

Article

Assessing Chatbot Acceptance in Policyholder's Assistance Through the Integration of Explainable Machine Learning and Importance–Performance Map Analysis

Jaume Gené-Albesa ¹ and Jorge de Andrés-Sánchez ^{2,*}

¹ Department of Business Administration, Campus de Bellissens, University Rovira i Virgili, 43204 Reus, Spain; jaume.gene@urv.cat

² Social and Business Research Laboratory, Campus de Bellissens, University Rovira i Virgili, 43204 Reus, Spain

* Correspondence: jorge.deandres@urv.cat

Abstract

Companies are increasingly giving more attention to chatbots as an innovative solution to transform the customer service experience, redefining how they interact with users and optimizing their support processes. This study analyzes the acceptance of conversational robots in customer service within the insurance sector, using a conceptual model based on well-known new information systems adoption groundworks that are implemented with a combination of machine learning techniques based on decision trees and so-called importance–performance map analysis (IPMA). The intention to interact with a chatbot is explained by performance expectancy (PE), effort expectancy (EE), social influence (SI), and trust (TR). For the analysis, three machine learning methods are applied: decision tree regression (DTR), random forest (RF), and extreme gradient boosting (XGBoost). While the architecture of DTR provides a highly visual and intuitive explanation of the intention to use chatbots, its generalization through RF and XGBoost enhances the model's explanatory power. The application of Shapley additive explanations (SHAP) to the best-performing model, RF, reveals a hierarchy of relevance among the explanatory variables. We find that TR is the most influential variable. In contrast, PE appears to be the least relevant factor in the acceptance of chatbots. IPMA suggests that SI, TR, and EE all deserve special attention. While the prioritization of TR and EE may be justified by their higher importance, SI stands out as the variable with the lowest performance, indicating the greatest room for improvement. In contrast, PE not only requires less attention, but it may even be reasonable to reallocate efforts away from improving PE in order to enhance the performance of the more critical variables.

Keywords: chatbots; insurance; decision tree regression; random forest; XGBoost; Shapley additive explanations (SHAP)



Academic Editors: Grzegorz Dudek and Arkadiusz Tomczyk

Received: 19 July 2025

Revised: 11 August 2025

Accepted: 15 August 2025

Published: 17 August 2025

Citation: Gené-Albesa, J.; de Andrés-Sánchez, J. Assessing Chatbot Acceptance in Policyholder's Assistance Through the Integration of Explainable Machine Learning and Importance–Performance Map Analysis. *Electronics* **2025**, *14*, 3266. <https://doi.org/10.3390/electronics14163266>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Industry 4.0 has revolutionized industrial production through the integration of advanced technologies such as artificial intelligence (AI), the Internet of Things (IoT), automation, and real-time data analytics [1]. The transformative power of Industry 4.0 has extended to other sectors, including finance, giving rise to Finance 4.0. Finance 4.0 leverages digitalization to provide innovative solutions, enhancing the efficiency of the global financial system [2]. These solutions are commonly referred to as Fintech and have driven

disruptive models such as digital banking, cryptocurrencies, and smart contracts, marking the beginning of a new era [3].

Insurance 4.0 is a specialized branch of Finance 4.0, focusing on the insurance industry. Thus, applications of Industry 4.0 technologies in the insurance field are labelled as Insurtech [4]. Insurtech impacts all areas of the insurance business, enabling automation in risk assessment, policy personalization, and the streamlining of claims processing through smart contracts and connected devices to prevent losses. Insurtech is redefining how insurers operate, making them more efficient, accessible, and customer-centric [5].

Among the most significant Insurtech applications, chatbots stand out as a key tool in customer service, improving both the operational efficiency of financial and insurance organizations and user accessibility [6]. Chatbots can respond to queries in real time, streamline policy management, facilitate claims reporting, and guide customers through complex processes without requiring immediate human intervention. Furthermore, they contribute to delivering a personalized experience by analyzing user data and providing tailored recommendations to meet specific needs [7]. They also help reduce operational costs and response times, ultimately increasing customer satisfaction. Their ability to operate 24/7 makes them an indispensable tool for enhancing customer service in an increasingly digitalized market [8].

Despite their benefits, many customers remain skeptical about chatbots in the insurance sector, perceiving them as providing impersonal and limited assistance. The lack of empathy in interactions, automated responses that sometimes fail to resolve complex queries, and difficulties in quickly escalating issues to a human agent contribute to user frustration [8].

This study analyzes the drivers of chatbot acceptance among policyholders for managing tasks related to their active insurance policies (e.g., filing a claim). The analysis of conversational robot acceptance in customer service is of particular interest in the insurance industry, as the use of an insurance policy—entailing claims notification—always requires communication with the insurer, both during the initial contact and in the subsequent transmission of relevant details [9].

The approach adopted in this study is based on the technology acceptance model (TAM) [10] and the unified theory of acceptance and use of technology (UTAUT) [11]. Specifically, it seeks to explain the intention to use (IU) chatbots for managing active policies through the following constructs: performance expectancy (PE), effort expectancy (EE), social influence (SI), and trust (TR). While the first three constructs are the most relevant in the literature on conversational robot acceptance [12], the inclusion of trust is justified by its key role both in the economic function of insurance [13] and in the adoption of conversational robots [14].

The analytical framework is presented in Figure 1. Concretely, this paper addresses two research questions:

- RQ1: What is the explanatory and predictive power of the proposed model?
- RQ2: What are the constructs that require greater attention for the successful implementation of chatbots?

The first contribution of this study is methodological. Unlike most of the existing literature, which typically addresses RQ1 using structural equation modeling (SEM), this research employs machine learning techniques (MLT). Concretely, it uses decision tree regression (DTR) and its ensemble generalizations—random forest (RF) and extreme gradient boosting (XGBoost). While SEM requires the specification of linear and pre-defined interactions among constructs, DTR offers a data-driven approach that identifies decision thresholds and interaction effects that may not be apparent in linear models. Additionally, it enables decision-makers to intuitively visualize relationships between variables while ac-

counting for nonlinearity [15]. This approach is particularly useful for analyzing behavioral phenomena such as consumer behavior [16,17] and the acceptance of new technologies, where the literature often lacks consensus on how explanatory variables interact to shape attitudes [18–20].

