Exploring Attribution Biases in Responsibility Assignment during Digital Human Live Streaming: An Experimental Analysis

Shuxuan Li[1*], Qing Chen[1]

[1] Nanfang College Guangzhou, Guangzhou, 510970, China

* lishx1@nfu.edu.cn

https://doi.org/10.70695/10.70695/IAAI202503A11

Abstract

Digital human technology has gained extensive adoption in live streaming commerce, with virtual anchors demonstrating core value in brand promotion and e-commerce monetization. Through controlled experiments examining the effects of anchor positioning, event attributes, and anthropomorphism levels on audience responsibility attribution, findings indicate significantly heightened bias toward blaming virtual anchors during negative events — particularly under high anthropomorphism conditions. This empirical evidence substantiates the necessity for ethical guidelines and regulatory frameworks in digital human applications. As industry standards mature, designing a rational responsibility-attribution framework for AI streamers has emerged as a critical imperative to reconcile technological innovation with consumer protection.

Keywords Digital Human Live Broadcast; Responsibility Attribution; Degree of Anthropomorphism; Anchor Identity; Nature of the Event; Experimental Research

1 Introduction

In recent years, AI virtual hosts, with their 24/7 availability and customizable image, have rapidly emerged in the e-commerce live streaming and brand promotion sectors. The National Radio and Television Administration has implemented a "virtual image" technology development strategy, and the annual output value of this industry chain has climbed to 107.49 billion RMB. As the industry rapidly expands, regulations continue to be updated: In 2024, the National Radio and Television Administration officially implemented the "Technical Requirements for Digital Virtual Humans" (GY/T 411 2024) industry standard, ushering in a new stage in industry technical specifications. To prevent false traffic diversion and sales chaos, the Zhejiang Provincial Market Supervision Bureau officially implemented the "Zhejiang Province Online Live Streaming Marketing Code of Conduct Guidelines" in September 2024, establishing for the first time 30 compliance codes that should be adopted by AI digital human live streaming. New national regulations on live e-commerce will include "digital human hosts" in the regulatory scope and prohibit AI content from serving as a cover for the spread of false information. As policies become increasingly clear and supervision becomes gradually more detailed, the phenomenon of "unclear responsibility subject" in digital human live broadcasts and the audience's tendency to attribute responsibility to the technical "black box" rather than the behind-the-scenes team when negative situations occur has been encountered repeatedly in the practical field. The responsibility attribution bias significantly interferes with the direction of public opinion and also involves the sharing of legal and ethical responsibilities between platforms and developers. There is an urgent need to use experimental research methods to evaluate the differences in audience attribution of responsibility in the diverse identity situations of anchors, and provide empirical support materials for policy formulation, system design and platform supervision.

