

From Intention To Actual Use: Explaining The Adoption Of Generative-AI Whatsapp Chatbots For Financial And Health Information With An Extended UTAUT2 Model And Behavioral Log Data

Indri Anjar Kartika Sari¹, Agung Budi Prasetyo², Hudan Studiawan³, Onno W. Purbo³, Eldiana Tri Narulita³

¹ Universitas Pembangunan Nasional "Veteran"; indri.cs2026@ids.ac.id

² Institut Teknologi Tangerang Selatan; agung@itts.ac.id

³ Institut Teknologi Sepuluh Nopember; hudan@its.ac.id

⁴ Institut Teknologi Tangerang Selatan; onno@indo.net.id

⁵ Institut Seni Indonesia Bali; eldiana.narulita@isi-dps.ac.id

ARTICLE INFO

Keywords:

Keyword 1; Conversational Agents

Keyword 2; Generative AI

Keyword 3; Intention

Keyword 4; Behavior Gap

Article history:

Received 2026-09-30

Revised 2026-10-02

Accepted 2026-12-01

ABSTRACT

Generative artificial intelligence is increasingly embedded in familiar messaging platforms, expanding access to intelligent information services. This study develops an extended UTAUT2 model to examine adoption of an AI-enabled WhatsApp bot integrating deterministic information utilities with a Gemini-powered stock analysis feature. The model distinguishes behavioural intention from actual use and adds trust in AI-generated information, perceived risk, and perceived information accuracy. Its main methodological contribution is linking survey-based perceptions with pseudonymised behavioural records, enabling direct assessment of intention behaviour correspondence without treating self-reported use as observed interaction. The open-source artifact, wa bot quote, provides cryptocurrency, stock, health and fitness, football, QR-code, and other utilities, while its repository documents Gemini integration for stock analysis. The manuscript therefore separates capabilities verified from the repository from outcomes requiring future empirical deployment. It specifies a reproducible protocol covering informed consent, data minimisation, secure logging, measurement validation, partial least squares structural equation modelling, measurement invariance, multi-group analysis, predictive assessment, and robustness testing. By integrating perceptions with behavioural evidence, the framework supports transparent evaluation of technology acceptance, user-group differences, and discrepancies between stated intentions and actual interactions. Accordingly, the manuscript constitutes a rigorous empirical research framework rather than a completed results paper and avoids presenting uncollected evidence or fabricated findings.

This is an open access article under the [CC BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



Corresponding Author:

Indri Anjar Kartika Sari

Universitas Pembangunan Nasional "Veteran"; indri.cs2026@ids.ac.id

1. INTRODUCTION

Generative AI assistants are moving into the messaging applications that people already use every day. WhatsApp reaches roughly nine in ten Indonesians online every month, users spend close to two hours a day in the app, and more than a third of Indonesians have begun using ChatGPT monthly (We Are Social, 2025). These two habits are now converging: assistants answer questions inside the same chat threads where people talk to family, colleagues and sellers, with no new app to install and no interface to learn. With 212 million internet users (Kemp, 2025), Indonesia is one of the largest settings in which this convergence is unfolding.

Convenience, however, is not the whole story, because the greatest challenge of generative AI arises when people ask for decision-relevant information. When the question is "Should I buy this stock?" or "How many calories should I eat?", a fluent but wrong answer carries a financial or bodily cost. Large language models are known to generate plausible yet factually incorrect content (Ji et al., 2023), and scholars have warned that generative conversational AI raises new questions of accuracy, accountability and trust (Dwivedi et al., 2023). In high-stakes domains, whether people adopt an AI assistant may therefore depend less on how easy it is to use and more on whether they trust what it says (Glikson & Woolley, 2020).

This raises a critical gap between what is known about chatbot acceptance and what is needed to understand these assistants. Firms have long used chatbots to serve customers, and research on customer acceptance of these agents is well established (Sheehan et al., 2020; Yatawara et al., 2025). Studies grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT) show that performance expectancy, social influence, trust and privacy concerns shape intentions to use ChatGPT and similar chatbots (Joshi, 2025; Menon & Shilpa, 2023; Shahzad et al., 2024). Yet the approach taken is often partial: almost all studies measure behavioral intention through a cross-sectional survey and treat it as a proxy for use, although intention explains only part of the variance in behavior (Sheeran, 2002) and habit can override intention once use becomes routine (Limayem et al., 2007). Whether stated intention predicts what people actually do with an AI chatbot remains largely untested.

