

THE INFLUENCE OF VOICE ANTHROPOMORPHISM IN GENERATIVE AI-BASED SHORT VIDEO CONTENT ON PURCHASE INTENTION



Prisalia Amanatus Sholihah¹
Universitas Indonesia, Jakarta, Indonesia
prisalia.amanatus@ui.ac.id

Karto Adiwijaya²
Universitas Indonesia, Jakarta, Indonesia
karto.adiwidjaya@ui.ac.id

Abstract

The rapid advancement of generative artificial intelligence has transformed digital marketing through AI-generated voice-overs in short-form video content. However, empirical understanding of how voice anthropomorphism influences consumer behavior remains limited. This study investigates the impact of short video content characteristics and AI voice anthropomorphism on purchase intention, with cognitive trust as a mediating variable. A quantitative approach was employed, surveying 384 Indonesian social media users exposed to AI-generated voice content. Data were analyzed using Partial Least Squares Structural Equation Modeling (PLS-SEM). Results indicate that content characteristics (usefulness, ease of use, entertainment) and all four anthropomorphism dimensions (appearance, behavioral, cognitive, emotional) significantly and positively influence cognitive trust and purchase intention. Emotional anthropomorphism emerged as the strongest predictor of cognitive trust. Cognitive trust significantly mediates the relationships between independent variables and purchase intention. These findings highlight the critical role of human-like vocal traits and rational trust evaluation in driving consumer purchase decisions, offering actionable insights for optimizing AI-driven marketing strategies in short-video ecosystems.

Keywords: Generative Artificial Intelligence, Voice Anthropomorphism, Short Video Content, Cognitive Trust, Purchase Intention

INTRODUCTION

The proliferation of short-form video platforms such as TikTok, Instagram Reels, and YouTube Shorts has fundamentally reshaped digital marketing communication. These platforms prioritize rapid, snackable content that captures declining user attention spans through multimodal stimuli, including dynamic visuals and engaging audio. Concurrently, the emergence of generative artificial intelligence (GAI) has enabled the mass production of high-quality, human-like synthetic voice-overs, significantly reducing production costs and scaling content creation. Despite rapid industry adoption, academic research on the psychological impact of AI-generated voices remains scarce. Most existing studies focus on visual anthropomorphism in virtual influencers or text-based chatbots, leaving a critical gap in understanding how auditory human-likeness (voice anthropomorphism) shapes consumer responses in short-video ecosystems.

To address this gap, this study examines how the characteristics of short video content and the anthropomorphic qualities of AI-generated voices influence purchase intention. Drawing on the Stimulus-Organism-Response (S-O-R) framework, we position cognitive trust as a key psychological mediator. Cognitive trust reflects consumers' rational evaluation of the competence, reliability, and credibility of the AI voice as an information source before forming purchase intentions. This research integrates insights from technology acceptance models and anthropomorphism theory to empirically test whether human-like vocal traits, combined with functional and entertaining content attributes, can foster trust and drive purchasing behavior. The findings aim to contribute theoretically to digital marketing literature and provide practical guidelines for marketers leveraging AI voice technologies in short-video campaigns.

REVIEW OF LITERATURE

Anthropomorphism refers to the attribution of human characteristics, emotions, or intentions to non-human entities (Epley et al., 2007). In digital marketing, anthropomorphism enhances social presence and emotional engagement, particularly when AI systems exhibit human-like behaviors. Within the context of AI-generated voice, anthropomorphism is operationalized through four dimensions: appearance (auditory human-likeness and natural vocal quality), behavioral (natural intonation, pacing, and contextual adaptability), cognitive (perceived logical reasoning, creativity, and information processing), and emotional (expression of empathy, enthusiasm, and affective resonance) (Li & Huang, 2024; Dabiran et al., 2024). Although appearance anthropomorphism traditionally relates to visual cues, in voice-based interfaces it translates to the perceived acoustic realism and human-like tonal qualities of synthetic speech (Schreibelmayer & Mara, 2022).