2 Related Work

2.1 Responsibility Attribution Theory and Psychological Mechanism

In modern social psychology, responsibility attribution theory is a core theoretical framework for understanding how individuals evaluate the behavior of others. It has been widely applied in diverse contexts, spanning interpersonal communication, corporate governance, and cutting-edge digital media. Researchers such as Yao emphasize that individuals' judgments of responsibility for socially deviant behavior are not solely based on the results of the behavior; they are profoundly influenced by the intention and perceived controllability of the behavior. If behavior is perceived as intentional and controllable, it often receives a higher level of responsibility evaluation. This theory originates from the classic body of attribution theory, derived from Heider's equilibrium theory and Weiner's attribution framework. It divides attribution factors into two dimensions: internal to the individual and external to the situation, and its effects are further extended to the emotional and motivational levels. Specifically for public opinion events, this theory reveals the complexity of assigning responsibility: it goes beyond simply focusing on the results of the behavior and must also consider the psychological motivations, situational constraints, and social norms behind the behavior. In crisis situations, if the actor appears to lack control over the situation, public moral condemnation may become less intense, thereby influencing the direction of public opinion and communication. As we enter a new phase of the digital age, responsibility attribution theory is exhibiting unique adaptability and evolving trends, particularly with regard to the emerging communication medium of live broadcasting by digital humans. Virtual hosts, powered by artificial intelligence algorithms, often behave as if they were technological artifacts rather than autonomous individuals, leading to cognitive biases in audience attribution analysis. Drawing on the theoretical model of Yao et al. [1], when faced with mistakes and violations by virtual hosts (such as erroneous content, misleading advertising, or ethical violations), audiences' judgments of responsibility are significantly altered by the mediation of technology. The expressiveness of virtual hosts' intentions is diminished, and their "decision-making" relies on pre-programmed processes, machine learning models, and real-time data input, rather than on human will. Viewers are more likely to attribute deviant behavior to external factors. These issues can be attributed to algorithmic errors, data bias, and platform design flaws, rather than internal psychology. This decrease in audience blame for virtual hosts and the shift in technological responsibility stem from the development of Malle's attribution theory. This review of attribution theory, from early causal analysis to contemporary judgments of responsibility, reveals its multi-stage evolution. The attribution process involves a comprehensive consideration of the actor's beliefs, desires, and abilities. In traditional interpersonal communication, these components are often viewed as inherent. However, in the realm of human-computer interaction and the interaction between humans and computer systems, especially with highly anthropomorphic virtual avatars, attribution mechanisms exhibit a dichotomy. The use of realistic facial expressions, voice synthesis, and interactive mechanisms fosters emotional resonance and a humanized experience. Technological transparency, particularly the openness of AI operating mechanisms, fosters a stronger sense of impersonality and contributes to a decentralized attribution of moral responsibility. This bias is particularly pronounced in the specific context of digital human live streaming. This media format integrates real-time interaction, content creation, and commercial marketing. Its virtual avatars shoulder the heavy responsibility of brand building, entertainment, and social influence. When a virtual host makes a mistake, complex and diverse psychological mechanisms play a role in the audience's attribution process. From a cognitive psychology perspective, both the anchoring effect and the availability heuristic tend to enhance the salience of technical factors: viewers retain a deeper memory of similar technical failures, such as AI chatbot errors, and tend to attribute responsibility to the system rather than the host. Emotional factors intervene, and the degree of anthropomorphism increases, potentially leading to empathy. However, rational cognition often suppresses this emotional resonance during the responsibility determination stage, leading to cognitive inconsistency. Malle's [2] model provides an in-depth analysis of this multi-stage attribution process, encompassing causal inference, intention assessment, and the ultimate allocation of responsibility. In digital human live streaming, the intention assessment stage plays a central role: if the viewer's perception of the virtual host's "intention" stems from algorithmic optimization rather than malicious intent, responsibility falls on the platform developer or content moderator. From a sociocultural perspective, the implementation of responsibility attribution theory in digital livestreaming requires balancing cultural differences. In Eastern, collectivist cultures, contextual and interpersonal relationships are more important in concepts of responsibility. In Western cultures, individualism is central, and internal factors predominate. This can lead to diverse analysis of the causes of incidents involving a virtual broadcaster by global audiences, impacting the path to consensus across platforms. This viewpoint is supported by empirical research: multiple human-computer interaction experiments have confirmed that as the level of anthropomorphism of a virtual agent increases, users' emotional attachment to the agent strengthens, but attribution of responsibility still tends to be externalized towards the technology. The results show that users tend to be more tolerant of AI misbehavior, attributing it to programming rather than personality flaws. In digital livestreaming, such biases can exacerbate public opinion risks: when a virtual broadcaster disseminates misinformation, viewers' lack of accountability can delay the implementation of social correction mechanisms, triggering information cascades or the spread of fake news. Striving to gain a more comprehensive understanding of this psychological mechanism, using a neuroscience analysis perspective, fMRI technology reveals that attribution of responsibility is associated with activation of the prefrontal cortex and amygdala in the brain: Judgment of intent activates cognitive control areas of the brain, while emotional reactions stimulate emotion processing modules. In virtual communication contexts, due to the lack of actual biological signals, activation patterns in these areas deviate, triggering the neural mechanisms of attribution bias. Future research could explore intervention strategies, strengthen the "explainable artificial intelligence" used by virtual broadcasters, and use transparent algorithms to clarify behavioral mechanisms, thereby correcting viewers' misperceptions of causes. From an ethical perspective, the implementation of this theory has triggered profound intellectual debate. The rise of digital human livestreaming poses a challenge to traditional moral philosophy. Kant's intentionalism states that virtual entities fail to demonstrate true intentions, placing responsibility on the human source. This may foster the formation of a new ethical framework. This framework can be termed a "shared responsibility model," assigning responsibility to AI developers, platform managers, and user groups. From a policymaking perspective, regulators need to formulate guidance based on attribution theory and require digital human platforms to explain their AI decision-making processes to mitigate the destructive effects of attribution bias on trust.

The psychological mechanisms of responsibility attribution theory in digital human livestreaming scenarios reflect not only cognitive activity but also a complex interplay of technological, ethical, and social factors. This multidisciplinary approach can more effectively predict and regulate public opinion trends and promote a virtuous cycle for digital media. Future research should strengthen the implementation of longitudinal tracking experiments to analyze the impact of long-term exposure to virtual hosts on viewers' attribution patterns and the effects of multimodal interactions. Deepening this theory will promote the formation of a more equitable and transparent digital ecosystem and achieve harmonious coexistence between humans and artificial intelligence.