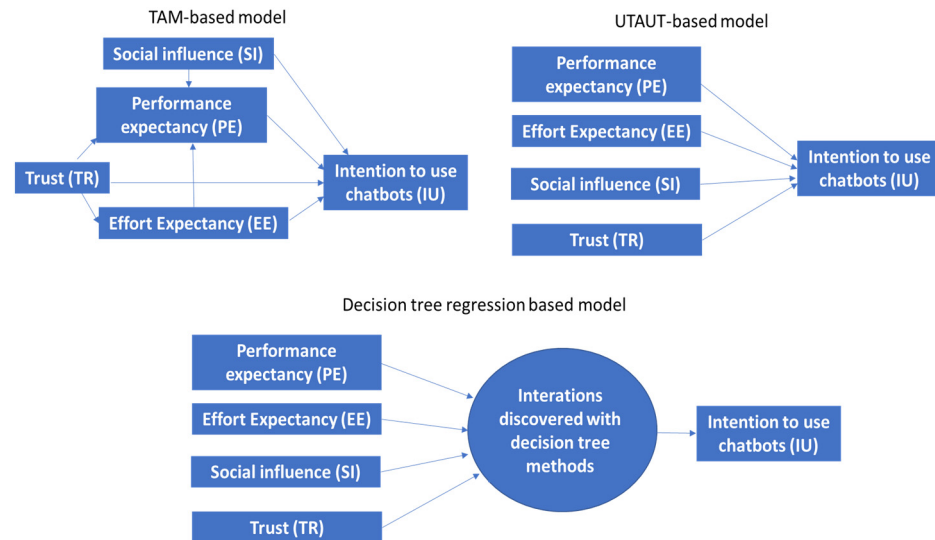


Figure 1. Conceptual framework of this paper versus TAM and UTAUT approaches.

In extended versions of the technology acceptance model (TAM), such as TAM2 [21], the influence of effort expectancy (EE) and social influence (SI) on intention to use (IU) is partially mediated by perceived usefulness (PU). On the other hand, the unified theory of acceptance and use of technology (UTAUT) assumes only direct effects. In the more specific context of chatbot adoption, some studies adopt a TAM-based perspective in which trust influences IU indirectly through PE and EE [22,23], while others posit a direct effect of trust on IU [24,25].

This study does not posit specific interaction hypotheses between predictors. Instead, we examine how explanatory variables correlate with IU, allowing the decision tree architecture to uncover the interaction patterns present in the data. Furthermore, combining DTR with ensemble models such as RF [26] or XGBoost [27] enhances the model's predictive performance. In this context, DTR can be interpreted as a representative or average tree within the ensemble of trees generated by RF and XGBoost. Figure 1 presents two conceptual modeling approaches previously used to explain chatbot acceptance and highlights the architecture ultimately adopted in this study.

To address RQ2, the study applies Shapley additive explanations (SHAP) [28], which enable interpretation of the relative importance of each explanatory variable in the predictions generated by the best-performing decision tree-based model—whether DTR, or more likely, RF or XGBoost. Subsequently, an importance–performance map analysis (IPMA) will be used, based on the diagonal partitioning of the IPM [29]. This will be adapted from the structural equation modeling approach proposed by [30] to the use of SHAP as a measure of the relative importance of variables. The value of IPMA lies in recognizing that the variable with the greatest impact does not necessarily warrant the most attention. If the performance of a given variable is already high compared to the other explanatory variables, improving it further may require substantial effort. Thus, it may be more effective to focus on variables with lower importance but where performance improvements are more feasible, potentially generating a greater overall impact on the target variable.

A key innovative contribution of this study lies in the integration of explainable machine learning techniques—specifically, decision tree regression, random forest, and

XGBoost—with Shapley additive explanations (SHAP) and importance–performance map analysis (IPMA) within the context of technology acceptance research. While previous studies on technology adoption have largely relied on structural equation modeling or other traditional statistical approaches, our method enables both high predictive accuracy and an intuitive understanding of the interaction patterns among explanatory variables without needing hypotheses about mediations and moderations. This combined approach not only bridges the gap between interpretability and predictive performance but also offers a replicable framework for future research in other technology adoption domains.

In addition, this research expands the scope of recent work by applying the proposed methodology to the insurance sector—a service industry where trust plays a central role—thereby capturing sector-specific behavioral drivers often overlooked in generalist studies. By comparing our findings with those from the latest empirical research across diverse geographical contexts and chatbot applications, we demonstrate how the explanatory hierarchy of variables can shift depending on cultural, industrial, and service-channel factors. This comparative perspective reinforces the originality of the study, as it provides nuanced insights that are both theoretically relevant and practically actionable for AI-powered service implementation.

2. Framework

Performance expectancy (PE) refers to the degree to which users perceive that a system enhances their performance in carrying out a task [11]. There are several reasons to conclude that chatbots are useful in insurance procedures. On the one hand, basic administrative tasks can be completed more quickly than when relying solely on human assistance [31]. Similarly, conversational robots do not substitute conventional interaction ways with the insurer but rather are an additional instrument that allows the improvement of policyholders' assistance [32]. That variety of communication channels is frequently valued and helps build customer satisfaction [33].

PE is likely the most influential construct in the acceptance of chatbots for customer service in both banking [6,34–37] and insurance contexts, where its impact has been observed both directly [8,38,39] and indirectly through mediating mechanisms [7]. Therefore, we propose:

Hypothesis 1. *Performance expectancy positively influences the intention to use chatbots for managing active insurance policies.*

Venkatesh et al. [11] define effort expectancy (EE) as the extent to which an individual believes that using a technology requires little effort. In the use of chatbots to provide customer service in insurance companies, EE refers to the absence of drawbacks for policyholders when carrying out procedures related to in-force contracts. In theory, conversational bots have specific benefits compared to other communication methods. They offer round-the-clock support and are more available than human agents [31]. Furthermore, they present fewer usability barriers compared to other digital technologies, as they can be accessed from multiple devices, including smartphones, tablets, computers, and landlines [22].

At present, there is widespread agreement that chatbots have not yet reached a level of sophistication that allows smooth and flawless interaction in every situation. Frequently, chatbots deliver unclear replies to users, which undermines their perceived usability and effectiveness. This, in turn, leads to a decline in user acceptance and satisfaction [40]. Notable issues in this regard include technological anxiety towards robots, which significantly influences usability perceptions and customer attitudes towards chatbots [41], as well as

the inability of chatbots to recognize vocal tones and inflections that help determine the direction of a conversation [40].

The relevance of effort expectancy in the acceptance of chatbots has been well documented in financial contexts, particularly in the use of banking and insurance services. In the banking sector, its influence has been reported in various studies [6,35,37,42], while in the insurance domain, it has been shown to play both a direct role [8,38] and a mediated one [7,41]. Therefore:

Hypothesis 2. *Effort expectancy positively influences the intention to use chatbots for managing active insurance policies.*

SI refers to the extent to which individuals perceive that important people believe they should use a new technology [11]. It is a well-established fact that peer opinions, such as those of friends or family members, have a significant impact on overall technology acceptance, as individuals tend to seek social approval [11].

The opinion of close insurance advisers is often relevant in policyholders' decision-making [43]. Despite the widespread adoption of chatbots in business practices, most consumers remain skeptical and reluctant to engage with them [44]. In fact, chatbots are primarily used to provide initial help to users and consumers [43]. The relevance of social influence in explaining the acceptance of conversational robots has been demonstrated in the contexts of both banking [37] and insurance procedures [8]. So, we suggest:

Hypothesis 3. *Social influence positively influences the intention to use chatbots for managing active insurance policies.*

The importance of TR in policyholders' acceptance of Insurtech solutions, including chatbots, should be analyzed from a dual perspective. It embeds the unique nature of the financial and insurance industry and the interactions between companies and policyholders that are facilitated by robots. Therefore, TR becomes a critical factor in understanding customer attitudes and behavioral intentions [13].