Short video content characteristics serve as external stimuli that shape consumer perceptions. Usefulness reflects the extent to which video content provides practical, product-related information that aids decision-making (Luo et al., 2025). Ease of use pertains to the clarity, pacing, and intuitive accessibility of the content, minimizing cognitive load during rapid consumption (Davis, 1989). Entertainment captures the content's ability to engage, delight, and create an immersive viewing experience, which reduces defensive attitudes toward promotional messages (Luo et al., 2025).

Cognitive trust is defined as a rational, competence-based assessment of a source's reliability, expertise, and consistency (McAllister, 1995). In AI-mediated communication, consumers evaluate whether synthetic voices demonstrate sufficient credibility and logical coherence before accepting marketing messages. Prior research confirms that anthropomorphic cues and content utility significantly enhance cognitive trust, which subsequently drives behavioral intentions (Li & Huang, 2024; Luo et al., 2025). Purchase intention, the dependent variable, represents the consumer's willingness and likelihood to acquire a promoted product, serving as a proximal indicator of actual buying behavior (Ajzen & Fishbein, 1969). Within the S-O-R paradigm, short video attributes and voice anthropomorphism act as stimuli, cognitive trust functions as the organism's internal processing, and purchase intention emerges as the behavioral response.

RESEARCH METHOD

This study employs a quantitative, explanatory research design to test the proposed theoretical model. The target population consists of Indonesian social media users aged 18 and above who have been exposed to short-form video content featuring AI-generated voice-overs within the past three months. A total of 384 valid responses were obtained through purposive sampling, meeting the minimum sample size requirements for PLS-SEM analysis (Hair et al., 2021). Data were collected via an online questionnaire distributed through social media platforms and targeted advertisements.

The research instrument utilized a seven-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree) to measure all constructs. Independent variables included usefulness, ease of use, entertainment, and the four dimensions of voice anthropomorphism (appearance, behavioral, cognitive, emotional). Cognitive trust served as the mediating variable, while purchase intention was the dependent variable. Prior to main data collection, a wording test and pretest (n=30) were conducted to ensure clarity, content validity, and internal consistency. Confirmatory factor analysis (CFA) and reliability testing confirmed that all indicators met standard thresholds (factor loading > 0.50; Cronbach's Alpha > 0.60).

Data analysis was performed using Partial Least Squares Structural Equation Modeling (PLS-SEM) via SmartPLS 4. The analytical procedure followed a two-stage approach: evaluation of the measurement model (outer model) for convergent and discriminant validity, and assessment of the structural model (inner model) for path coefficients, R-squared values, effect sizes (f^2), and predictive relevance (Q^2). Hypotheses were tested using bootstrapping with 5,000 subsamples, applying a one-tailed significance threshold of $p < 0.05$ and t-statistics > 1.645. Mediation effects were examined through specific indirect effect analysis.

RESULTS AND DISCUSSION

The measurement model demonstrated strong psychometric properties. All outer loadings exceeded 0.70, Average Variance Extracted (AVE) values ranged from 0.746 to 0.869 (exceeding the 0.50 threshold), and discriminant validity was confirmed via the Fornell-Larcker criterion and HTMT ratios (< 0.90). Composite reliability and Cronbach's Alpha values for all constructs surpassed 0.70, indicating high internal consistency. The structural model showed substantial explanatory power, with R^2 values of 0.659 for cognitive

trust and 0.789 for purchase intention. Predictive relevance was confirmed ($Q^2 > 0$ for both endogenous constructs), and model fit indices (SRMR = 0.032; NFI = 0.899) indicated good overall fit.

Table 1.
Hypothesis Testing Results and Path Coefficients

Hypothesis	Path Relationship	Original Sample (β)	T-Statistics	P-Value	Conclusion
H1a	Usefulness → Cognitive Trust	0.138	2.499	0.012	Accepted
H1b	Ease of Use → Cognitive Trust	0.157	3.115	0.002	Accepted
H1c	Entertainment → Cognitive Trust	0.169	3.195	0.001	Accepted
H2a	Appearance Anthropomorphism → Cognitive Trust	0.123	3.440	0.001	Accepted
H2b	Behavioral Anthropomorphism → Cognitive Trust	0.147	3.339	0.001	Accepted
H2c	Cognitive Anthropomorphism → Cognitive Trust	0.122	2.660	0.008	Accepted

