

VALIDATED INSTRUMENT FOR INSURANCE CHATBOT ACCEPTANCE

Dana Jašková¹, Katarína Kráľová²

¹ *Asst. Prof., Faculty of Social and Economic Relations, A. Dubcek University of Trencin, Studentska 2, Trencin, Slovakia, E-mail address: dana.jaskova@tmuni.sk*

² *Asst. Prof., Faculty of Social and Economic Relations, A. Dubcek University of Trencin, Studentska 2, Trencin, Slovakia, E-mail address: katarina.kralova@tmuni.sk*

Received 01 06 2026; Accepted 21 06 2026

Abstract

The rapid digital transformation of the insurance industry and the growing implementation of artificial intelligence have significantly increased the use of chatbots in customer communication. However, despite their expanding application, there is still a lack of validated instruments specifically designed to measure customer acceptance of insurance chatbots. Therefore, this study aimed to develop and content-validate a measurement instrument for assessing chatbot acceptance in the insurance sector. The instrument was developed based on the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) and included five dimensions: perceived usefulness, perceived ease of use, privacy concern, trust, and behavioural intention. Content validity was evaluated by an expert panel using the Content Validity Index (CVI), Content Validity Ratio (CVR), modified kappa coefficient, Aiken's V, and Kendall's coefficient of concordance. The results confirmed high content validity across all dimensions and statistically significant agreement among experts. The proposed instrument fills an important methodological gap in insurance research by providing a theoretically grounded and content-validated tool suitable for subsequent psychometric validation and practical application in evaluating customer acceptance of AI-supported communication.

Keywords: *Artificial Intelligence, Chatbot Acceptance, Content Validity, Insurance Sector, Measurement Instrument, TAM; UTAUT.*

JEL Code: *C18, C83, M15, G22.*

Introduction

The rapid development of artificial intelligence (AI) significantly transforms communication between organizations and customers across various sectors, including the insurance industry. Chatbots increasingly represent an important tool of digital communication that improves service availability, response speed, and operational efficiency. Despite the growing implementation of chatbot technologies, their successful application largely depends on customer acceptance and willingness to use AI-supported communication systems. Research on technology acceptance is commonly based on established theoretical frameworks, particularly the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT). These models explain how users perceive and adopt new technologies. However, although chatbot adoption has attracted increasing research attention, there is still a lack of validated

measurement instruments specifically designed for the insurance sector. Existing instruments often insufficiently reflect sector-specific issues such as trust, privacy concerns, and acceptance of automated communication. The identified research gap highlights the need to develop a context-specific, methodologically validated instrument to measure customer acceptance of chatbots in the insurance sector. The research problem lies in the absence of a content-validated instrument specifically designed to assess customer acceptance of chatbots in insurance services. Therefore, this study aims to develop and content-validate a measurement instrument for assessing customer acceptance of chatbots in the insurance sector. The research focuses on customer interactions with insurance chatbots, with the research subject comprising perceived usefulness, perceived ease of use, privacy concerns, trust, and behavioural intention regarding chatbot usage. The study contributes to

technology acceptance research by providing a theoretically grounded and practically applicable instrument suitable for future empirical research and digital communication management in insurance organizations.

Although numerous studies have investigated chatbot acceptance in banking, e-commerce, healthcare, and general online services, considerably less attention has been devoted to the insurance sector. Insurance services differ from many other digital service environments because they involve highly sensitive personal information, complex contractual relationships, and increased requirements for customer trust and data protection. Consequently, measurement instruments developed for general digital services may not adequately capture the determinants influencing chatbot acceptance in insurance. The present study therefore addresses this methodological gap by developing a context-specific measurement instrument whose content validity is verified through a comprehensive expert evaluation process. By providing a validated instrument, the study creates a foundation for future psychometric validation and empirical investigations of chatbot acceptance in insurance organisations.

Literature review

The rapid development of digital technologies and artificial intelligence (AI) has significantly transformed communication between organizations and customers, including within the insurance sector. Insurance companies increasingly implement AI-based technological solutions to improve operational efficiency, customer experience, and service accessibility. Among the most important digital communication tools are chatbots, which enable automated customer interaction, real-time information provision, and personalized customer services. Insurance organizations also intensively use digital platforms to support online communication, product distribution, and customer support services.