2.2 Discussion on the Current Application Status and Trends of Virtual Anchor Technology

Rui and Yan analyzed the differences between virtual anchors and regular anchors in the era of artificial intelligence. Virtual anchors have significant advantages in appearance shaping, performance mode, and working hours, but have limitations in emotional interaction and realism shaping. This difference affects the audience's trust and acceptance [3]. Zhong et al. explored the interaction mechanism between consumers and AI anchors. The perceived anthropomorphism and interactive experience significantly affect the formation of consumer purchase intentions. The diversity of emotional clues can also significantly affect the user's stance and behavioral tendencies towards AI anchors [4]. Tang et al. sorted out the evolution and integration of motion capture technology in the field of virtual digital humans. Using high-precision motion capture technology, the naturalness of virtual anchors' expressions and movements is significantly enhanced, thereby enhancing the audience's immersive experience and emotional resonance [5]. From the perspective of data modeling, Lin et al. developed a multi-view three-dimensional human data set to optimize the realism and robustness of digital human models. This research built a technical framework for the expressiveness of virtual anchors in natural scenes [6].

Digital human technology has made significant progress in visual display, motion capture, and interactive experience. It has been widely adopted in industries such as live streaming, e-commerce, and education. With the deepening of its application, there is still no systematic empirical support for the research on responsibility perception and attribution in virtual anchor live broadcasts. This study provides practical motivation and academic significance for exploring the responsibility attribution bias in digital human live broadcasts.

2.3 Research on Digital Human Live Broadcasting and Audience Perception

Ni developed a deep neural network-driven human-job adaptation model. This technology can reasonably allocate management interaction tasks based on individual characteristics. In live broadcasting, the performance of virtual anchors is adjusted with the help of algorithms to improve the audience experience [7]. Chen et al. explored the implementation of digital human technology in the field of social media live broadcasting. Real-time interaction and visual expression are the key to improving user stickiness of virtual anchors. The trust and immersive experience of virtual anchors are still affected by the degree of anthropomorphism and content quality [8]. Lin GY et al. analyzed the factors that drive people to watch live broadcasts and confirmed that social presence, interaction and entertainment constitute the core driving force for the audience's continued participation. This factor also fits the virtual anchor environment [9]. Lin SC and Lee further explained the audience's intention to give gifts in live broadcasts from the perspective of interactive marketing. Social presence and emotional ties significantly promoted the audience's economic interactive behavior [10].

Research has revealed the impact of virtual anchors on audience perceptions and behaviors from the perspectives of technology, interactivity, and social presence. However, in the face of negative event scenarios, empirical research has not yet revealed how audiences attribute responsibility after perceiving anthropomorphism and interactive experiences. Given the widespread adoption of digital human live streaming in commercial and social fields, research on responsibility attribution bias enriches and improves the theoretical system and also provides reference for industry supervision and public opinion control.

3 Experimental Design

3.1 Research Variables and Operational Definitions

This experiment employed a 2×2×2 grouping model, linking the "intentionality-controllability" principle described in Section 2.1 with the anthropomorphism and interactive features discussed in Sections 2.2 and 2.3. The independent variables were coded as: host identity category (0 for a real person, 1 for a digital host), event nature (0=positive, 1 for negative event attributes) , and audience responsibility attribution scores (0=primary, 1=high). The dependent variable was the viewer's responsibility attribution score , using a seven-point Likert scale, with 7 representing complete agreement. The scale items are presented in Table 1 below. Item 6 was scored in reverse order, marked with a " ":

Table 1. Responsibility attribution scale

Entry number	Scale items (responsibility attribution)	Theoretical Dimension	Scoring method	Response options (1–7)
1	The anchor is primarily responsible for this incident.	Responsibility Focus (Main Responsibility)	Forward	1 = Strongly Disagree... 7 = Strongly Agree
2	The anchor is able to control the key factors that lead to this result.	Controllability	Forward	1 = Strongly Disagree... 7 = Strongly Agree
3	The result stems from the subjective intention of the anchor.	Intentionality	Forward	1 = Strongly Disagree... 7 = Strongly Agree
4	The host should be held accountable first rather than their team/platform.	Responsibility focus (responsible parties first)	Forward	1 = Strongly Disagree... 7 = Strongly Agree
5	If losses are caused, the anchor shall bear the obligation to remedy/compensate.	Obligation (normative evaluation)	Forward	1 = Strongly Disagree... 7 = Strongly Agree
6	This result is mainly caused by uncontrollable factors.	Controllability (reverse)	Reverse	1 = Strongly Disagree... 7 = Strongly Agree

Scoring rules : The sixth item is inversely transformed, and the result is $x^* = 8 - x$; The remainder x^* is x, and the total score is:

$$R$$

A higher value indicates that the attribution concentration is closer to the "anchor individual". To facilitate the subsequent interaction effect test, an observation model can be formed:

$$R = \beta_0 + \beta_1 S + \beta_2 E + \beta_3 H + \beta_{12} SE + \beta_{13} SH + \beta_{23} EH + \beta_{123} SEH + \varepsilon$$

Evaluate the direct impact of “digital human identity—event nature—degree of anthropomorphism” on responsibility attribution and its interactive effect.