H2d	Emotional Anthropomorphism → Cognitive Trust	0.193	3.998	0.000	Accepted
H3	Cognitive Trust → Purchase Intention	0.221	3.738	0.000	Accepted
H4a-c	Content Characteristics → Purchase Intention	0.105–0.113	2.568–2.655	0.008–0.010	Accepted
H5a-d	Anthropomorphism Dimensions → Purchase Intention	0.099–0.218	2.496–5.829	0.000–0.013	Accepted

Source: Author's Analysis (2026)

All hypothesized direct effects were statistically significant and positive. Emotional anthropomorphism exerted the strongest influence on cognitive trust ($\beta = 0.193$, $p < 0.001$), indicating that vocal expressions of empathy, enthusiasm, and affective resonance are critical for establishing rational trust in AI voices. Cognitive trust, in turn, strongly predicted purchase intention ($\beta = 0.221$, $p < 0.001$). Furthermore, indirect effect analysis confirmed that cognitive trust significantly mediates the relationships between all independent variables and purchase intention ($p < 0.05$), supporting a partial mediation model.

These findings align with and extend the S-O-R framework and technology acceptance literature. The results demonstrate that AI-generated voices are not merely functional tools but social stimuli that trigger cognitive evaluations. When synthetic voices exhibit human-like emotional and behavioral cues, consumers perceive them as more credible and competent, reducing skepticism toward AI-mediated marketing. The dominance of emotional anthropomorphism corroborates recent studies on affective AI design, which emphasize that vocal prosody and emotional authenticity significantly enhance user engagement and trust (Efthymiou & Hildebrand, 2024; Thimm et al., 2024). Practically, marketers should prioritize AI voice configurations that balance informational clarity with emotional expressiveness, leveraging text-to-speech platforms that allow fine-tuning of pitch, pacing, and affective tone. Integrating entertaining and easily digestible content further amplifies trust formation, ultimately converting viewer attention into purchase intent.

CONCLUSION

This study empirically demonstrates that voice anthropomorphism in Generative AI-based short video content significantly influences consumer purchase intention, with cognitive trust acting as a crucial mediating mechanism. Content characteristics (usefulness, ease of use, entertainment) and all four anthropomorphism dimensions (appearance, behavioral, cognitive, emotional) positively shape cognitive trust, which subsequently drives purchase decisions. Emotional anthropomorphism emerged as the most powerful predictor, underscoring the importance of affective vocal design in AI marketing. Theoretical contributions include extending the S-O-R framework into auditory AI contexts and validating cognitive trust as a rational evaluation pathway in human-AI commercial interactions. Practically, the findings advise digital marketers to optimize AI voice parameters for emotional expressiveness, logical coherence, and content entertainment value to enhance credibility and conversion rates.

Limitations include the predominance of young, digitally active respondents, reliance on self-reported quantitative data, absence of direct comparisons between AI and human voices, and lack of product-category or platform-specific segmentation. Future research should employ mixed-method designs, incorporate affective trust and perceived risk variables, conduct experimental comparisons across voice types, and examine category-specific or platform-specific moderating effects to deepen understanding of consumer-AI voice dynamics in digital commerce.

REFERENCES

- Aquilino, L., di Dio, C., Manzi, F., Massaro, D., Bisconti, P., & Marchetti, A. (2025). Decoding trust in artificial intelligence: A systematic review of quantitative measures and related variables. *Informatics*, 12(3). <https://doi.org/10.3390/informatics12030070>
- Blut, M., Wunderlich, N. V., & Brock, C. (2024). Facilitating retail customers' use of AI-based virtual assistants: A meta-analysis. *Journal of Retailing*, 100(2), 293–315. <https://doi.org/10.1016/j.jretai.2024.04.001>
- Chang, S. H., Chih, W. H., Liou, D. K., & Yang, Y. T. (2016). The mediation of cognitive attitude for online shopping. *Information Technology and People*, 29(3), 618–646. <https://doi.org/10.1108/ITP-08-2014-0172>
- Chang, Y. H., Silalahi, A. D. K., Eunike, I. J., & Riantama, D. (2024). Socio-technical systems and trust transfer in live streaming e-commerce: Analyzing stickiness and purchase intentions with SEM-fsQCA. *Frontiers in Communication*, 9. <https://doi.org/10.3389/fcomm.2024.1305409>
- Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction*, 39(9), 1727–1739. <https://doi.org/10.1080/10447318.2022.2050543>
- Dabiran, E., Farivar, S., Wang, F., & Grant, G. (2024). Virtually human: Anthropomorphism in virtual influencer marketing. *Journal of Retailing and Consumer Services*, 79, 103797. <https://doi.org/10.1016/j.jretconser.2024.103797>
- de Andrés-Sánchez, J., & Gené-Albesa, J. (2023). Explaining policyholders' chatbot acceptance with a unified technology acceptance and use of technology-based model. *Journal of Theoretical and Applied Electronic Commerce Research*, 18(3), 1217–1237. <https://doi.org/10.3390/jtaer18030062>

