphism: A multi-layered evolution from physical resemblance to emotional resonance

In this study, anthropomorphism is defined as the degree to which AI-powered digital humans simulate human external characteristics and emotional expressions in terms of visual appearance, speech, facial expressions, and movements. This definition goes beyond the approach taken in traditional research, which treats anthropomorphism as a single dimension, and reveals its theoretical implications as a multi-layered construct.

Visual anthropomorphism

Visual Anthropomorphism constitutes the most fundamental level of anthropomorphism, encompassing both static and dynamic visual manifestations, such as the AI anchor's facial features, body proportions, texture of their clothing, lip-sync accuracy, and naturalness of their voice. Consumers are extremely sensitive to such characteristics, often forming overall judgements based on these details.

"Although the presenter looks artificial at first glance, you will notice that when she smiles, the corners of her mouth turn up, and there are slight fine lines at the corners of her eyes—the kind of natural creases you see when a real person smiles. Just this detail alone makes her feel far more comfortable to watch than those other presenters with stiff facial expressions." (Consumer C04)

In addition to visual characteristics, the naturalness of the voice is a key component of a human-like appearance, directly influencing the perception of how human it sounds.

"Voice is the biggest pitfall. Many AI voices sound instantly like TTS (text-to-speech), with a flat pitch for every syllable. But a good voice has variations in intonation and even emphasizes certain words when highlighting a product's selling points. That's the key to sounding human." (Practitioner P02)

Behavioral anthropomorphism

Behavioral anthropomorphism, in contrast to static appearance, refers to the degree to which an AI-powered digital human's movements, postures and reactions resemble human Behavioral patterns. Respondents generally believed that Behavioral anthropomorphism contributed even more to a sense of presence than visual anthropomorphism.

"The moment that really made me feel it was conscious was when, after I posted a comment, it paused briefly, gave a slight nod, and then said, 'That's a good question, friend.' The timing of that nod and pause made it feel as though it was genuinely listening to me, rather than following a pre-programmed script." (Consumer C06)

Achieving this requires subtle control of the timing and motion design, striking an optimal balance between zero latency and anthropomorphic delay.

"From a technical perspective, response latency is a double-edged sword. Zero latency makes users feel that it is a preset script; however a variable latency of 200–400 milliseconds, combined with appropriate head movements, creates the illusion that it is thinking. This is what we refer to as the anthropomorphism threshold." (Practitioner P03)

Emotional anthropomorphism

Emotional anthropomorphism represents the highest level of anthropomorphism, referring to an AI-powered digital human's ability to convey emotions, express empathy and establish emotional connections. Consumers' descriptions often carry strong emotional undertones, reflecting a deep-seated desire to be understood.

"I know it's just a programme, but once I asked about three different models of robot vacuum cleaners across various price points. It finally said, 'Based on your questions just now, I get the feeling you're quite focused on value for money; this model at two thousand three hundred might be the best fit for you.' When it used the word *feel*, I was genuinely taken aback—can AI really have feelings? But in that very moment, I felt it understood me." (Consumer C05)

Academics further have pointed out, from a theoretical perspective, that emotional anthropomorphism is the key to crossing the Uncanny Valley and establishing quasi-social relationships, and its importance surpasses that of physical realism.

"Anthropomorphism is not about making AI look more and more like a human, but about making people willing to treat it as a social partner. A 60 out of 100 for appearance is sufficient, but emotional expression must score over 80; otherwise, it will fall into the uncanny valley." (Academic A02)

Intelligent interactivity: Cognitive interaction beyond merely responding

In this study, intelligent interactivity is defined as the comprehensive ability of an AI-powered digital human to understand complex user intentions, maintain coherent dialogue with contextual relevance, and provide timely, accurate and contextually appropriate responses. The core of this dimension lies in intelligence; it goes beyond the emphasis on response speed found in traditional human-computer interaction research and points toward a deeper level of semantic understanding.

Response immediacy

Response immediacy is a fundamental element of intelligent interactivity. Although consumers expect interactions to occur instantly, they cannot tolerate a sacrifice in the quality of answers in the pursuit of speed.

"I asked it, 'What's the difference between this phone and the Pro version?' It replied within three seconds but only mentioned the screen size; it did not mention the differences

in the chip and camera, which are the very things I care about the most. I might as well have looked at the specs myself." (Consumer C01)

E-commerce practitioners, however, emphasize that true intelligence lies in the sustainability of the conversation rather than merely the speed of the initial response; conversation retention rates are a more critical operational metric.