Despite the growing implementation of chatbots, their adoption in the insurance sector is more complex compared with many other industries, mainly due to regulatory requirements and the high sensitivity of processed customer

data (Gebert-Persson et al., 2019). Insurance services belong to data-intensive sectors in which privacy protection, data security, and customer trust represent critical determinants of successful digital technology implementation. Insufficient protection of personal data may lead to financial, psychological, or reputational consequences for customers, increasing the importance of trust and privacy concerns in chatbot adoption (Przegalinska et al., 2019).

Trust represents one of the most important factors influencing interaction between humans and AI-supported technologies. It is considered a necessary condition for effective communication between users and AI systems (Yang & Wibowo, 2022). In the insurance industry, chatbots are primarily used for customer service automation, information provision, and service personalization. Their implementation may improve customer satisfaction, increase process efficiency, and reduce customer complaints (Yang & Zhou, 2021). However, the successful implementation of chatbot technologies largely depends on customer acceptance and willingness to use such systems.

Technology acceptance represents an important research area within information systems research. One of the most frequently applied theoretical frameworks is the Technology Acceptance Model (TAM), which emphasizes the role of perceived usefulness and perceived ease of use in shaping behavioural intention toward technology usage. Previous studies focused on chatbots confirm that these constructs are among the most important determinants of chatbot acceptance (Andrés-Sánchez & Gené-Albesa, 2023a). In the insurance context, customers positively evaluate rapid responses, continuous service availability, and the ability of chatbots to solve simple customer requests without direct interaction with human employees.

In addition to TAM, more comprehensive frameworks such as the Unified Theory of Acceptance and Use of Technology (UTAUT) are also commonly applied in chatbot acceptance research. Previous studies indicate that trust, satisfaction, and prior user experience significantly influence chatbot acceptance in the insurance sector (Andrés-Sánchez & Gené-Albesa, 2023b). Emotional factors, including communication comfort and perceived

naturalness of interaction, also play an important role in shaping customer attitudes toward chatbot technologies.

Another important determinant of chatbot acceptance is communication personalization. Chatbots capable of adapting communication to individual customer needs achieve higher levels of user acceptance (Andrés-Sánchez & Gené-Albesa, 2024). However, increased personalization simultaneously increases requirements for secure data processing, algorithm transparency, and trust building.

Recent research has also focused on chatbot anthropomorphism. Anthropomorphism refers to the attribution of human characteristics to technological systems in order to increase the naturalness of interaction between humans and technology (Pfeuffer et al., 2019). Studies suggest that anthropomorphic elements may support more natural communication and increase users' willingness to interact with chatbots (Salles et al., 2020). At the same time, communication quality, response speed, and user experience significantly influence customer satisfaction and future behavioural intention toward chatbot usage (Rese et al., 2020).

An important research area also concerns chatbot usability evaluation. For this purpose, standardized instruments such as the Chatbot Usability Scale (BUS-15) were developed to evaluate interaction quality and user satisfaction (Borsci et al., 2022). However, most existing instruments were designed for general digital service environments and insufficiently reflect the specific characteristics of the insurance sector. The development of large language models and advanced AI systems has further expanded chatbot capabilities and enabled more sophisticated forms of communication. However, these technological innovations also introduce new challenges related to algorithm transparency, ethics, and user trust (Nov et al., 2023). As chatbot systems become increasingly complex, the need for robust and validated instruments measuring chatbot acceptance continues to increase.

Despite the growing number of studies focused on chatbot acceptance, there remains a lack of validated measurement instruments specifically designed for the insurance sector. Most existing studies focus on general digital services or other economic sectors, limiting the

applicability of their findings in insurance services. Therefore, the present study addresses the identified research gap through the development and content validation of a measurement instrument focused on customer acceptance of chatbots in the insurance industry.