Trust is the foundation of any financial transaction and is even more crucial in the insurance market, where both the insurer and the policyholder must rely on mutual trust in an environment characterized by a high degree of adverse selection and moral hazard [45]. A policyholder's trust in an insurance company can be defined as the perception that its services will offer reliable compensation in the event of a loss and that interactions related to claims will be satisfactory [45]. The relevance of trust in the acceptance of chatbots has been observed in contexts related to insurance, such as banking services [6,42], as well as within the insurance sector itself—both directly [7,39,41] and indirectly, mediated by performance expectancy and effort expectancy [22]. Therefore, we propose:

Hypothesis 4. *Trust positively influences the intention to use chatbots for managing active insurance policies.*

3. Materials and Methods

3.1. Sample and Sampling

The paper analyzed an online survey distributed via social media platforms (LinkedIn, Facebook, Telegram) and moderated mailing lists, conducted between 20 December 2022 and 12 March 2023.

Respondents were encouraged to share the survey hyperlink with others, meaning the sampling methodology used was mixed, combining convenience sampling and snowball sampling. The estimated time to complete the questionnaire was 10–15 min.

Considering the duration of data collection, it was adequate for a cross-sectional study such as ours. Such studies require a certain window of time to obtain an adequate number of responses, but they are ultimately a snapshot at a specific point in time; so the survey must be anchored within a defined moment. According to the literature reviewed on cross-sectional studies focused on chatbot acceptance—among those authors who actually reported how long it took them to collect data, which not all did—the time frame ranged from half a month [22] to three months [38,41].

Regarding the focus on a specific cultural context, this was also common in social sciences and human behavior studies. Such research sought responses within a defined geographical area, either to inform action or to gain insights without “contamination” from other contexts. For example, in Asia, Ref. [22] focused on Korea, Ref. [24] on India, Ref. [35] on China, and [42] on Bangladesh. In Europe, Ref. [23] was set in Spain, Ref. [25] in the United Kingdom, Ref. [34] in Romania, and Ref. [39] in Germany.

As we sought opinions from genuinely informed consumers, only responses from individuals who held at least two insurance policies were accepted. Given that the survey targeted a very specific population segment, convenience sampling could be considered appropriate [46]. Moreover, respondents were not compensated, making it reasonable to assume they were genuinely motivated to answer the questions and paid attention to their responses.

The initial number of observations was 252. We subsequently discarded incomplete responses, resulting in a final sample of 226 responses, which was considered statistically adequate according to the heuristic “ten times rule” [47], which suggests that the minimum required sample size should be 40, given that there were only four explanatory variables for behavioral intention. Additionally, using the G*Power 3.1 software [48], we verified that this sample size provided a statistical power of 80% for a linear regression with four variables, assuming a significance level of 5% and an effect size of at least 0.05, which corresponded to a minimum coefficient of determination of 4.76%. The profile of the individuals in the sample is shown in Table 1.

Table 1. Sociodemographic profile of the sample (*n* = 226).

Variable	Responses
Gender	53.10% of responses came from men and 44.69% from women, and 2.21% provided other responses.
Age	14.16% of responses came from individuals under 40 years old, 53.98% from those aged between 40 and 55, 30.09% from individuals over 55 years old and 1.77% did not answer.
Academic background	87.17% of respondents reported having completed a university degree and 12.83% reported being undergraduate.
Income level	30.09% reported an income not exceeding EUR 1750, 38.94% reported an income level between EUR 1750 and EUR 3000, a 30.09% reported earning over EUR 3000 and 0.88% did not answer.
Number of insurance policies	47.79% of respondents reported holding between 2 and 4 policies, while 52.21% held more than 4 policies.

3.2. Measurement Model

The survey was conducted using a structured questionnaire written in Spanish, with the items presented in Table A1 of the Appendix A. Initially, the questionnaire was distributed among six professionals from the insurance industry in Spain. After receiving their feedback and incorporating it into a revised version, it was assessed by an additional twelve voluntaries that were not professionally linked with the insurance industry.

Regarding the scales used, the IU, PE, EE, and SI measures were based on the proposals of [11], adapted to the use of chatbots in the policyholder–insurer relationship. The trust scale was based on [49].

Responses were collected using an eleven-point Likert scale (ranging from 0, indicating strong disagreement with the statement, to 10, representing strong agreement), where 5 was the neutral value.

3.3. Data Analysis

First step: Since the study dealt with latent variables, the first step involved assessing the internal reliability and the discriminant validity of scales [47]. This includes calculating Cronbach’s alpha, the composite reliability index, and the average variance extracted (AVE) and conducting factor extraction through exploratory factor analysis. Additionally, the correlation matrix of the constructs was analyzed to provide a first assessment of the consistency of the hypotheses regarding the direction of the relationships between the model variables. This step was performed using the *psych* package in R.

Second step: The final scores for each construct were determined by calculating the weighted average of the items based on the factor extraction, which was then rescaled to a 100-point reference system. This approach follows [30] for evaluating the performance of latent variables.

If construct X is composed of I items x_i , where $i = 1, 2, \dots, I$, and we denote w_i as the percentage of variance extracted for the i th item, then the value of the construct for the j th observation, X_j , was calculated as the weighted average of the scores for each item of that observation (x_{ij}), weighted by the factor extraction of that item (w_i). Since x_i takes values between 0 and 10, and X_j is, following [30], referenced on a 100-point scale, we calculate:

$$X_j = \frac{\sum_{i=1}^I x_{i,j} w_i}{\sum_{i=1}^I w_i} \cdot 10.$$

This step was carried out using the *psych* and *dplyr* packages.

Third step: Subsequently, a decision tree regression (DTR) model was fitted with all explanatory constructs. Since the variables were already measured on a 100-point scale, rather than using standardized factor extractions, this facilitates the interpretation of the tree and the cut-off values at the nodes. The sign of the relationship between an explanatory variable X and IU was inferred from how observations are distributed across the nodes in which it participates. If a threshold $X < X_a$ is required to reach terminal nodes associated with lower acceptance, then the relationship is positive. Conversely, if reaching these nodes requires $X > X_a$, a negative relationship can be inferred. It is important to note that, in assessing the sign of the relationship, we considered not only the primary splits but also the surrogate splits—that is, the alternative splits that would be used if the observation for the variable responsible for the primary split were missing. This step was conducted using the *rpart* and *rpart.plot* packages in R.

Fourth step: When the objective extends beyond model explanation to achieve more accurate fits and predictions, ensemble methods such as RF and XGBoost generalize decision trees in a way that enhances predictive performance—albeit at the cost of interpretability, which is a key strength of single decision trees.