"Many technical teams focus solely on optimising the response latency for the first sentence; however, true user satisfaction stems from conversation retention rates, that is whether users are willing to continue asking follow-up questions. This requires an understanding of context; it is not merely a matter of speed." (Practitioner P04)

Response relevance

Response relevance is a hallmark feature distinguishing intelligent interaction from traditional interaction. Respondents repeatedly mentioned local concepts such as "keeping up with the banter" and "staying on topic across multiple rounds", emphasizing the precise alignment of answers with the intent of the question and the ability to retain contextual memory.

"I chatted with it for five minutes, asking about everything from lipstick shades to foundation shades, and it actually remembered that I'd mentioned at the start that I have dry skin. At one point, I digressed to ask about delivery, and after answering, it proactively said, 'Let's get back to the foundation you were asking about earlier.' At that moment, I genuinely felt that I was being taken seriously." (Consumer C02)

This ability to retain contextual information relies on complex intent recognition and dialogue management technologies, requiring a balance between open-domain dialogue and task-oriented interactions.

"Currently, integrating large language models into live streaming is a growing trend, but the challenge lies in maintaining conversational coherence whilst preventing the bot from being sidetracked by users onto completely unrelated topics. This requires a highly refined design of intent recognition and dialogue strategies." (Practitioner P05)

Conversational coherence

Conversational Coherence is reflected in an AI-powered digital human's ability to perceive the conversational context and proactively adjust its interaction strategy, marking a transition from passive responses to active service.

"During one live stream, many people in the comment section asked if there were any discount vouchers for the same product. The AI anchor first explained the process for claiming the voucher, then proactively pinned the voucher link to the top of the screen, adding, 'I see many of you are still asking, so I've placed the link here for everyone to claim with a single click.' This was no longer merely answering a question, but solving a problem." (Consumer C04)

Academics regard this capability as the ultimate form of intelligent interaction, in which the system is not only capable of responding but also of anticipating and adapting, thereby truly possessing the value of a host rather than a mere tool.

"Adaptability is the ultimate form of intelligent interaction. It involves not only understanding what the user says but also what they do not say; for instance, sensing the collective sentiment in the comments section to determine whether the pace of the explanation needs to be adjusted. Only an AI capable of this truly possesses the value of a presenter rather than a mere tool." (Academic A03)

Personalised recommendation: A mechanism for a 'Thousand faces for one person' experience

In this study, personalized recommendation are defined as a service mechanism whereby AI-powered digital humans provide highly customized information content and product recommendations based on users' long-term historical behavior, real-time interactive context, or explicitly expressed preferences. The unique value of this dimension lies in the fact that it represents an area with the greatest potential for differentiated competitive advantage for AI anchors compared to human anchors.

Content relevance

Content relevance forms the basis of personalized recommendation, namely, the degree to which recommended results align with a user's explicit or implicit interests. Consumers are acutely aware of the fine line between precision and excessive intrusion, reflecting the complexities of privacy computing.

"Once it recommended a pair of Bluetooth headphones that I had searched for the previous week but had not bought. My first reaction wasn't that I was being watched, but rather, 'Oh, it knows I need this.' I don't mind recommendations like this; in fact, I feel it saves me the time of searching for them again." (Consumer C02)

However, e-commerce practitioners must carefully balance recommendation accuracy with user privacy protection to avoid a sense of surveillance triggered by cross-domain data retrieval.

"Personalisation must not cross the privacy line. Our current strategy is to make real-time recommendations 'based on behavior within the live stream,' while minimizing cross-domain access to user data from other apps. Although this reduces accuracy, it increases users' sense of security." (Practitioner P01)

Perceived personal attention

Perceived personal attention emphasizes the user's subjective experience of being treated as special, rather than merely the accuracy of algorithmic matching. This perception stems from the recognition of the social cue that something is "specifically for me."

"It says, 'These coffee beans have been specially prepared for you because you previously browsed pour-over equipment.' I know it's an algorithmic match, but when it says 'specially for you', it really makes me feel as though this streamer is my personal shopping assistant." (Consumer C08)

In operational practice, simple strategies such as repeating a user's nickname or remembering past preferences can

significantly enhance users' sense of being noticed, thereby improving user satisfaction.

"The essence of the perception of personalisation is making the user feel that 'this AI remembers me.' Our tests have shown that even simply repeating a user's nickname, or recalling preferences mentioned in their previous conversation, leads to a significant increase in user satisfaction." (Practitioner P02)

Interaction customizability

Interaction customization manifests as the ability to make dynamic adjustments based on real-time interactions. This represents the unique value of AI over static algorithmic recommendations, marking a leap from a *one-size-fits-all* approach to a *thousand faces for one person* approach.

"I told it that I did not like overly sweet snacks, and it immediately removed the chocolate biscuits from the recommendation list and replaced them with savoury nuts. This real-time adjustment convinced me more than any algorithmic claim that 'it understands me'." (Consumer C06)

Academics have pointed out from a theoretical perspective that this real-time interactive customization capability constitutes a competitive moat for AI anchors relative to human anchors, and represents the core competitive advantage of AI-powered digital human live streaming.