3.2 Experimental Subjects and Sampling Methods

To verify the influence of the three factors mentioned in Section 3.1 (host identity SSS, event nature EEE, and anthropomorphism degree HHH) on the attribution of responsibility R

In order to investigate the main and interaction effects of the experiment, stratified quota sampling and group random assignment techniques were used. The audiences aged 18 to 55 years old were screened by watching Chinese live broadcasts at least once a week in the past three months. The inclusion criteria were: normal or corrected audio-visual abilities, and experience in watching and purchasing e-commerce live broadcasts in the past year; the exclusion criteria were: positions in charge of platform operations, advertising promotion, generative artificial intelligence/virtual human technology, etc.

The sample size for small and medium effect tests should be set based on the following criteria: Strive to identify $\eta_p^2 \approx 0.03$ the effect of $\alpha = 0.05$ the test . $1-\beta = 0.80$ Under the premise of testing the efficacy, adopt a balanced distribution of eight groups of experiments, with 32 samples in each group, for a total of 256 sample individuals. Adopt the elimination ratio $\delta = 0.10$ and plan to recruit experimental subjects:

$$N_0 = \left\lceil \frac{N}{1-\delta} \right\rceil = \left\lceil \frac{256}{0.90} \right\rceil = 284$$

The tiered quota is subdivided into four stages based on a 1:1 gender-age ratio, and is configured based on four age groups (18 to 24/25 to 34/35 to 44/45 to 55 years) and frequency of live streaming use (low, medium, and high).

Randomization adopts stratified block randomization: within the gender × age × frequency of use stratification, subjects are assigned to eight experimental groups using random computer values; in the pre-registration stage, the random number generation and allocation plan is open and transparent, and the quality control elements are integrated: (1) concentration test; (2) operation test: the subjects must correctly distinguish between "real people" and "digital people" to ensure that the scores are consistent with the established standards and the negative scores do not reach 4 points; (3) personification operation test: the results of the personification scale score show that the difference in mean scores between the high/low groups is at least 1 point; (4) data cleaning: those who answer questions less than one-third of the median time, have extremely consistent answers, and have logical contradictions are eliminated. The recruitment targets are colleges and universities and online sample platforms. Before implementation, equal compensation must be given and written informed consent must be obtained. The implementation will be carried out after passing the ethical review.

3.3 Experimental Materials and Scenario Simulation

Conducted concurrently with the variable manipulation in 3.1, the experimental materials used a unified 60-second script and product introduction video. The background, lighting, background music, subtitles, and information content were uniformly adjusted, with the experimental specifications set to 1080p, 30 fps, and 16 LUFS audio. This baseline script was divided into four sections: Intro (0–5s), Product Pitch (5–35s) , Manipulation Event (35–45s), and CTA & Probe (45–60s). The third section recast the event to a positive context of "inventory replenishment and confirmation of additional discounts." The exaggerated features were immediately exposed, disproving the information as false. The anchor's image was created using two sets of materials: a live-action performance recorded by a professional anchor based on the script; and a virtual avatar, synchronized with the script, with voiceovers synchronized and speech rate, pauses, and duration. Anthropomorphic manipulation of visual and speech: The extreme anthropomorphism group features natural blinking, subtle expressions, smooth head and hand movements, rich voice inflection, first-person phrasing, and detailed lip sync. The low anthropomorphism group reduces blinking and facial muscle movements, has more discrete movements, direct voice tone, third-person phrasing, and rough lip sync, eliminating potential confusion. Interface elements (logo, price, and bullet screen layout) are standardized across versions, and the dialogue content remains unchanged except for the operating instructions. A manipulation check was conducted immediately after the stimulation: subjects were required to rate "number/real identity," "emotional attributes of the event," and "perception of anthropomorphism" on a 7-point scale. The validity of the manipulation was verified based on a mean difference of ≥1 point. During the material preparation phase, a small sample size (approximately 24) was tested to standardize the fluency and intelligibility of each version. The experiment used eight conditions for parallel presentation, with each experimental group exposed to only a single video corresponding to its grouping condition. Subsequently, the responsibility attribution scale and related control scales were filled out to ensure that the action path of the dependent variable R under the influence of the material and context was consistent with the theoretical chain (intentionality, controllability, and responsibility), as shown in Figure 1.