For RF and XGBoost, all explanatory variables were also included, and a hyperparameter tuning process was performed [50,51]. The dataset was randomly split into 80% for training and 20% for testing. Within the training set, 70% of the full dataset was used for actual model training, and hyperparameter tuning was performed via 10-fold cross-validation applied to this 70%. This internal cross-validation step acted as the validation phase, replacing the need for a fixed 10% hold-out.

Hyperparameter tuning for RF focused on three parameters [51,52]: the number of variables randomly selected at each split (*mtry*), the number of trees in the forest (*ntree*), and the minimum number of observations in a terminal node (*nodesize*).

For XGBoost, we tuned the learning rate (*eta*), the maximum tree depth (*max_depth*), the minimum child weight (*min_child_weight*), and the number of boosting rounds (*nrounds*) [51,52].

Notice that, in contrast, DTR was fitted directly using the *rpart* package on the entire dataset without additional tuning, as the goal was to preserve interpretability.

This step is executed using the *caret* package in R, in combination with the *randomForest* and *xgboost* packages to implement the respective algorithms.

Fifth step: Once the RF and XGBoost models had their hyperparameters tuned (Step 4), and the DTR was fitted as described above, all models were evaluated using the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) to assess in-sample fit.

Out-of-sample predictive performance was assessed using Monte Carlo cross-validation with repeated random subsampling (80/20 split, 5000 repetitions). In each repetition, 80% of the data was randomly selected for training and 20% for testing. Predictive performance was quantified using Stone–Geisser’s Q^2 statistic, along with RMSE and MAE, averaged across all repetitions to ensure robustness and mitigate the variability associated with a single train–test split. This step is performed using the *caret*, *rsample*, *randomForest*, and *xgboost* packages.

This evaluation addressed research question 1, enabling a robust assessment of the models’ predictive and explanatory capabilities, as well as the visualization of interaction patterns that drive chatbot acceptance, thereby facilitating the evaluation of Hypotheses H1–H4. An overall overview about fourth and fifth step is provided in Table 2.

Table 2. Data partitioning and cross-validation methods used in Steps 4 and 5.

Method	Training Set	Validation Set	Testing Set	Cross-Validation Method (Step 5)
DTR	100%	—	20%	Monte Carlo CV, 5000 reps
RF	70%	10% CV on training	20%	Monte Carlo CV, 5000 reps
XGBoost	70%	10% CV on training	20%	Monte Carlo CV, 5000 reps

Note: The first three columns refer to the data partitioning used for model training and hyperparameter tuning (Step 4). The last column indicates the cross-validation method applied for final evaluation (Step 5).

Sixth step: To test whether the differences in predictive performance between models were statistically significant, we conducted paired-sample *t*-tests and ANOVA on the prediction metrics. This analysis provided an evidence-based comparison of the models and helped identify whether certain decision tree-based methods consistently outperformed others. This step was implemented using the *rstatix* R package.

Seventh step: To address research question 2, we first computed SHAP values for each variable across all observations. These values allowed us to calculate the mean absolute SHAP values, which represent the average contribution of each variable to the

model predictions. This enabled the construction of a hierarchy of relevance among the explanatory variables, offering insights into their relative importance in explaining IU chatbots. This step was carried out using the *iml* package in R.

Eighth step: Finally, to complete the analysis of research question 2, an importance–performance map analysis (IPMA) was conducted. The performance of the constructs was stated simply as the sample mean of the items, that were rescaled in 100 [30]. On the other hand, the importance of variables was their average absolute value of SHAP. The interpretation of the importance–performance map followed the diagonal partitioning approach proposed by [29] and is illustrated in Figure 2.

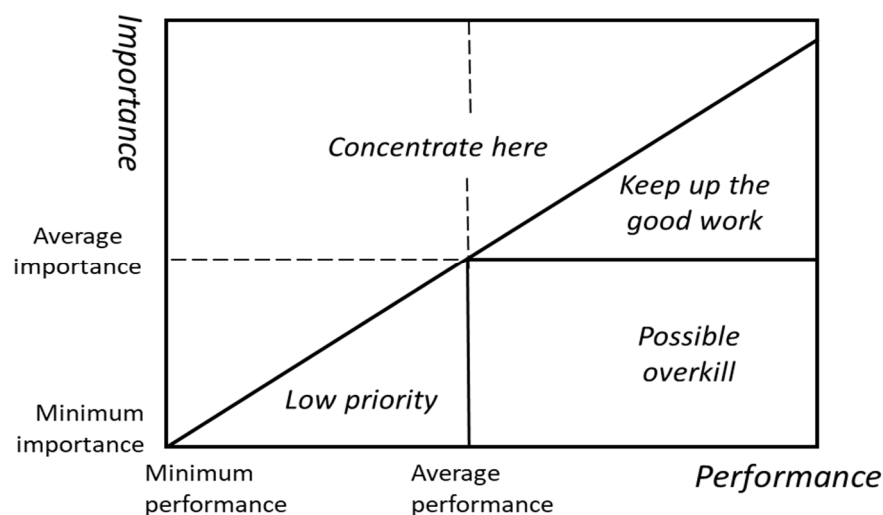


Figure 2. Importance–performance map interpretation used in this study. Note: Adapted from Abalo et al. [29].

4. Results

4.1. Analysis of Research Question 1

Table 3 shows the descriptive statistics of the items that make up the scales and the measures of their internal validity. It can be observed that every item received a score considerably less than 5, indicating a very low evaluation among users. The scales exhibit internal consistency (Cronbach’s alpha and the composite reliability index > 0.7) and convergent validity, as the factor extraction for the constructs is > 0.7 , and the average variance extracted is, in all cases, > 0.5 . Table 4 shows that the constructs have discriminant validity according to the Fornell–Larcker criterion, as the correlations between the constructs never exceed the square root of the average variances extracted.

In Table 4, it can be also checked that the hypotheses regarding the positive relationships between PE, EE, SI, and TR, with IU, are supported by the Pearson correlations, which are consistently positive and significantly different from zero. This is further confirmed by Figure 3, where all variables contribute to the splitting of at least one node. In each case, values below the threshold lead to a lower level of chatbot acceptance. Indeed, Figure 3 and Table 5 show that TR, EE, and SI each serve as the primary split into two partitions, while PE serves as the primary split into one. Table 5 presents not only the primary splits but also the surrogate splits. It can also be observed that when the explanatory variables act as surrogate splits, their direction of influence suggests a positive relationship of all variables with IU. Observations with values below the threshold are classified into nodes associated with lower levels of acceptance. So, from Tables 4 and 5 we can conclude that Hypothesis H1, H2, H3, and H4 can be accepted.

Table 3. Descriptive statistics and measures of internal reliability of scales.