"Interactive customisation is the only way AI anchors can surpass human anchors. When broadcasting to an audience of tens of thousands, a human presenter cannot remember what each individual has said; however, an AI anchor can. If utilised effectively, this capability serves as the moat for AI-powered digital human live streaming." (Academic A01)

Discussion

Theoretical contributions

This study provides a feature-dimensional framework rooted in user language for the field of AI-powered digital human live streaming, bridging the conceptual gap between macro-theoretical concepts and micro-level user perceptions. Specifically, this study makes breakthroughs at the following three theoretical levels.

Multidimensional deconstruction of anthropomorphism theory. Using empirical data, this study reveals the existence of emotional anthropomorphism as an independent dimension, identifying it as the key mechanism for overcoming the uncanny valley and establishing emotional connections between humans and machines. Furthermore, the concept of Behavioral anthropomorphism proposed in this study (such as the anthropomorphism threshold for reaction time) provides a concrete operational pathway for Moriuchi's (2021) call for hierarchical measurement, thereby overcoming the theoretical limitations of traditional one-dimensional continua.

Cognitive shift in the interactivity construct. This study shifts theoretical focus from formal characteristics such as response speed to cognitive quality, proposing that intelligent interactivity comprises three elements: response immediacy, response relevance, and conversational coherence. This construct expands upon the dimensional classification proposed by Li et al. (2025), emphasizing that consumers expect proactive intelligence rather than passive responses, thereby extending the applicability of the technology acceptance model (TAM) in AI contexts.

Emotion-technology integration in personalization research. This study finds that consumers' perception of personalization transcends algorithmic accuracy; the core essence is constituted by a sense of being specially treated and the interactive process of real-time adjustment. Specifically, this hierarchical structure comprises three levels: content relevance, perceived personal attention, and interaction customization, thereby broadening the technical research perspective on personalized recommendation. More importantly, this study reveals the tension between privacy concerns and expectations of personalization: consumers anticipate the efficiency and sense of exclusivity brought by highly customized services, while remaining vigilant about the risk of data breaches. This ambivalence is particularly pronounced in contexts of uncertainty regarding privacy breaches (Lim & Kim, 2025).

Implications for practice

The findings of this study offer direct practical guidance for enterprises in the design, operation and iteration of AI-powered digital human live streaming. Based on the independence of these three dimensions, this study proposes the following differentiated resource allocation and optimization strategies.

Adopt a principle of moderation in anthropomorphic design. Consumers do not seek perfect human likeness but rather seek recognizable intelligent agents. Enterprises should adopt a resource-balancing strategy regarding appearance and emotion, prioritizing investment in emotional expression capabilities (such as empathetic responses and emotion recognition). In terms of technical implementation, it is recommended that Behavioral latency be controlled within a variable range of 200–400 milliseconds to create a natural sense of thought and avoid the robotic feel of instant responses. Concurrently, the AI's identity should be clearly labelled to prevent ethical controversies and trust crises arising from excessive anthropomorphism.

Prioritize conversational coherence in intelligent interactions. Industry feedback indicates that conversation retention rates are a better predictor of user satisfaction than response speed to the first sentence. It is recommended to prioritize optimizing multi-turn contextual memory and intent anchoring capabilities, rather than simply reducing latency; furthermore, collective emotional perception should be utilized to proactively trigger service actions, elevating interactions from mechanical question answering to companion-style service.

Balance privacy concerns with recommendation accuracy. It is recommended to establish a tiered customization strategy in which the foundational tier ensures privacy and security awareness based on real-time behavior within the live stream (non-cross-domain data); the emotional tier enhances the sense of being cared for through lightweight per-

sonalization methods such as repeating nicknames; the customization tier triggers real-time adjustments only when users explicitly express preferences, thereby avoiding the intrusive feeling caused by algorithmic predictions.

Research limitations and future directions

Although this study employed maximum difference sampling, the sample remained predominantly concentrated among young consumers in first- and second-tier cities, as well as specific types of practitioners and academics. Its generalizability requires careful validation when extrapolating to other groups (such as users in lower-tier markets and the elderly) and scenarios (such as B2B livestreaming and cross-border livestreaming). Furthermore, retrospective interviews may be subject to memory reconstruction bias; in particular, recollections of emotional experiences and interactive details may not be entirely accurate. Future research could employ the experience sampling method to validate these findings.

Future research could be expanded in three directions: first, developing standardized scales based on this framework to establish construct validity; second, conducting cross-cultural comparisons to examine the differential manifestations of the three-dimensional characteristics across different cultural contexts (such as high-context and low-context cultures); and third, exploring the moderating effects of product type (high-involvement or low-involvement goods) and platform environment (private-domain or public-domain livestreaming) on the perception of these characteristics.