- Efthymiou, F., & Hildebrand, C. (2024). Empathy by design: The influence of trembling AI voices on prosocial behavior. *IEEE Transactions on Affective Computing*, 15(3), 1253–1263. <https://doi.org/10.1109/TAFFC.2023.3332742>
- ElevenLabs. (2026). *Voice library*. Retrieved from <https://elevenlabs.io/voice-library/asmr>
- Fakhimi, A., Garry, T., & Biggemann, S. (2023). The effects of anthropomorphised virtual conversational assistants on consumer engagement and trust during service encounters. *Australasian Marketing Journal*, 31(4), 314–324. <https://doi.org/10.1177/14413582231181140>
- Fiveable. (2026). *Acoustic and auditory phonetics*. Introduction to Linguistics. Retrieved from <https://fiveable.me/introduction-linguistics/unit-2/acoustic-auditory-phonetics/>
- Goel, P., & Garg, A. (2025). Virtual personalities, real bonds: Anthropomorphised virtual influencers' impact on trust and engagement. *Journal of Consumer Marketing*, 42(4), 431–447. <https://doi.org/10.1108/JCM-05-2024-6915>
- Golossenko, A., Pillai, K. G., & Aroean, L. (2020). Seeing brands as humans: Development and validation of a brand anthropomorphism scale. *International Journal of Research in Marketing*, 37(4), 737–755. <https://doi.org/10.1016/j.ijresmar.2020.02.007>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial least squares structural equation modeling (PLS-SEM) using R*. Springer. <https://doi.org/10.1007/978-3-030-80519-7>
- Isbahi, M. B., Zuana, M. M. M. ., & Mariana, E. R. . (2022). The Technology Strategy in Website Communication Media in Improving Business Activities. *Majapahit Journal of Islamic Finance and Management*, 1(2), 126–138. <https://doi.org/10.31538/mjifm.v1i2.17>
- Ho, J. (2025, March 17). *150+ AI-generated video creation statistics for 2025*. Zebracat Blog. <https://www.zebracat.ai/post/ai-video-creation-statistics>
- Hsu, W. C., & Lee, M. H. (2023). Semantic technology and anthropomorphism: Exploring the impacts of voice assistant personality on user trust, perceived risk, and attitude. *Journal of Global Information Management*, 31(1). <https://doi.org/10.4018/JGIM.318661>
- Huang, Y., & Yu, Z. (2023). Understanding the continuance intention for artificial intelligence news anchor: Based on the expectation confirmation theory. *Systems*, 11(9), 438. <https://doi.org/10.3390/systems11090438>
- Kane, J., Johnstone, M. N., & Szewczyk, P. (2024). Voice synthesis improvement by machine learning of natural prosody. *Sensors*, 24(5), 1624. <https://doi.org/10.3390/s24051624>
- Knödler, L., & Rudeloff, C. (2024). Look who's talking now: The effects of pre-recorded and AI-generated synthetic brand voices on brand anthropomorphism and brand equity. *Journal of Creative Communications*, 19(3), 295–311. <https://doi.org/10.1177/09732586241253651>
- Legood, A., van der Werff, L., Lee, A., & den Hartog, D. (2021). A meta-analysis of the role of trust in the leadership-performance relationship. *European Journal of Work and*