Item	Mean	SD	Factor Loading	CA	CR	AVE
IU1	1.27	1.87	0.921	0.891	0.894	0.822
IU2	2.24	2.7	0.862			
IU3	1.38	2.06	0.935			
PE1	2.44	2.63	0.877	0.92	0.932	0.76
PE2	2.71	2.66	0.91			
PE3	2.57	2.58	0.904			
PE4	2.46	2.61	0.914			
PE5	3.29	2.86	0.742			
EE1	2.88	2.82	0.864	0.885	0.893	0.813
EE3	2.16	2.27	0.922			
EE4	2.64	2.64	0.917			
SI1	1.75	1.94	0.922	0.927	0.929	0.872
SI2	1.61	2.05	0.953			
SI3	2.03	2.15	0.927			
TR1	2.07	2.5	0.912	0.83	0.865	0.745
TR2	3.46	3.04	0.836			
TR3	2.08	2.18	0.839			

Note: CA = Cronbach’s alpha, CR = composite reliability measure, AVE = average variance extracted.

Table 4. Matrix with Pearson correlations and the square root of the average variance extracted.

	IU	PE	EE	SI	TR
IU	0.906				
PE	0.719	0.872			
EE	0.732	0.816	0.901		
SI	0.713	0.697	0.648	0.934	
TR	0.734	0.861	0.796	0.690	0.863

Note: (a) The square root of AVE is on the main diagonal. (b) All Pearson correlations are significant with $p < 0.001$.

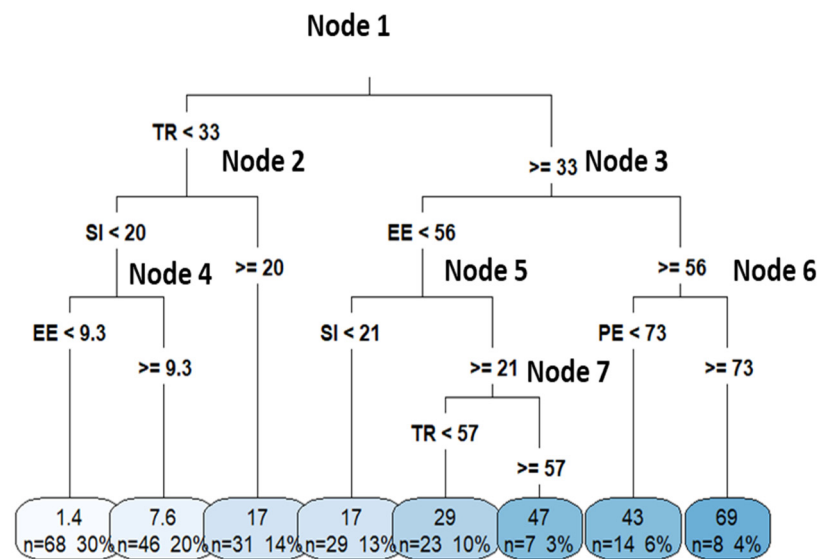


Figure 3. Results of decision tree regression. Note: $R^2 = 69.23\%$, RMSE = 11.010, and MAE = 7.830.

Table 5. Principal splits (first row) and subrogate splits in the decision tree nodes (Figure 3).

Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7
TR < 32.92	SI < 19.81	EE < 56.46	EE < 9.27	SI < 21.29	PE < 72.76	TR < 56.88
PE < 35.81	PE < 29.63	PE < 58.76	PE < 10.56	PE < 25.65	TR < 63.50	PE < 55.96
EE < 33.09	TR < 29.56	TR < 64.70	TR < 1.62	EE < 30.01	EE < 73.30	SI < 63.35
SI < 31.71	EE < 50.09	SI < 50.06	SI < 6.59	TR < 46.49	SI < 51.67	EE < 50.09

Moreover, Figure 3 shows that the coefficient of determination indicates that the DTR explains nearly 70% of the variability in the response variable, which can be considered substantial. However, both the explanatory and predictive capabilities of the DTR can be enhanced by applying RF and XGBoost.

Following the fourth step described in Section 3.3, the best-performing RF model was obtained by tuning the number of variables randomly selected at each split ($mtry = 1$), the total number of trees in the ensemble ($ntree = 100$), and the minimum number of observations required in a terminal node ($nodesize = 1$). Figure 4 shows how the error decreases as the key parameter $ntree$ increases, and that beyond the value of 100, it stabilizes.

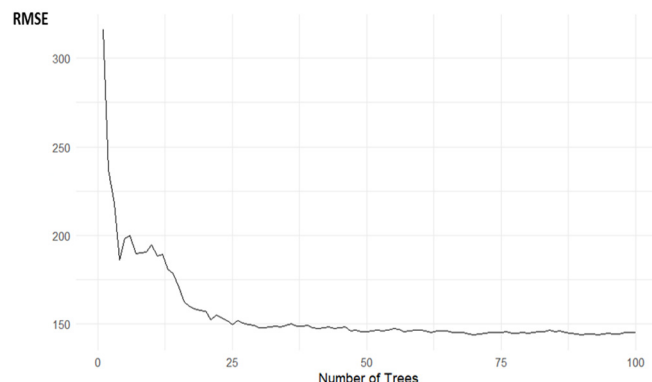


Figure 4. RMSE evolution of the optimal number of trees in the tuning of the random forest model.

Similarly, the optimal XGBoost model was achieved by tuning the learning rate ($eta = 0.1$), the maximum depth of the trees ($max_depth = 4$), the minimum child weight ($min_child_weight = 5$), and the number of boosting rounds ($nrounds = 42$). In both cases, fine-tuning was performed using 10-fold cross-validation, selecting the configuration that minimized the root mean squared error (RMSE). Figure 5 illustrates how the XGBoost error behaves as a function of the hyperparameters eta and $nrounds$.

Table 6 shows that RF, followed by XGBoost, achieves the highest R^2 values and substantially lower error metrics compared to DTR. Furthermore, the results of the Monte Carlo cross-validation presented in Table 7 indicate that RF exhibits the best generalization performance, followed by XGBoost and, lastly, DTR. It is also worth noting that in all cases the Q^2 values exceed 50%, indicating that all decision tree-based methods demonstrate a high level of generalizability [47].

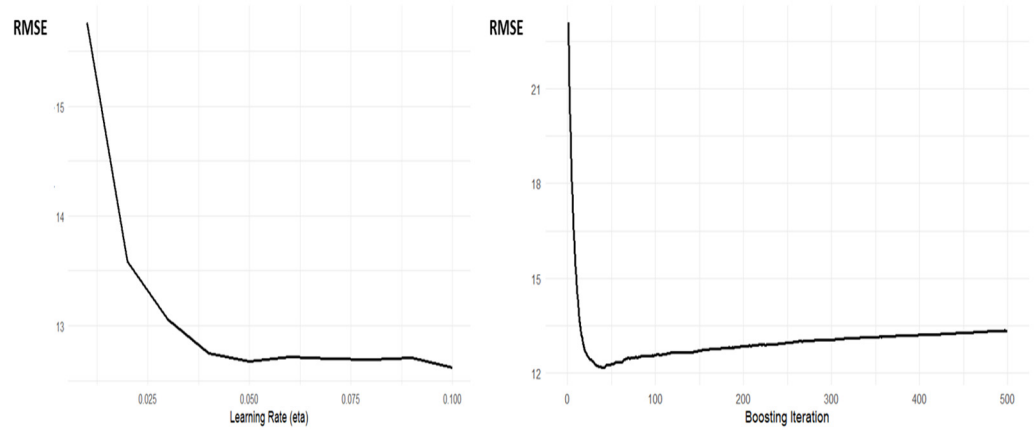


Figure 5. RMSE evolution of the learning rate and boosting iteration in the tuning of the optimal XGBoost model.