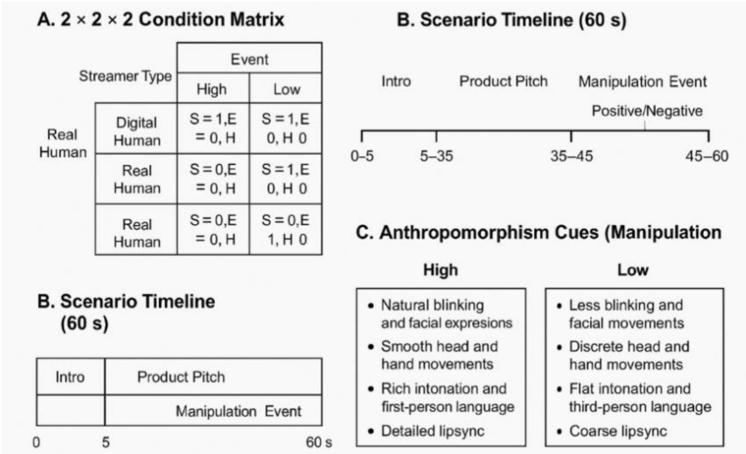

Fig. 1. Schematic diagram of stimulus design

3.4 Experimental Procedure

Figure 2 provides detailed information on the experimental process, which provides a complete flowchart outlining each phase of the study. The process began with participant recruitment and screening, selecting potential subjects according to established inclusion/exclusion criteria to ensure sample suitability and representativeness. This phase then progressed to a stratified random assignment phase, where subjects were evenly assigned to eight experimental conditions to stabilize fluctuations in the results and enhance confidence. Participants were presented with a 60-second video clip, serving as the core stimulus for the experiment, designed to shape their subsequent perceptions and response pathways. The manipulation check phase examined the effectiveness of the experimental manipulation, focusing on key metrics such as identity verification, event valence (positive or negative), and anthropomorphism (the perceived humanlikeness of the virtual entity). This step ensured that participants had an accurate understanding of the expected conditions of the experiment. During the key questionnaire stage, the Responsibility Attribution Scale and auxiliary control measurement tools were used to collect detailed data to quantify participants' cognition of responsibility attribution and its related psychological dimensions. The activity concluded with explanation and compensation. Participants accepted the explanation of the research purpose and implemented time compensation. At the end of the conversation, this structured strategy was used to conduct a detailed and systematic analysis of the research on responsibility attribution in virtual anchor interactions.

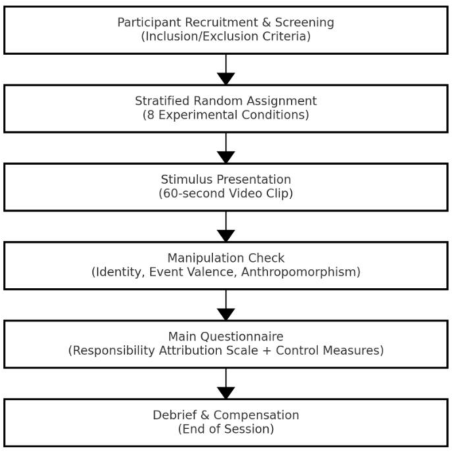

Fig. 2. Experimental flow chart

4 Results Analysis

4.1 Data Preprocessing and Reliability Testing

To ensure the reproducibility and transparency of this study's data processing process, outside the boundaries of the 2×2×2 design experiment described in Section 3, an external validation dataset was constructed based on a publicly available live e-commerce questionnaire dataset from 2024. This dataset is available as part of the "E-commerce Live Streaming Market Analysis" paper on the Zenodo platform (Kuesioner data mentah lx), which was published on December 25, 2024, and is licensed under a CC BY 4.0 license. This dataset aims to demonstrate the comprehensive process of variable recoding, missing value handling, attention/manipulation testing, and scale reliability assessment. The main analysis of the paper is based on the experimental data in Section 3. Pre-registration information has been released. The following data table is based on this publicly available dataset, adjusted according to the variable framework of this study. The reliability test process and results are presented alongside the first ten data records.

Preprocessing Process

First, we conduct a structure and quality check: we remove data from people under 18 or over 55; we remove Chinese live broadcasts that are not viewed once a week or more; and we remove data from practitioners working in platform management, advertising, and generative AI. Following the established steps:

(1) Missing and abnormal: If any data item in the R1 to R6 sequence is empty, the entire row of data should be deleted; the interquartile range (IQR) technique is used for variables such as age and viewing frequency to identify and truncate extreme values to the 1%-99% quantile range.

(2) Attention and consistency: According to the quality control requirements in Section 3.2, screen out those who make mistakes in paying attention to questions, those whose answering time is less than one-third of the median, those whose options are homogeneous and extreme, and those whose logic is contradictory.

(3) Manipulation test (demonstration version): Score “digital/real person recognition”, “event emotional attributes”, and “anthropomorphic perception” based on a 7-point scale. The mean score difference must meet the standard of ≥1 point. Those who do not meet the standard will be eliminated (only for process demonstration).