Research Methods

The development and application of measurement instruments designed to assess complex social and behavioural constructs require systematic verification of their psychometric properties, particularly validity and reliability. The measurement of abstract constructs requires their operationalization through appropriately formulated items and scales that adequately represent the examined concept. To ensure the correct interpretation of results, measurement instruments should be developed through a methodologically rigorous process and subsequently subjected to psychometric evaluation.

In research on technology acceptance, it is necessary to assess multiple aspects of user perception that may influence the adoption of new technologies. One of the fundamental stages in the development of a new measurement instrument is the assessment of content validity, which reflects the extent to which the instrument items adequately and representatively cover the theoretical construct being measured. Content validity is based on expert evaluation conducted before pilot or main data collection. In classical psychometric theory, content validity represents the degree to which individual items reflect all relevant dimensions of the examined construct (Lynn, 1986; Lawshe, 1975). Unlike statistical validation approaches such as exploratory factor analysis or structural equation modelling, content validity represents a conceptual and expert-based form of instrument validation. Expert review is a methodological procedure in which a panel of specialists evaluates individual items in terms of relevance, clarity, comprehensibility, necessity, and representativeness in relation to the defined construct. The results of expert evaluation are subsequently quantified using content validity indices.

The most frequently applied indicators of content validity are the Content Validity Ratio

(CVR) and the Content Validity Index (CVI), each assessing different aspects of validity. The CVR evaluates the degree of agreement among experts regarding the necessity of individual items for measuring the examined construct. Experts typically assess each item using a three-point scale: “essential”, “useful but not essential”, and “not essential”. The CVR is calculated as: $CVR = (n_e - N/2) / [N/2]$, where n_e represents the number of experts who identified the item as essential and N is the total number of experts. The obtained CVR values are calculated for each item separately and subsequently compared with critical values recommended according to the number of experts (Lawshe, 1975).

The most used indicator for quantifying content validity is the Content Validity Index (CVI). Experts evaluate the relevance of individual items using a four-point ordinal scale ranging from 1 (“not relevant”) to 4 (“highly relevant”). The use of a four-point scale without a neutral midpoint minimizes the tendency toward neutral responses and supports more explicit expert judgments.

Content validity is evaluated both at the item level as the Item-Level Content Validity Index (I-CVI) and at the scale level as the Scale-Level Content Validity Index (S-CVI). The I-CVI represents the proportion of experts who rated the item as relevant or highly relevant (ratings 3 or 4) relative to the total number of experts. Polit et al. (2007) recommend a minimum acceptable I-CVI value of 0.78 when the number of experts exceeds five. The content validity of subdimensions and the entire instrument is subsequently assessed using S-CVI/Ave and S-CVI/UA indicators.

Recent recommendations suggest complementing traditional content validity indices with additional measures reflecting the probability of chance agreement among experts. The most frequently applied indicators include the modified Cohen’s kappa coefficient (κ^*), Aiken’s V coefficient, and Kendall’s coefficient of concordance (W) (Polit et al., 2007). The modified kappa coefficient corrects the level of expert agreement for random agreement and provides a more accurate estimate of expert consensus. The coefficient is calculated as: $\kappa^* = (I - CVI - P_c) / (1 - P_c)$, where P_c represents the probability of chance agreement calculated as: $P_c = [N! / A! (N - A)!] 0,5^N$, where N is the

number of experts, and A represents the number of experts who rated the item as relevant. Values of $\kappa^* < 0.40$ indicate poor agreement, values between 0.40 and 0.59 indicate moderate agreement, values between 0.60 and 0.74 indicate good agreement, and values above 0.75 indicate excellent agreement among experts.

Aiken’s V coefficient was additionally applied as a supplementary indicator of content validity because it reflects the ordinal nature of expert ratings and the intensity of expert agreement. The coefficient is calculated as: $V = \sum(r_i - 1) / (3N)$, where r_i represents the rating assigned by each expert, and N is the total number of experts. Values of $V \geq 0,80$ indicate high content validity, values between 0.70 and 0.79 indicate acceptable validity, and values below 0.70 indicate insufficient content validity. To assess the overall level of agreement among experts, Kendall’s coefficient of concordance (W) was employed. This coefficient evaluates the consistency of ratings provided by multiple independent experts assessing the same set of items. The purpose of this coefficient is not to evaluate the quality of individual items but rather to determine the level of consensus within the expert panel.