The need to examine this question is reinforced by limits in the theory and in the evidence. The extended UTAUT (UTAUT2; Venkatesh et al., 2012) predates generative AI and contains no construct for the accuracy of machine-generated information or for trust in it. Customer-chatbot studies usually examine a firm's support channel rather than an assistant that supplies decision-relevant information. Moreover, studies typically examine one system in one domain, so differences between domains are confounded with differences between systems, and samples are dominated by university students.

The novelty of this study lies in linking perceptions to observed behavior within a single deployed system. We built and deployed a WhatsApp chatbot that serves financial information (stock and cryptocurrency prices), health information (body mass index, calorie and nutrition tools) and free-text answers from a large language model. Over [X] weeks we recorded pseudonymised usage logs and then surveyed consenting users, which allows survey perceptions to be matched to logged use and allows the financial and health domains to be compared within the same system.

This study aims to explain what drives the intention to use a generative-AI WhatsApp chatbot for financial and health information, and to test whether stated intention translates into actual use. It is guided by three questions. First, which factors shape the intention to use the chatbot for financial and health information (RQ1)? Second, to what extent does intention predict actual use recorded in

system logs (RQ2)? Third, do these effects differ between the financial and health domains and across age groups (RQ3)? The main argument is that understanding the adoption of AI information services requires looking beyond what users say they intend to do, to what they actually do, and to whether they trust what the machine tells them.

The study makes three contributions. Theoretically, it extends UTAUT2 with trust in AI-generated information, perceived risk and perceived information accuracy, adapting the model to generative AI. Methodologically, it tests the intention-behavior link with objective log data rather than self-reported use, which also reduces common method bias. Practically, it offers evidence on how to design AI information services for messaging platforms in emerging markets. Section 2 reviews the theoretical background, Section 3 develops the hypotheses, Section 4 describes the method, and Sections 5 to 7 report and discuss the results.

2. METHODS

This study employed a quantitative, prediction-oriented design to examine the adoption and actual use of an AI-enabled WhatsApp information bot using an extended UTAUT2 model. The open-source Node.js chatbot provided financial information, health utilities, and free-text responses generated through Google Gemini, with appropriate investment and medical disclaimers. The research comprised system enhancement, public deployment, questionnaire administration, and statistical analysis. A consent mechanism and privacy-preserving logging module recorded only salted SHA-256 user hashes, timestamps, and feature categories without retaining health-message content. Eligible users completed a seven-point Likert-scale questionnaire measuring UTAUT2 constructs, trust in AI-generated information, perceived risk, perceived information accuracy, and behavioural intention, while actual use was derived from interaction frequency, active days, sessions, feature breadth, and retention. An a priori G*Power analysis indicated a minimum sample of approximately 160 respondents, although at least 300 participants were targeted to support financial-versus-health multi-group analysis. The questionnaire underwent translation, back-translation, expert validation, and pilot testing with 30 users. Data were analysed using PLS-SEM in SmartPLS 4 by evaluating indicator loadings, internal consistency, convergent and discriminant validity, collinearity, path coefficients based on 5,000 bootstrap subsamples, explanatory power, effect sizes, and PLSpredict performance. Measurement invariance was assessed through MICOM before multi-group analysis, while age was examined as a continuous moderator. Informed consent, voluntary participation, data minimisation, pseudonymisation, restricted data linkage, and post-analysis deletion were implemented in accordance with Indonesia's Personal Data Protection Law.

3. FINDINGS AND DISCUSSION

The empirical results are reported in three sequential stages: assessment of the sample profile and potential common method bias, evaluation of the measurement model, and examination of the structural model. This sequence permits the adequacy of the data and measures to be established before the hypothesised relationships are interpreted. The analysis further distinguishes behavioural intention from actual use, with the latter derived from pseudonymised behavioural logs rather than self-reported usage.

Table 4 summarises the demographic and technology-use characteristics of the analytic sample, including gender, age, education, prior experience with AI chatbots, and the dominant information domain observed in the logs. These characteristics provide the descriptive context required to assess the composition of the sample and to interpret subsequent estimates. Common method bias should be evaluated using the prespecified diagnostic procedure and reported alongside the completed sample profile.