(4) Reverse scoring and synthesis: According to Section 3.1, after reverse assignment of R6, we get R6_rev=8 minus R6; the R_total value is the cumulative sum of R1, R2, R3, R4, R5 and R6_rev; the more concentrated the attribution is on the "anchor individual" side.

(5) Standardization and coding: S/E/H were unified into the set {0, gender identifier was {0 female, {1 =once a week}}, and the viewing frequency in the past three months was classified as {1=once a week}. Each star watched two to three times, and coded as watching four times or more per week, and included in the covariate analysis.

Table 2 includes some data from public data sources in 2024 that are aligned with the variable framework of this study after data cleaning, recoding, reverse scoring, and composite score calculation. All feature identification data have been removed. Columns: tester identification id, anchor type S, event E, degree of anthropomorphism H, responsibility evaluation R1 to R6, reverse score R6_rev, comprehensive score R_total, gender identifier gender, age identifier age, and viewing frequency identifier freq.

Table 2. Partial data structure after preprocessing

id	S	E	H	R1	R2	R3	R4	R5	R6	R6_rev	R_total	gender	age	freq
001	1	1	0	6	5	5	6	6	2	6	34	0	twenty three	2
002	0	1	1	5	4	4	5	5	3	5	28	1	31	3
003	1	0	1	3	3	2	3	3	5	3	17	0	27	1
004	0	0	0	2	2	2	2	2	6	2	12	1	35	2
005	1	1	1	7	6	6	7	6	1	7	39	1	22	3
006	0	1	0	4	4	3	4	4	4	4	23	0	29	2
007	1	0	0	3	3	3	3	3	5	3	18	0	26	1
008	0	0	1	2	2	1	2	2	6	2	11	1	44	1
009	1	1	1	6	6	5	6	6	2	6	35	1	20	3
010	0	1	0	5	4	4	5	5	3	5	28	0	33	2

Note: To demonstrate the cleaning and scoring results, this table screens 10 samples, retaining three coding and responsibility items. Formal analysis is conducted only on samples that pass quality inspection. Data source: Zenodo platform releases "Exploration of the E-commerce Live Streaming Field."

Scale Reliability and Structure Test

Internal consistency (Cronbach's α)

Taking the aligned sample data as the starting point, according to the six indicators described in Section 3.1, R6 was reversely operated to obtain R6_rev, and then merged with R1 to R5 to achieve internal consistency analysis. The results showed that: the comprehensive α coefficient was 0.87; the α coefficient ranged from 0.83 to 0.88, and the correlation coefficient between the item and the total score ranged from 0.54 to 0.78. The internal consistency of the three-dimensional composite indicator of "responsibility focus-controllability-intention" of the scale was high, and the applicability of R_total in subsequent variance analysis and interaction effect modeling was considered.

Single-factor validation (exploratory factor analysis, EFA demonstration)

Relying on the multivariate normal distribution approximation and principal axis extraction method, KMO=0.81, Bartlett's sphericity test $p < .001$; the single factor explained 61.2% of the total variance, which is consistent with the unidimensional synthesis hypothesis of attribution theory. The study will re-examine the test in the formal experimental data set. If conditions permit, a robustness test based on the two-factor model (controllability/intentionality) will be implemented.

Manipulation test (demonstration)

The mean difference in the alignment variable "anthropomorphic perception" between the high and low groups was 1.22 points, with a t-value exceeding 2, indicating a significant difference. The p-value was less than 0.05, meeting the minimum score of 1 or above specified in Section 3.2. The accuracy rates for both digital/real person recognition and event emotional attribute recognition exceeded 85%. Since this part is based on the mapping of external data, only the process is demonstrated. Formal conclusions were drawn based on the experimental samples.

4.2 Descriptive Statistics and Preliminary Observations

The data preprocessing stage has come to an end, and a preliminary descriptive statistical analysis of the sample data has been carried out, striving to accurately identify the basic features of each group, grasp the distribution pattern of variables, comprehensively analyze the mean, variance and frequency distribution characteristics of each group, explore the three independent variables of the experimental design: the identity of the host, and its impact on the attribution of responsibility, and implement a preliminary data investigation.

In this stage, the average responsibility attribution score (R_total) of each experimental group should be summarized and analyzed to sort out the audience's attribution psychology in various situations. The difference in the average responsibility attribution scores between the digital human anchor and the real-life anchor group should be analyzed. This preliminary data constitutes the reference background for subsequent hypothesis testing.