It should be emphasized that content validity assessment does not verify the factor structure, reliability, or construct validity of the instrument. Therefore, content validity represents the initial stage of the validation process and should be followed by further psychometric analyses focused on reliability assessment, factor structure evaluation, and construct validity verification of the proposed measurement instrument.

Research Results and Discussion

The development of the measurement instrument followed a multi-stage procedure commonly applied in instrument design. The process began with the development stage, during which the key factors influencing the adoption of new technologies were identified based on an extensive review of relevant theoretical models and empirical studies. Attention was paid to constructs derived from established technology acceptance frameworks, such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology

(UTAUT). Based on this conceptual framework, the identified constructs and their respective dimensions were operationalized into measurable questionnaire items, resulting in the preliminary version of the instrument. In the subsequent stage, the content validity of the instrument was assessed through expert evaluation. The expert validation was conducted between January and March 2026. Experts were selected using purposive sampling based on their academic and professional experience in psychology, insurance, and artificial intelligence. The minimum inclusion criterion was five years of professional experience in the relevant field. The expert panel consisted of specialists in psychology, insurance, and AI-related information technologies. Seven experts

participated in the first evaluation round, and six experts participated in the second round. The experts reviewed the items and evaluated their relevance and representativeness with respect to the defined construct domains. The purpose of this assessment was to determine whether the proposed items adequately capture the theoretical constructs and sufficiently represent their conceptual dimensions. The conceptual model proposed for assessing the acceptance of chatbots in customer–insurance company interactions is illustrated in Figure 1. The model integrates constructs derived from TAM and UTAUT and serves as the theoretical basis for the development of the measurement instrument.

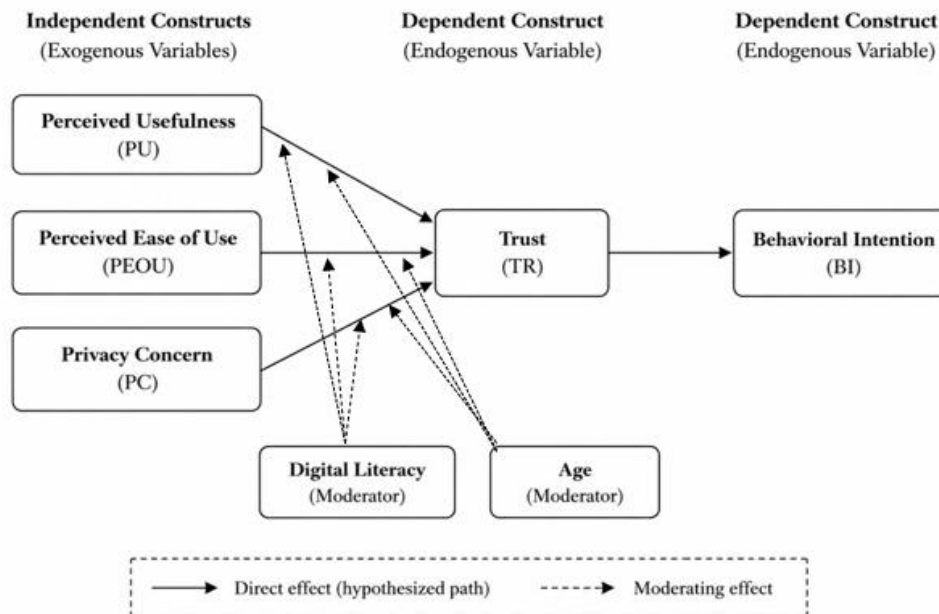


Figure 1. Conceptual model of research

*Source: own elaboration.

Figure 1 presents the conceptual framework of the study and illustrates the relationships among the five dimensions used to assess customer acceptance of insurance chatbots.