Table 6. Adjustment capability metrics of the three machine learning methods used.

	R ²	RMSE	MAE
Decision tree	69.23%	11.010	7.830
Random Forest	95.57%	4.692	3.406
XGBoost	80.95%	9.862	7.004

Table 7. Mean of the performance of predictive metrics of the three machine learning methods used.

	Q ²	RMSE	MAE
Decision tree	51.70%	14.04	9.91
Random Forest	63.40%	12.23	8.55
XGBoost	59.80%	12.9	8.98
ANOVA	F = 27.70 (<i>p</i> < 0.001)	F = 33.96 (<i>p</i> < 0.001)	F = 33.17 (<i>p</i> < 0.001)

4.2. Analysis of Research Question 2

Although Table 7 suggests that the method with the highest predictive performance is RF, we conducted a more in-depth analysis by performing pairwise comparisons of the significance of differences in the prediction metrics, which consistently favored RF. The mean difference analysis presented in Table 8 indicates that, regardless of the metric used, the superior predictive performance of RF compared to the other methods is statistically significant, with a *p*-value < 0.001. Conversely, the method with the poorest predictive performance is DTR. Therefore, the SHAP analysis is based on the random forest fit.

Table 8. Paired-sample *t*-tests for mean differences in predictive metrics.

	Q ²			RMSE			MAE		
	diff	t-Ratio	<i>p</i> -Value	diff	t-Ratio	<i>p</i> -Value	diff	t-Ratio	<i>p</i> -Value
DTR vs. RF	−11.70%	−26.40	<0.001	1.81	27.24	<0.001	1.36	24.32	<0.001
DTR vs. XGBoost	−8.10%	−18.13	<0.001	1.14	14.44	<0.001	0.93	16.39	<0.001
RF vs. XGBoost	3.60%	16.56	<0.001	−0.67	−20.83	<0.001	−0.43	−13.40	<0.001

Note: diff stands for the difference between mean performance metrics in Table 7.

The analysis of SHAP values in the beeswarm plot in Figure 6 reveals that TR is the most influential predictor of chatbot acceptance, showing consistently high contributions to the model’s output, especially when its original values are high. This indicates that as

users’ trust in the chatbot increases, so does the predicted intention to use it—highlighting a strong positive relationship. In contrast, PE exhibits the lowest SHAP values overall, with limited growth in explanatory power even at higher levels, suggesting that perceived usefulness plays a comparatively minor role in shaping user intention. EE and SI occupy an intermediate position. For both variables, SHAP values tend to increase moderately with higher original values, indicating a positive, but less dominant, impact.



Figure 6. Beeswarm plot of the RF adjustment of the conceptual framework in Figure 1.

Table 9 presents the mean absolute SHAP values for the four explanatory variables, along with the results of paired-sample comparisons. While TR displays the highest average absolute SHAP value, the differences between TR, EE, and SI are not statistically significant. However, PE is found to be significantly less relevant than the other three variables. Therefore, although TR appears to be the most important predictor, its contribution is not significantly greater than that of EE and SI. In contrast, the lower relevance of PE is statistically significant when compared to all other predictors.

Table 9. Mean absolute SHAP values for each predictor variable.

Var 1	Var 2	Mean Absolute SHAP (Var 1)	Mean Absolute SHAP (Var 2)	Difference	t-Ratio	p-Value
PE	EE	3.126	3.941	−0.815	−4.591	<0.001
PE	SI	3.126	3.846	−0.719	−3.634	<0.001
PE	TR	3.126	3.985	−0.859	−6.025	<0.001
EE	SI	3.941	3.846	0.096	0.384	0.702
EE	TR	3.941	3.985	−0.044	−0.197	0.844
SI	TR	3.846	3.985	−0.140	−0.715	0.476

Figure 7 presents the importance–performance map constructed based on the eighth step of Section 3 and the SHAP values shown in Table 9. While TR and EE have slightly higher importance than SI, they also exhibit higher performance levels, making their improvement considerably more challenging. In contrast, PE shows the lowest importance combined with the highest performance, placing it in the potential overkill zone. In other words, it clearly does not require immediate attention, and it may even be justified to reallocate efforts away from improving this item in order to focus on enhancing the three more relevant constructs.

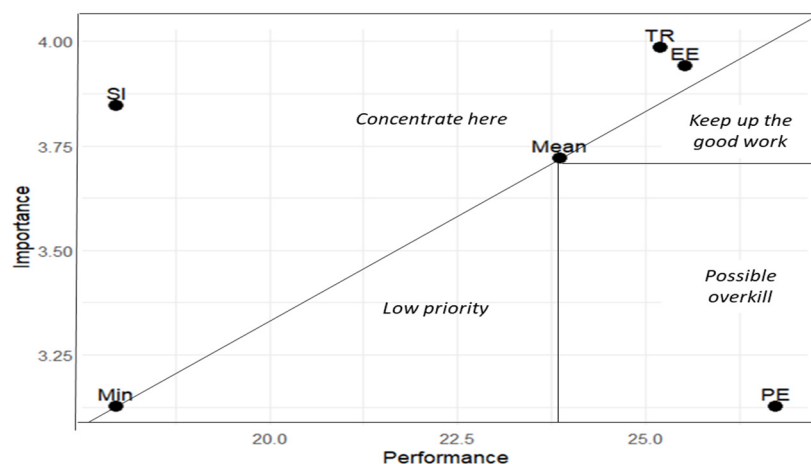


Figure 7. Importance–performance map of the assessed variables to produce acceptance of chatbots.

5. Discussion

5.1. General Considerations

Regarding the first research question (RQ1), we found that the model fit the data well across all machine learning methods, providing a detailed understanding of how the explanatory factors contributed to both acceptance and rejection. While the ensemble decision tree methods yielded better model fit and predictive performance, the simple decision tree regression allowed us to assess the extent to which the hypotheses proposed in Section 2 were supported, as well as to visualize how the explanatory variables interacted to segment the sample into seven distinct user types, ordered by their level of chatbot acceptance. The hypothesized positive relationship between the explanatory variables and acceptance was reflected in the fact that all variables contributed to the partitioning of at least one node of the decision tree, and the resulting partitions suggested a positive association with IU. The results obtained with the decision tree regression were consistent with the fact that all explanatory variables showed a significant correlation with IU in the expected (positive) direction.