- Organizational Psychology*, 30(1), 1–22.
<https://doi.org/10.1080/1359432X.2020.1819241>
- Li, C., & Huang, F. (2024). The impact of virtual streamer anthropomorphism on consumer purchase intention: Cognitive trust as a mediator. *Behavioral Sciences*, 14(12), 1228. <https://doi.org/10.3390/bs14121228>
- Li, H., Li, W., & Ma, T. (2025). Exploring the mechanism of AI-powered virtual idols' intelligence level on digital natives' impulsive buying intention in e-commerce live streaming: A perspective of psychological distance. *Journal of Theoretical and Applied Electronic Commerce Research*, 20(3), 173. <https://doi.org/10.3390/jtaer20030173>
- Li, W., Cujilema, S., Hu, L., & Xie, G. (2025). How social scene characteristics affect customers' purchase intention: The role of trust and privacy concerns in live streaming commerce. *Journal of Theoretical and Applied Electronic Commerce Research*, 20(2), 85. <https://doi.org/10.3390/jtaer20020085>
- Luo, C., Hasan, N. A. M., Zamri bin Ahmad, A. M., & Lei, G. (2025). Influence of short video content on consumers purchase intentions on social media platforms with trust as a mediator. *Scientific Reports*, 15(1), 94994. <https://doi.org/10.1038/s41598-025-94994-z>
- Malhotra, N. K., Nunan, D., & Birks, D. F. (2017). *Marketing research: An applied approach* (5th ed.). Pearson.
- Malik, N. I., Ramzan, M. M., Malik, Z., & Tariq, I. (2025). The rise of reels: Analyzing the communicative power of short-form videos on social media. *Qlantic Journal of Social Sciences*, 6(2), 138–145. <https://doi.org/10.55737/qjss.vi-ii.25350>
- Microsoft. (2022, May 9). *Azure Neural TTS releases 5 new voices and expands emotions in American English*. Microsoft Tech Community. <https://techcommunity.microsoft.com/blog/azure-ai-foundry-blog/azure-neural-tts-releases-5-new-voices-and-expands-emotions-in-american-english/3298435>
- Microsoft. (2026). *Text-to-speech overview – Speech service*. Azure AI Services. <https://docs.azure.cn/en-us/ai-services/speech-service/text-to-speech>
- Ozdemir, S., Zhang, S. J., Gupta, S., & Bebek, G. (2020). The effects of trust and peer influence on corporate brand—Consumer relationships and consumer loyalty. *Journal of Business Research*, 117, 791–805. <https://doi.org/10.1016/j.jbusres.2020.02.027>
- Saffarizadeh, K., Keil, M., Boodraj, M., & Alashoor, T. (2024). "My name is Alexa. What's your name?" The impact of reciprocal self-disclosure on post-interaction trust in conversational agents. *Journal of the Association for Information Systems*, 25(3), 528–568. <https://doi.org/10.17705/1jais.00839>
- Schreibelmayr, S., & Mara, M. (2022). Robot voices in daily life: Vocal human-likeness and application context as determinants of user acceptance. *Frontiers in Psychology*, 13, 787499. <https://doi.org/10.3389/fpsyg.2022.787499>
- The Voice Foundation. (2022). *Voice anatomy & physiology: Understanding voice production*. Retrieved from <https://voicefoundation.org/health-science/voice-disorders/anatomy-physiology-of-voice-production/>
- Thimm, C., Engelhardt, P., Nehls, P., Hens, J., Biehler, M., Hermanns, C., Huzel, P., Bennewitz, M., Mücke, J., & Dengler, N. (2024). Trust, (dis)comfort, and voice quality: Anthropomorphism in verbal interactions with NAO. *Communications in*

- Computer and Information Science*, 2114, 405–412. https://doi.org/10.1007/978-3-031-61932-8_46
- Xie, Q., Mahomed, A. S. B., Mohamed, R., & Subramaniam, A. (2023). Investigating the relationship between usefulness and ease of use of live streaming with purchase intentions. *Current Psychology*, 42(30), 26464–26476. <https://doi.org/10.1007/s12144-022-03698-4>
- Zhang, Z. (2016). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, 140(4), 2614–2635. <https://doi.org/10.1121/1.4964509>