Content validation was conducted using a systematic and comprehensive two-stage approach. The use of a single-stage process (either only the development stage or only the evaluation stage) is generally insufficient for assessing content validity and may result in a low-quality construct. The two-stage content validity process

includes the instrument development stage and the judgment–quantification stage (Almanasreh et al., 2022). These two stages are therefore considered essential in the development of instruments designed to measure complex phenomena and constructs. However, the judgment–quantification stage should be followed by a third stage involving revision and reconstruction/reformulation of the instrument, reflecting the results of expert evaluation and ensuring further improvement of the instrument.

The proposed measurement instrument included five domains: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Privacy Concern (PC), Trust (TR), and Behavioural Intention (BI). Each domain initially contained eight items, including several reverse-coded statements. Respondents evaluate the items using a five-point Likert scale. The development stage focused on identifying the construct domains, generating questionnaire items, and designing the instrument structure. Content validity assessment examined the relevance, representativeness, and conceptual clarity of individual items in relation to the examined constructs. Since the instrument was designed to measure affective perceptions related to chatbot acceptance, particular emphasis was placed on the conceptual definition of domains, representation of subdimensions, and relevance of item formulation.

The judgment–quantification stage involved expert evaluation of the extent to which

the proposed items adequately represented the examined construct domains. The expert panel consisted of specialists from psychology, insurance, and AI-related information technologies. Seven experts participated in the first evaluation round and six experts in the second round. Following the initial assessment, the instrument was revised based on item scores and content validity indices. Items with insufficient validity values or redundant content were removed or reformulated according to expert recommendations. The final version of the instrument consisted of 25 items grouped into five domains and was subsequently re-evaluated by the experts. Each item was assessed from two perspectives: relevance of the item to the corresponding domain and necessity of the item for measuring the construct. Relevance ratings were used for calculating CVI indices, while necessity ratings served as the basis for CVR calculation.

Table 1. Content Validity of the Instrument: CV Coefficients

Dimension	Item	Item score			Dim	
		Relevancy	Necessary	I-CVI	S-CVI/Ave	CVR
PU - Perceived Usefulness	PU_1	3,83	2,83	1,00	0,96	1,00
	PU_2	3,67	2,50	1,00		0,67
	PU_3	3,83	2,67	1,00		1,00
	PU_4	3,50	2,67	1,00		0,67
	PU_5	3,33	2,67	0,83		1,00
PEOU - Perceived Ease of Use	PEOU_1	3,83	2,83	1,00	0,94	0,67
	PEOU_2	3,67	2,67	1,00		0,67
	PEOU_3	3,50	2,67	0,83		1,00
	PEOU_4	3,50	2,50	1,00		1,00
	PEOU_5	3,83	2,67	1,00		1,00
PC - Privacy Concern	PC_1	3,83	2,67	1,00	0,94	1,00
	PC_2	3,83	2,83	1,00		1,00
	PC_3	3,83	2,83	1,00		0,67
	PC_4	3,83	2,83	1,00		0,67
	PC_5	3,83	3,00	1,00		1,00
TR - Trust	TR_1	3,67	3,00	1,00	0,85	1,00
	TR_2	3,83	2,83	1,00		0,67
	TR_3	3,33	2,83	0,83		1,00
	TR_4	3,33	2,83	0,83		1,00
	TR_5	4,00	2,83	1,00		0,67
BI - Behavioral Intention	BI_1	3,83	2,83	1,00	0,92	1,00
	BI_2	3,67	2,67	1,00		1,00
	BI_3	3,83	2,67	1,00		0,67
	BI_4	3,50	2,50	0,83		1,00
	BI_5	4,00	2,83	1,00		1,00

*Source: own elaboration.