Table 4. Sample profile (n = []).

Characteristic	Category	n	%
Gender	Female / Male	[]	[]
Age (years)	18–24 / 25–34 / 35–44 / ≥ 45	[]	[]
Education	Secondary / Bachelor / Postgraduate	[]	[]
Prior AI chatbot use	Never / Occasional / Regular	[]	[]
Dominant domain (logs)	Financial / Health	[]	[]

Table 5. Reliability and convergent validity.

Construct	Loadings	α	c	AVE
PE	[]	[]	]	[]
EE	[]	[]		[]
SI	[]	[]		[]
FC	[]	[]		[]
HM	[]	[]		[]
HT	[]	[]		[]
TR	[]	[]		[]
PR	[]	[]		[]
PIA	[]	[]		[]
BI	[]	[]		[]

Note. α = Cronbach's alpha; qc = composite reliability; AVE = average variance extracted. [Add Table 6 (HTMT matrix) and formative indicator weights and VIFs for actual use.]

The measurement model evaluates indicator reliability, internal consistency, convergent validity, and discriminant validity for performance expectancy (PE), effort expectancy (EE), social influence (SI), facilitating conditions (FC), hedonic motivation (HM), habit (HT), trust in AI-generated information (TR), perceived risk (PR), perceived information accuracy (PIA), and behavioural intention (BI). As reported in Table 5, the final manuscript should present the retained indicator loadings, Cronbach's alpha, composite reliability, and average variance extracted for each reflective construct. The accompanying HTMT matrix and the weights and variance inflation factors of the formative actual-use indicators are required to complete the assessment of construct distinctiveness and multicollinearity.

The structural model tests the proposed effects of the UTAUT2 determinants and the three AI-specific extensions on behavioural intention, trust, and objectively recorded actual use. Table 7 reports the standardised path coefficients (β), bootstrap t statistics, p values, effect sizes (f^2), and decisions for H1-H11. Statistical inference is based on 5,000 bootstrap subsamples and one-tailed

tests for directional hypotheses; explanatory and predictive performance should be reported using R^2 for BI, TR, and AU and $Q^2_{predict}$ for the relevant endogenous constructs.

Table 7. Results of hypothesis testing.

Hypotheses	Path	β	t	p	f_2	Supported
H1	PE → BI	[]	[]	[]	[]	[]
H2	EE → BI	[]	[]	[]	[]	[]
H3	SI → BI	[]	[]	[]	[]	[]
H4	FC → BI	[]	[]	[]	[]	[]
H5	HM → BI	[]	[]	[]	[]	[]
H6	HT → BI	[]	[]	[]	[]	[]
H7	TR → BI	[]	[]	[]	[]	[]
H8	PR → BI	[]	[]	[]	[]	[]
H9	PIA → TR	[]	[]	[]	[]	[]
H10	BI → AU	[]	[]	[]	[]	[]
H11	HT → AU	[]	[]	[]	[]	[]

Note. Bootstrapping with 5,000 subsamples; one-tailed tests for directional hypotheses. [Report R^2 for BI, TR and AU and $Q^2_{predict}$.]

DISCUSSION

This study advances technology-acceptance theory in three respects. First, it extends UTAUT2 to the context of generative AI by modelling trust in AI-generated information, perceived risk, and perceived information accuracy alongside the established determinants [state what the results show]. Second, the use of behavioural logs provides a comparatively rare test of the relationship between behavioural intention and actual AI-chatbot use, while reducing reliance on a single self-reported data source and addressing longstanding concerns regarding common method variance (Podsakoff et al., 2003). Third, comparing financial and health information use within the same technological system helps distinguish domain-specific influences from system-level effects, thereby clarifying the boundary conditions under which acceptance models explain the adoption of AI-enabled messaging services.

The findings also have direct implications for organisations that deliver customer services through WhatsApp, including banks, insurers, retailers, and clinics, as well as for developers of AI services embedded in messaging platforms. Subject to completion of the empirical estimates [adapt to results], the model indicates four design priorities: displaying data sources and timestamps for financial responses to reinforce perceived accuracy and trust; developing concise, repeatable commands that facilitate habit formation rather than depending on novelty; implementing stronger safeguards and clearer disclaimers for health-related information; and communicating privacy

protections explicitly because perceived risk may inhibit adoption. At the policy level, the study underscores the need for regulatory guidance governing AI-generated financial and health information disseminated through consumer messaging channels.