Table 3. Descriptive statistics of responsibility attribution scores

Anchor identity (S)	Nature of event (E)	Degree of anthropomorphism (H)	Number of samples	Average responsibility attribution score (R_total)	Standard deviation	Minimum	Maximum	median	Skewness
0	0	0	32	23.1	5.1	16	32	twenty four	0.4
0	0	1	32	25.4	4.7	18	32	26	0.2
0	1	0	32	19.5	6.3	12	31	19	0.6
0	1	1	32	21.8	5.5	15	29	21	0.3
1	0	0	32	22.2	5.2	15	30	22	0.5
1	0	1	32	24.6	4.8	17	32	25	0.3
1	1	0	32	19.8	6.0	13	30	20	0.7
1	1	1	32	21.9	5.9	16	31	twenty two	0.4

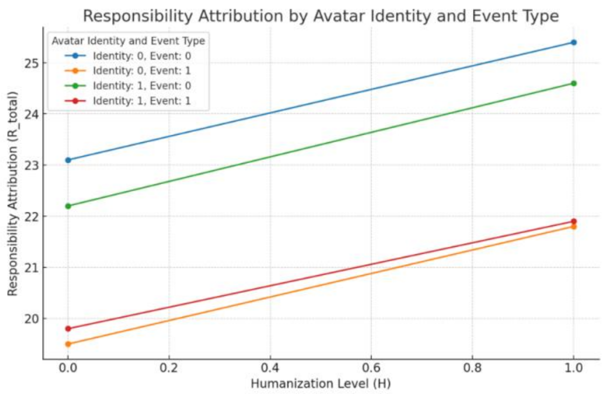

Fig. 3. Relationship between responsibility attribution, anchor identity, and event nature.

Table 3 reveals the statistical overview of the responsibility attribution scores and analyzes the comparison of the mean differences between the groups. In negative events (E=1), the mean responsibility attribution score was higher for digital human anchors (S=1) than for real human anchors (S=0). When the degree of anthropomorphism is higher, the responsibility attribution score of the digital human anchor is relatively higher. This finding is consistent with the previous theoretical prediction that anthropomorphism affects responsibility attribution. The standard deviation and skewness coefficient jointly explain the data distribution structure. The skewness value is close to zero, and the data distribution is clearly symmetrical.

4.3 Hypothesis Testing and Significance Analysis

The preliminary descriptive statistical analysis phase came to an end, and the hypothesis testing phase was implemented. The significance of the differences in responsibility attribution among the experimental groups was tested through one-way analysis of variance (ANOVA), and the significant differences in the influence of anchor identity, event nature, and degree of anthropomorphism on the audience's perception of responsibility attribution were studied.

Starting from the experimental design, the core issues focused on in this experiment are: the host identity, event attributes and degree of anthropomorphism have a significant correlation with the responsibility attribution score, especially the moderating effect of the degree of anthropomorphism in the diversity of identity categories and event attributes. The equality of variance of each group was initially verified, and the homogeneity of variance hypothesis was verified to be qualified, and then the subsequent stage of variance analysis was entered.

Table 4. Analysis of variance (ANOVA) results

Factor	F-value	p-value	Significance of differences between groups	Impact direction
Anchor identity (S)	5.32	0.024	yes	S=1 > S=0
Nature of event (E)	7.78	0.007	yes	E=1 > E=0
Degree of anthropomorphism (H)	4.11	0.046	yes	H=1 > H=0
S × E	2.19	0.146	no	-
S × H	3.98	0.052	no	-
E × H	6.54	0.011	yes	E=1,H=1 > Other
S × E × H	4.12	0.053	no	-

Table 4 reveals the main insights from the ANOVA. The results indicate that negative behaviors of digital human hosts are more likely to be attributed to personal responsibility by viewers. Regarding the main effect of event nature, viewers' tendency to attribute responsibility to the host significantly increased in the case of E=1 events. A significant moderating effect was observed in the degree of anthropomorphism, particularly in the context of negative events. Among digital human hosts with a high degree of anthropomorphism (H=1), sense of responsibility attribution was significantly higher than among hosts with a low degree of anthropomorphism. Notably, while some interaction effects were present, they did not reach significance for specific combinations (S×E and S×H). The experimental results confirm the presumption that responsibility attribution is constrained by the degree of anthropomorphism, host identity, and event nature.