The Item-Level Content Validity Index (I-CVI) values for the individual items of the proposed instrument ranged from 0.83 to 1.00. Validation studies generally agree that an acceptable I-CVI value for newly developed instruments should be 0.80 or higher. The average scale-level content validity index (S-CVI/Ave) reached 0.96 for the Perceived Usefulness (PU) dimension, 0.94 for Perceived Ease of Use (PEOU), 0.85 for Trust (TR), and 0.92 for Behavioural Intention (BI). To assess the overall suitability of the proposed measurement instrument, the total average content validity index reached $S-CVI/Ave = 0.92$, while the universal agreement index was $S-CVI/UA = 0.60$. For the evaluated panel of six experts, the acceptable threshold for I-CVI is above 0.78

(Almanasreh et al., 2022). The average item score was 3.83, ranging from 3.50 to 4.00. Based on the obtained results and comparison with recommended threshold values, the items of the proposed instrument can be considered acceptable in terms of both relevance and necessity. After computing the I-CVI, for all items, the Kappa modified (κ^*) was calculated by using the value P_c . Several standards for evaluating kappa recommend whether the value for each κ^* is fair (0.40-0.59), good (0.60-0.74), or excellent ($\kappa^* > 0.74$) (Polit et al., 2007). Table 2 presents the calculated values of the modified kappa coefficient (κ^*) and the supplementary content validity index, Aiken's V.

Table 2. Supplementary Content Validity Coefficients

Dimension	Item	Relevant	I-CVI	Pc	Aiken's V	Interpretat	
PU - Perceived Usefulness	PU_1	6	1,00	0,02	1,00	0,94	excellent
	PU_2	6	1,00	0,02	1,00	0,89	excellent
	PU_3	6	1,00	0,02	1,00	0,94	excellent
	PU_4	6	1,00	0,02	1,00	0,83	excellent
	PU_5	5	0,83	0,09	0,82	0,78	excellent
PEOU - Perceived Ease of Use	PEOU_1	6	1,00	0,02	1,00	0,94	excellent
	PEOU_2	6	1,00	0,02	1,00	0,89	excellent
	PEOU_3	6	0,83	0,02	0,83	0,83	excellent
	PEOU_4	6	1,00	0,02	1,00	0,83	excellent
	PEOU_5	6	1,00	0,02	1,00	0,94	excellent
PC - Privacy Concern	PC_1	6	1,00	0,02	1,00	0,94	excellent
	PC_2	6	1,00	0,02	1,00	0,94	excellent
	PC_3	6	1,00	0,02	1,00	0,94	excellent
	PC_4	6	1,00	0,02	1,00	0,94	excellent
	PC_5	6	1,00	0,02	1,00	0,94	excellent
TR - Trust	TR_1	6	1,00	0,02	1,00	0,89	excellent
	TR_2	6	1,00	0,02	1,00	0,94	excellent
	TR_3	5	0,83	0,09	0,82	0,78	excellent
	TR_4	5	0,83	0,09	0,82	0,78	excellent
	TR_5	6	1,00	0,02	1,00	1,00	excellent
BI - Behavioral Intention	BI_1	6	1,00	0,02	1,00	0,94	excellent
	BI_2	6	1,00	0,02	1,00	0,89	excellent
	BI_3	6	1,00	0,02	1,00	0,94	excellent
	BI_4	5	0,83	0,09	0,82	0,83	excellent
	BI_5	6	1,00	0,02	1,00	1,00	excellent

*Source: own elaboration.

For all items, the value of the modified kappa coefficient (κ^*) ranged from 0.82 to 1.00, which is interpreted as excellent agreement. The values of Aiken's V ranged from 0.83 to 1.00, while all items achieved at least the minimum acceptable value. The results confirm a high level of content validity of the instrument.

The average value of CVR for the entire instrument reached 0.88. Most items were considered essential by the experts, although complete agreement was not achieved for some items. To assess the level of agreement among experts in evaluating the individual items of the proposed measurement instrument, Kendall's coefficient of concordance (Kendall's W) was employed. This coefficient measures the degree of consensus among raters who evaluate the same set of items. The results of the analysis showed a value of Kendall's W = 0.55, indicating a good level of agreement among the experts in their assessment of the relevance of the questionnaire items. According to commonly used interpretation guidelines for Kendall's W, values within the interval 0.50–0.70 indicate a good level of consensus, suggesting that the experts demonstrated a substantial level of agreement in their evaluations. The statistical significance of the agreement was verified using the chi-square test, with the result $\chi^2(5) = 27.5$, which was highly statistically significant ($p < 0.05$). This finding confirms that the observed level of agreement among the experts is not due to chance but reflects a systematic consensus in the evaluation of the assessed items. Based on these results, it can be concluded that the expert assessment demonstrates a statistically significant and methodologically satisfactory level of agreement, thereby supporting the content validity of the proposed measurement instrument. The obtained degree of concordance indicates that the experts largely agreed in their evaluation of the relevance and representativeness of the items with respect to the examined constructs. Overall, the identified level of agreement among the experts provides an important methodological justification for the further use and empirical validation of the proposed questionnaire in subsequent stages of the research.