Several limitations delimit the interpretation and generalisability of the findings. Recruitment through online communities may disproportionately represent users with relatively high levels of digital literacy, while a deployment period of [X] weeks captures initial adoption more effectively than sustained or routinised use. In addition, the behavioural logs originate from a single non-commercial chatbot; consequently, the findings may not transfer directly to branded services with broader functionality, different governance arrangements, or stronger institutional reputations. Future research should employ longitudinal designs to examine continuance, experimentally test the effects of source transparency, and replicate the extended model across other emerging-market settings and service domains.

4. CONCLUSION

This study advances understanding of generative-AI chatbot adoption by extending UTAUT2 to incorporate trust in AI-generated information, perceived risk, and perceived information accuracy while distinguishing behavioural intention from objectively recorded actual use. By linking survey responses with pseudonymised behavioural logs, the study provides a more robust basis for examining the intention behaviour relationship than approaches relying exclusively on self-reported usage. The comparison of financial and health information services within a single WhatsApp-based system further clarifies how domain-specific sensitivities may shape user acceptance, trust, and continued interaction. The study consequently offers theoretical value by refining the explanatory boundaries of UTAUT2 in generative-AI environments and practical value by highlighting the importance of transparent information sources, perceived accuracy, privacy protection, appropriate safeguards, and interaction designs that encourage habitual use. Nevertheless, the interpretation and generalisability of the findings should account for the characteristics of the recruited sample, the limited deployment period, and the use of a single non-commercial chatbot. Future research should therefore employ longitudinal and experimental designs, examine continued use across broader populations, and replicate the model in different institutional, cultural, and technological contexts.

REFERENCES

- Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, 100006.
- Alt, M.-A., Vizeli, I., & Săplăcan, Z. (2021). Banking with a chatbot – A study on technology acceptance. *Studia Universitatis Babeş-Bolyai Oeconomica*, 66(1), 13–35. <https://doi.org/10.2478/subboec-2021-0002>
- Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In I. Kompatsiaris et al. (Eds.), *Internet Science: INSCI 2017 (Lecture Notes in Computer Science, Vol. 10673, pp. 377–392)*. Springer.
- DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems*, 19(4), 9–30.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160.
- Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets

- perspective. *International Journal of Human-Computer Studies*, 59(4), 451–474.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51–90.
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- Hsu, C.-L., & Lin, J. C.-C. (2023). Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services*, 71, 103211. <https://doi.org/10.1016/j.jretconser.2022.103211>
- Huang, D., Markovitch, D. G., & Stough, R. A. (2024). Can chatbot customer service match human service agents on customer satisfaction? An investigation in the role of trust. *Journal of Retailing and Consumer Services*, 76, 103600. <https://doi.org/10.1016/j.jretconser.2023.103600>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1–38.
- Joshi, H. (2025). Integrating trust and satisfaction into the UTAUT model to predict chatbot adoption – A comparison between Gen-Z and Millennials. *International Journal of Information Management Data Insights*, <https://doi.org/10.1016/j.ijime.2025.100332>
- Limayem, M., Hirt, S. G., & Cheung, C. M. K. (2007). How habit limits the predictive power of intention: The case of information systems continuance. *MIS Quarterly*, 31(4), 705–737.
- Menon, D., & Shilpa, K. (2023). "Chatting with ChatGPT": Analyzing the factors influencing users' intention to use the Open AI's ChatGPT using the UTAUT model. *Heliyon*, 9(11), e20962. <https://doi.org/10.1016/j.heliyon.2023.e20962>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903.
- Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research*, 115, 14–24. <https://doi.org/10.1016/j.jbusres.2020.04.030>
- Sheeran, P. (2002). Intention–behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1–36.
- Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, 53(11), 2322–2347.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.
- Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178.
- We Are Social. (2025, November). Digital 2026: Top digital and social media trends in Indonesia. <https://wearesocial.com/id/blog/2025/11/digital-2026-top-digital-and-social-media-trends-in-indonesia/>
- Yatawara, K., Sampath, T., Kalupahana, P. L., Rathnayake, S., Jayasuriya, N., & Rathnayake, N. (2025). A systematic review on consumer adoption of AI-driven chatbots. *Vision*. Advance online publication. <https://doi.org/10.1177/09722629251332349>