In RQ2, we investigated which explanatory variables were most relevant in explaining the intensity of chatbot acceptance. This analysis was conducted using the Shapley additive explanations (SHAP) measure. The results showed that the most influential variables, in order of importance, were TR, EE, SI, and, lastly, PE. It is also worth noting that while the mean absolute SHAP values for the first three variables did not differ significantly from each other, the SHAP value for PE was significantly lower than those of the top three.

The use of DTR enabled a deep understanding of how variables interacted to classify potential chatbot users into different levels of usage—eight, in this case, corresponding to the number of terminal nodes shown in Figure 3. We observed that while TR was used in the initial splits and was therefore decisive in differentiating among all user typologies, SI and EE acted as discriminating factors at low and intermediate levels of acceptance. In contrast, PE only contributed to refining the classification of users who exhibited the highest levels of acceptance.

Trust had a positive and significant influence on behavioral intention. It was, in fact, the variable with the greatest influence on the acceptance of conversational robots. This outcome could be attributed to two key elements: first, the inherent characteristics of the insurance industry, which relies heavily on trust [45], and second, the importance of this concept in the adoption of robotic technologies, making trust a highly important factor in AI-powered Insurtech [13]. Our findings aligned with those from previous studies on

conversational robots in various fields and in countries such as Korea [22], Lebanon [53], and Germany [54].

The positive relationship between EE and IU in the context of the insurer–insured relationship was expected, as convenience could be a relevant factor in the acceptance of chatbots in this area, as reported in Sweden [7]. This result was also in line with studies on customer interactions with chatbots in several countries across Asia, Europe, and North America [53,55–59].

The relevance of SI in influencing IU was consistent with the acceptance of conversational robots by consumers in many cultural contexts and customer service settings, including Korea [22], Lebanon [53], India [56], China [60], and Romania [61].

However, it should be noted that the relevance of PE was secondary compared with the rest of the variables. Regarding PE, this finding could be explained by the fact that the relevance of utilitarian motivations in the use of technologies is paramount when the use of the information system is mandatory [21]. However, chatbot services for customer service should be understood in a multichannel interaction context [43], where their use is optional. In fact, Refs. [54,62] in two different European countries did not observe a significant influence of PE on IU.

The IPMA provided a deeper understanding of the key variables to increase the acceptance of chatbots. The results indicated that social influence, as well as trust and effort expectancy, were not only the most important variables, but also those that, based on their current performance levels, offered the greatest scope for improvement. Therefore, these variables should be prioritized in implementation and improvement strategies. In contrast, performance expectancy, although relevant, was more consolidated among users, requiring secondary attention.

5.2. Theoretical Implications of the Findings in This Paper

We showed that a UTAUT model with four explanatory variables (PE, EE, SI, and TR) explained 70% of the variability in IU using decision tree regression, and this explanatory power increased up to 95% when using random forest. Furthermore, the predictive capacity of the model remained high regardless of the machine learning method applied. This should not be interpreted as decision tree regression being inferior to random forests or XGBoost; rather, these methods belong to the same family and can be used in a complementary manner.

One of the greatest strengths of DTR was its interpretability. Through its structure of nodes and branches, it was possible to understand how an average decision-maker assigned a particular rating or evaluated a level of acceptance, as in the case of adopting technologies such as robots. Although this method did not provide p -values, its interpretation was intuitive. The distribution of observations within the nodes allowed the inference of the direction of the relationship between factors [15]. Moreover, the predictive performance of DTR could be enhanced through decision tree-based techniques such as random forests or XGBoost [63]. To the best of our knowledge, explanatory approaches in consumer behavior analysis were scarce, and within the literature on chatbot acceptance in B2C interactions, virtually non-existent. Therefore, from a methodological standpoint, this study offered a novel perspective based on the use of explainable machine learning techniques.

The use of Shapley additive explanations (SHAP) enabled the quantification of each explanatory variable's importance. This importance was combined with the performance level of each variable to conduct an importance–performance matrix analysis (IPMA), similar to the approach proposed by Ringle and Sarstedt [30], thereby enhancing the explanatory power of partial least squares structural equation modeling. In summary, this study demonstrated that combining machine learning decision tree methods with IPMA

could be highly beneficial for economic and business analyses. This constituted the second methodological contribution of the study, as—although the use of IPMA was common in business analysis—the application of SHAP to quantify variable importance, to the best of our knowledge, represented an original methodological focus.

5.3. Practical Implications of the Findings in This Paper

This has significant implications for the insurance industry. The IPMA results indicate that the variables requiring focused attention for a successful chatbot implementation in the insurer–insured relationship are social influence, trust, and effort expectancy, which are summarized in Table 10 and elaborated on in the following paragraphs. So, regarding improving social perception, effective measures could include:

Table 10. Practical recommendations for enhancing chatbot implementation in insurance.

Focus Area	Recommendation
Social Influence	<ul style="list-style-type: none"> • Humanize the chatbot through empathetic language, a name, and visual identity. • Launch educational campaigns showing benefits and use cases via videos and testimonials.
Trust	<ul style="list-style-type: none"> • Ensure seamless escalation to human agents when needed, with transparent handovers. • Clearly inform users they are interacting with a chatbot and explain its limitations.
Effort Expectancy	<ul style="list-style-type: none"> • Design a simple, intuitive interface with natural language processing and mobile compatibility. • Provide onboarding support (tutorials, tooltips, step-by-step guidance). • Incorporate accessibility features (e.g., voice commands, screen-reader support). • Enable memory of past interactions and reduce redundant questions. • Train human agents to avoid repeating information during handovers. • Offer personalization (pre-filled data, smart suggestions) and communicate convenience benefits.

- Humanizing the chatbot [58], giving it natural, empathetic language, assigning it a name and visual identity to make it more recognizable and friendly, and programming responses that reflect understanding and empathy, especially in sensitive situations like claims management.
- Educating and familiarizing customers with chatbot use. This can be achieved through informative campaigns that share details on how to use the chatbot and its benefits via the insurer’s channels. Videos or interactive guides showing how the chatbot can assist in various processes, along with testimonials from policyholders who have had positive experiences with the system, could also be useful.

Some measures to increase trust in the chatbot include:

- Emphasizing the need for the chatbot to handle complex cases and errors appropriately. It is crucial to implement systems that automatically detect when the chatbot cannot resolve a request and must seamlessly refer the case to a human agent. Furthermore, it is important to clearly explain to users the transition from bot to human to avoid frustrations.
- Ensuring transparency and clear communication between the chatbot and the policyholder. This involves informing users from the start that they are interacting with a bot, clarifying when they will be transferred to a human agent, and ensuring the client understands the chatbot’s capabilities and limitations from the outset.