Discussion

The study aimed to assess the content validity of a newly developed instrument for

measuring customer acceptance of chatbots in the insurance sector. The validation process was conducted through a structured expert evaluation procedure combined with multiple content validity indices. The results confirmed a high level of content validity and satisfactory agreement among experts, supporting the suitability of the instrument for further empirical testing. The expert evaluation process resulted in the refinement and reduction of the initial item pool, leading to a final instrument consisting of 25 items across five theoretical dimensions: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Privacy Concern (PC), Trust (TR), and Behavioural Intention (BI). These dimensions are consistent with established technology acceptance frameworks, particularly TAM and related models explaining user perception of digital technologies.

The calculated I-CVI values ranged from 0.83 to 1.00, exceeding the recommended threshold for newly developed instruments. The overall scale-level content validity index (S-CVI/Ave = 0.92) also indicated a high level of content validity. Additional indicators, including the modified kappa coefficient and Aiken's V, confirmed excellent agreement among experts, while the average CVR value demonstrated that most items were considered essential for measuring the examined construct. Kendall's coefficient of concordance (W = 0.55) further confirmed a statistically significant level of expert consensus.

Overall, the findings suggest that the proposed instrument represents a theoretically grounded and methodologically robust tool suitable for future empirical research focused on chatbot acceptance in insurance services. Future studies should focus on verifying the factor structure, reliability, and construct validity of the instrument using larger respondent samples.

The obtained findings are consistent with previous studies emphasizing the importance of theoretically grounded instruments for investigating chatbot acceptance. Like Andrés-Sánchez and Gené-Albesa (2023a, 2023b, 2024), the present instrument is based on the core determinants of technology acceptance, particularly perceived usefulness, perceived ease of use, trust, and behavioural intention. However, unlike previous studies that primarily examined

structural relationships among these constructs, the present study focuses on the methodological development and validation of the measurement instrument itself. This distinction is important because reliable empirical research depends on psychometrically sound instruments.

The findings are also consistent with recommendations by Polit et al. (2007), who advocate combining several complementary content validity indices rather than relying solely on the Content Validity Index. The simultaneous application of CVI, CVR, modified kappa coefficient, Aiken's V, and Kendall's coefficient of concordance increase the methodological robustness of the validation process and provides stronger evidence supporting the adequacy of the proposed questionnaire.

From a practical perspective, the validated instrument may support insurance companies in identifying factors influencing customers' willingness to interact with AI-supported communication systems. Better understanding of customer perceptions regarding usefulness, trust, ease of use, and privacy concerns may contribute to more effective chatbot implementation strategies and improved customer experience. At the same time, the instrument creates opportunities for comparative studies across different insurance markets and customer groups.

Conclusion

This study contributes to the growing body of research on artificial intelligence adoption by addressing an important methodological gap in the insurance sector. Unlike previous studies that mainly focused on explaining chatbot acceptance, the present research developed and content-validated a dedicated measurement instrument specifically designed for insurance customer interactions. The proposed instrument provides a reliable foundation for future empirical studies investigating the determinants of chatbot acceptance in insurance services.

Acknowledgments

The EU NextGenerationEU funded this research through the Recovery and Resilience Plan for Slovakia under the project No. 09I02-03-V01-00029.