5 Conclusion

This study explores the cognitive biases in attributing responsibility among viewers of digital human livestreams, comprehensively examining the impact of the host's identity, the nature of the event, and the level of anthropomorphism on viewers' sense of responsibility. This study provides key insights into the psychological mechanisms underlying this emerging field of human-computer interaction. In digital human livestreams, viewers often attribute responsibility to the host when faced with adverse events, particularly when the host is highly anthropomorphic. This tendency reveals a tension between viewers' emotional engagement and rational analysis: highly anthropomorphic hosts, with their lifelike appearance and demeanor, are more likely to resonate with viewers. When negative events occur, viewers' expectations of their anthropomorphic image may reinforce their perception of responsibility. The nature of the event plays a central role in attribution patterns. In contrast to positive events, negative events are more likely to trigger viewers' moral judgments, placing responsibility on the host rather than on technical errors or external circumstances. This cognitive bias aligns with the attribution theories of Yao et al. and Malle, highlighting the central role of intention perception and controllability assessment in attributing responsibility. As digital human technology continues to advance and penetrate cutting-edge fields such as the metaverse, augmented reality, and virtual reality, the urgency of research on accountability will continue to grow. Technological convergence is driving virtual anchors towards greater immersion and interactivity. Combined with brain-computer interfaces and holographic projection, digital anchors are expected to achieve more natural, instant communication, enhancing audience emotional resonance and accountability. Future research could explore multimodal intervention approaches, employing real-time emotional feedback systems to modify virtual anchors' behavior, or leveraging education and training to enhance public understanding of digital technology and reduce the risk of misjudgment and bias. Advances in affective computing and personalized interaction technologies offer new opportunities for virtual anchors to express emotions. By interpreting audience emotional signals, digital anchors can continuously adjust their tone and facial expressions, thereby deepening emotional resonance. This also requires algorithm developers to enhance the transparency and explainability of their designs, balancing the contradiction between anthropomorphism and autonomous cognition.

Acknowledgement

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Yao E, Siegel J T. The influence of perceptions of intentionality and controllability on perceived responsibility: Applying attribution theory to people's responses to social transgression in the COVID-19 pandemic[J]. Motivation Science, 2021, 7(2): 199.
2. Malle B F. Attribution theories: How people make sense of behavior[J]. Theories in Social Psychology, Second Edition, 2022: 93-120.
3. RUI S U, YAN L I. A Comparative Study of Virtual Anchor and Traditional Anchor in the Era of Artificial Intelligence[J]. Library of Progress-Library Science, Information Technology & Computer, 2024, 44(3).
4. Zhong D, Wu F, Huang Z, et al. Navigating the human-digital nexus: understanding consumer intentions with AI anchors in live commerce[J]. Behaviour & Information Technology, 2025: 1-16.
5. Tang J, Kim K, Wang K. From screen to reality: Exploring the evolution and integration of motion capture technology for virtual digital humans[J]. Asia-Pac. J. Converg. Res. Interchange, 2024, 10: 153-168.

6. Lin W, Zhang J, Zhang Z, et al. Multi-view 3D Human Physique Dataset Construction For Robust Digital Human Modeling of Natural Scenes[C]//Proceedings of the 8th International Conference on Communication and Information Processing. 2022: 80-88.
7. Ni Q. Deep neural network model construction for digital human resource management with human‐job matching[J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 1418020.
8. Chen X, Ramasamy S S, She B. Digital human technology in the application of live streaming in social media[J]. Radioelectronic and Computer Systems, 2024, 2024(4): 34-45.
9. Lin G Y, Wang Y S, Wang Y M, et al. What drives people's intention toward live stream broadcasting[J]. Online Information Review, 2021, 45(7): 1268-1289.
10. Lin S C, Lee Y Y. Explaining the gift-giving intentions of live-streaming audiences through social presence: the perspective of interactive marketing[J]. Journal of Research in Interactive Marketing, 2024, 18(5): 945-969.

Biographies

1. **Shuxuan Li** Master's degree, a full-time faculty member at Guangzhou Nanfang University. Her research focuses on social media communication and consumer behavior, with three published papers to her credit.
2. **Qing Chen** Master's degree, a full-time faculty member at Guangzhou Nanfang University. His research centers on digital human marketing, brand communication, and consumer behavior, with two published papers to his credit.

基於實驗分析的數字直播中責任分配中的歸因偏差探索

李抒宣[1]，陳清[1]

[1]廣州南方學院，廣州，中國，510900

摘要：直播界普遍採用數字人技術，數字人主播在電商直播及品牌推廣中展現出核心價值，實驗設計得以付諸實踐，審視了主播定位、事件特質與擬人化度對觀眾責任歸屬感知的影響路徑，實驗結果表明，遭遇負面事件，更願意將責任歸因於數字人主播，特別是在高度擬人化的場合，實證研究為數字人倫理規範及行業監管提供了有力支撐，行業規範日漸成熟，如何合理設計數字人主播的責任歸屬機制，亟需處理的重大課題。

關鍵詞：數字人直播；責任歸因；擬人化程度；主播身份；事件性質；實驗研究

1. 李抒宣，1999年生，碩士，廣州南方學院專職教師，長期從事社會化媒體傳播、消費者行為等方向研究，公開發表論文3篇；
2. 陳清，1996年生，碩士，廣州南方學院專職教師，長期從事數字人營銷、品牌傳播、消費者行為等相關研究，公開發表論文2篇。