Conclusion

Through in-depth interviews with 19 multi-stakeholders and the application of thematic analysis, this study systematically identified three core characteristic dimensions of AI-powered digital human live streaming: anthropomorphism (the progressive presentation of appearance, behavior and emotions), intelligent interactivity (the integration of immediacy, relevance and coherence), and personalized recommendation (the stratification of content precision, attention perception and interactive customization). This integrated framework transcends the fragmented examination of isolated technical variables found in existing literature, revealing the unique structure of AI anchors as carriers of experience with dual socio-technical attributes and bridging the conceptual gap between macro-level technical discourse and micro-level user perception. By providing operational definitions that combine local relevance with academic dialogue, this study not only lays the conceptual foundation for subsequent scale development and mechanism testing, but also offers a theoretical framework for explaining how AI anchors simultaneously trigger cognitive trust and emotional resonance. Future research should develop standardized measurement tools based on this framework and, through longitudinal tracking and cross-cultural comparisons, examine the dynamic evolution and situational specificity of the three-dimensional structure, thereby supporting the paradigm shift in this field from exploratory to confirmatory research.

References

1. Blumer, H. (1954). What's wrong with social theory? American Sociological Review, 19(1), 3–10. https://doi.org/10.2307/2088165
2. Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. SAGE Publications.
3. Chen, X. (2000). Qualitative research in social sciences. Beijing: Educational Science Publishing House.
4. China International E-commerce Center Research Institute. (2025, May 14). High-quality development report of live streaming e-commerce. Beijing.
5. Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.
6. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008
7. Griffor, E. R., Greer, C., Wollman, D. A., & Burns, M. J. (2017). Framework for cyber-physical systems: Volume 1, overview. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1500-201
8. Hetherington, R., & McRae, R. (2017). Make-believing animated films featuring digital human: A qualitative inquiry using online sources. Animation, 12(2), 166–180. https://doi.org/10.1177/1746847717710738
9. Hwang, J., & Lim, Y. (2025). Adoption of AI Services Based on the Technology Acceptance Model: A Meta-Research Approach. International Journal of Technology Diffusion, 15(1), 1–14. https://doi.org/10.4018/IJTD.394262
10. Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions: A multidimensional conceptualization. Communication Theory, 33(1), 42–52. https://doi.org/10.1093/ct/qtac020
11. Li, R., Ma, B., Zhang, P., & Gao, X. (2025). Research on Mechanisms, Trends, and Impediments to Civil-military Technology Transfer in Emerging Fields. Journal of Beijing Institute of Technology (Social Sciences Edition). 27(1), 127–144, 187. https://doi.org/10.15918/j.jbitss1009-3370.2024.6959
12. Lim, S. E., & Kim, M. (2025). AI-powered personalized recommendations and pricing: Moderating effects of ethical AI and consumer empowerment. International Journal of Hospitality Management, 130, 104259. https://doi.org/10.1016/j.ijhm.2025.104259
13. Machidon, O. M., Duguleana, M., & Carrozzino, M. (2018). virtual human in cultural heritage ICT applications: A review. Journal of Cultural Heritage, 33, 249–260. https://doi.org/10.1016/j.culher.2018.01.007
14. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81
15. McMillan, S. J., & Hwang, J. S. (2002). Measures of perceived interactivity: An exploration of the role of direction of communication, user control, and time in shaping perceptions of interactivity. Journal of Advertising, 31(3), 29–42. https://doi.org/10.1080/00913367.2002.10673674
16. Mori, M. (1970). Bukimi no tani [The uncanny valley]. Energy, 7(4), 33-35.
17. Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. IEEE Robotics & Automation Magazine, 19(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811
18. Moriuchi, E. (2021). An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior. Psychology & Marketing, 38(1), 21–34. https://doi.org/10.1002/mar.21407

19. Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). SAGE Publications.
20. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. Quality & Quantity, 52(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8
21. Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London: John Wiley & Sons.
22. Xu, Y. (W.), Chi, C. G., Gursoy, D., & Cai, R. (R.). (2026). Rethinking AI anthropomorphism: A holistic conceptualization and scale across AI systems and service contexts. Technology in Society, 85, 103189. https://doi.org/10.1016/j.techsoc.2025.103189
23. Wang, N. (2007). The representativeness of case study and its sampling logic. Gansu Social Sciences. (5), 1–4. https://doi.org/10.15891/j.cnki.cn62-1093/c.2007.05.001
24. Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005
25. Zhang, L., Zhang, J., Wang, D., & Mu, J. (2024). Development and validation of an AI virtual streamer scale for livestream e-commerce. International Journal of Human-Computer Interaction, 41(14), 8525–8538. https://doi.org/10.1080/10447318.2024.2411088