The study developed and content-validated a measurement instrument for assessing customer acceptance of chatbots in the insurance sector. The developed instrument achieved high content validity, with I-CVI values ranging from 0.83 to 1.00 and an overall S-CVI/Ave of 0.92. Kendall's coefficient of concordance ($W = 0.55$) confirmed statistically significant expert agreement. Therefore, the instrument can be considered suitable for subsequent psychometric validation. The final instrument consists of 25 items organized into five domains: perceived usefulness, perceived ease of use, privacy concern, trust, and behavioural intention. The obtained results indicate that the instrument adequately represents the examined construct and may be applied in future empirical studies focused on AI-supported customer communication in insurance organizations. The study contributes to technology acceptance research by providing a context-specific and methodologically validated measurement instrument for the insurance sector. At the practical level, the instrument may support insurance companies in evaluating customer perceptions of chatbot technologies and improving digital communication strategies. The study is subject to several limitations. First, the expert panel was relatively small, which may have affected the stability of some content validity indicators, particularly the Content Validity Ratio (CVR). Second, although multiple complementary indices were applied to strengthen the validation process, expert evaluation remains inherently subjective. Third, the study focused exclusively on content validity assessment; therefore, the factor structure, construct validity, and reliability of the proposed instrument were not examined. Future research should focus on pilot testing, reliability assessment, exploratory and confirmatory factor analysis, and verification of construct validity using larger and more diverse respondent samples.

References

- Almanasreh, E., Moles, R. J., & Chen, T. F. (2022). A practical approach to the assessment and quantification of content validity. In *Contemporary research methods in pharmacy and health services* (pp. 583-599). Academic Press.
- Andrés-Sánchez, J., & Gené-Albesa, J. (2023a). Assessing Attitude and Behavioral Intention toward Chatbots in an Insurance Setting: A Mixed Method Approach. *International Journal of Human-Computer Interaction*, 40, 4918–4933. <https://doi.org/10.1080/10447318.2023.2227833>
- Andrés-Sánchez, J., & Gené-Albesa, J. (2023b). Explaining Policyholders' Chatbot Acceptance with an Unified Technology Acceptance and Use of Technology-Based Model. *J. Theor. Appl. Electron. Commer. Res.*, 18, 1217–1237. <https://doi.org/10.3390/jtaer18030062>
- Andrés-Sánchez, J., & Gené-Albesa, J. (2024). Drivers and Necessary Conditions for Chatbot Acceptance in the Insurance Industry. Analysis of policyholders' and professionals' Perspectives. *Journal of Organizational Computing and Electronic Commerce*, 35, 189–216. <https://doi.org/10.1080/10919392.2024.2435118>
- Borsci, S., Malizia, A., Schmettow, M., van der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2022). The Chatbot Usability Scale: The Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. *Personal and Ubiquitous Computing*, 26(1), 95–119. <https://doi.org/10.1007/s00779-021-01582-9>
- Gebert-Persson, S., Gidhagen, M., Sallis, J. E., & Lundberg, H. (2019). Online insurance claims: When more than trust matters. *International Journal of Bank Marketing*, 37(2), 579–594. <https://doi.org/10.1108/IJBM-02-2018-0024>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
- Nov, O., Singh, N., & Mann, D. (2023). Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. *JMIR Medical Education*, 9(1), e46939. <https://doi.org/10.2196/46939>
- Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic Information Systems. *Business & Information Systems Engineering*, 61(4), 523–533. <https://doi.org/10.1007/s12599-019-00599-y>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. *Business Horizons, Digital Transformation & Disruption*, 62(6), 785–797. <https://doi.org/10.1016/j.bushor.2019.08.005>
- Rese, A., Ganster, L., & Baier, D. (2020). Chatbots in retailers' customer communication: How to measure their acceptance? *Journal of Retailing and Consumer Services*, 56, 102176. <https://doi.org/10.1016/j.jretconser.2020.102176>
- Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. *AJOB Neuroscience*, 11(2), 88–95. <https://doi.org/10.1080/21507740.2020.1740350>
- Yang, R., & Wibowo, S. (2022). User trust in artificial intelligence: A comprehensive conceptual framework. *Electronic Markets*, 32(4), 2053–2077. <https://doi.org/10.1007/s12525-022-00592-6>
- Yang, W., & Zhou, J. (2021). Service Innovation of Insurance Data Based on Cloud Computing in the Era of Big Data. *Complexity*, 2021(1), 2303129. <https://doi.org/10.1155/2021/2303129>