To improve effort expectancy, the following can be suggested:

- First, simplifying the user interface is essential. The chatbot should offer a clear and intuitive design that guides users through tasks with minimal effort. Leveraging natural language processing (NLP) allows users to interact using everyday language, eliminating the need to learn specific commands. Ensuring compatibility across devices—particularly smartphones—is also key to promoting ease of access.
- In addition, providing onboarding support can greatly reduce perceived effort. Interactive tutorials, embedded tooltips, and step-by-step instructions for common procedures (e.g., filing a claim) help users feel confident from the start. Offering multi-lingual support and using clear, jargon-free language ensure that a broader range of users can engage effectively with the chatbot. Accessibility features, such as voice commands and screen-reader compatibility, should also be incorporated to accommodate users with diverse needs.
- Moreover, the chatbot's functionality should be reliable and consistent. This includes avoiding repetitive requests for the same information, enabling memory of previous interactions, and offering seamless handovers to human agents when needed. In this case the company has to provide training and procedures to the agents to whom the policyholder will be transferred, with the aim of avoiding repetition of information already given and preventing users from feeling that their time is wasted when assisted by a chatbot. Personalization features—such as pre-filled data and smart suggestions—can further reduce user effort. Finally, communicating the benefits of using chatbots, including time savings and convenience, and sharing testimonials from satisfied users, can positively shape expectations and reduce perceived difficulty.

6. Conclusions

6.1. Principal Takeaways

This study offers several insights into the drivers of chatbot acceptance in the insurance sector. It builds on the well-known TAM and UTAUT frameworks, enriched with the construct of trust and analyzed through machine learning techniques based on decision trees.

The first key takeaway is that the proposed model—incorporating performance expectancy (PE), effort expectancy (EE), social influence (SI), and trust (TR)—is both theoretically robust and empirically sound. Using decision tree regression methods, the model explains about 70% of the variability in the intention to use (IU) chatbots. When ensemble methods such as random forest and XGBoost are applied, predictive performance rises sharply, with R^2 values up to 95%. These findings show the relevance of the selected constructs and the effectiveness of decision tree-based methods for capturing complex interaction patterns and improving prediction accuracy in technology adoption research.

The second major insight concerns the relative importance of the explanatory variables. SHAP analysis shows that TR, EE, and SI are the most influential predictors of chatbot acceptance, with no statistically significant differences among them. In contrast, PE—though traditionally a central driver in technology acceptance models—has significantly lower importance, despite a high-performance rating among users. This challenges assumptions inherited from earlier models like TAM and UTAUT, especially in contexts where chatbot use is optional and part of a multichannel service environment.

A third takeaway comes from the importance–performance map analysis (IPMA). TR and EE rank highest in importance, but their strong performance levels suggest that further improvements could be costly and bring diminishing returns. SI, however, shows high importance and lower performance, making it a priority for strategic action. PE falls into the “overkill” zone—low importance but high performance—suggesting that further enhancement may be unnecessary or inefficient at this stage.

From an analytical perspective, we found that the use of DTR provides an empirical view of how the explanatory variables of a phenomenon—in this case, PE, EE, SI, and TR—interact to produce acceptance or rejection of chatbots. In our hypothesis development, we did not specify mediated or moderated relationships; instead, these are empirically uncovered through the DTR model. Moreover, the consideration of surrogate splits offers a deeper understanding of the interaction between explanatory variables and the direction of their influence on the outcome. To the best of our knowledge, this application of DTR has not yet been leveraged in consumer behavior studies focused on the acceptance of new communication channels with firms.

Finally, the study confirms that combining machine learning techniques with explainability tools such as SHAP and managerial instruments like IPMA provides a powerful and integrated approach to understanding user behavior in technology adoption contexts. This methodological synergy not only improves predictive accuracy, but also bridges the gap between complex algorithmic outputs and practical and intuitive managerial interpretation. By translating model results into interpretable and strategically relevant insights, the approach facilitates more informed decision-making regarding technology design, communication strategies, and resource allocation. Furthermore, it demonstrates the potential of hybrid methodologies to move beyond mere prediction and toward prescriptive guidance for implementation, especially in settings where user acceptance is critical to success. The integration of SHAP and IPMA proves especially valuable in dynamic business environments where decision-making must rely on real-time diagnostics rather than historical patterns, which may hold limited relevance. This is particularly true in the context of emerging technologies—such as AI-powered chatbots—where rapid innovation outpaces the applicability of past data and requires timely, actionable insights to guide effective implementation.

In the case of chatbot deployment in insurance services, and in the moment of the study, efforts should be directed primarily toward improving users' trust, perceived ease of use, and social perceptions, rather than focusing solely on performance-related expectations. This nuanced approach is essential for increasing user acceptance and optimizing the integration of artificially driven tools in customer service.

6.2. Limitations and Future Research Directions

We recognize the constraints of this empirical research. This study was carried out in a specific territory, Spain, with the majority of answers gathered from platforms like LinkedIn. Users of these platforms tend to have higher education and professional experience, typically ranging from mid-level management to executive roles. As a result, the educational and economic backgrounds of the participants may influence our findings on the behavioral intent to use chatbots. The study of other social groups and cultures may yield completely different results. A study on the acceptance of ChatGPT by Zoomers for general uses in Croatia shows, unlike our study, that performance expectancy is the most relevant variable for chatbot acceptance [59].

Therefore, caution is advised when generalizing our results to policyholders from other cultures or those with professional and educational profiles that differ from the sample group. To draw broader conclusions, it would be essential to include a more diverse range of countries and socio-economic profiles of the respondents.

It should also be noted that the study focuses on a specific economic sector and a very particular type of customer service: the management of in-force insurance policies. Extrapolating the findings of this study to other sectors (e.g., non-financial industries) or even to different contexts within the insurance sector—such as providing advice on future contracts—should be approached with caution. Within the insurance domain, this study

could be extended to other potential services offered to customers, such as suggesting new products to existing policyholders or assisting individuals interested in initiating new contracts with the company.

While the RF model demonstrated a high explanatory capacity (R^2), it is important to acknowledge the potential risk of model overfitting, particularly in complex, non-parametric algorithms. Although the Monte Carlo cross-validation procedure mitigates this concern by evaluating predictive performance on multiple holdout samples, the possibility of overly optimistic estimates cannot be entirely ruled out. It should also be noted that RF achieved a high predictive capacity ($Q^2 > 50\%$)—substantially higher than that of DTR—yet still notably lower than its own explanatory capacity. Future research could further address this issue by testing alternative regularization techniques, tuning strategies, or simpler model specifications to confirm the robustness of the findings.

The analysis in this paper is based on a cross-sectional survey, meaning the conclusions cannot be extended to long-term trends and are limited to a specific geographical area (Spain). However, the data collection period is consistent with other cross-sectional studies, which range from half a month in Croatia [59], one month in Canada [38], two months in Romania [34,61], to three months [38] in Canada. The findings therefore represent a snapshot tied to a specific moment in the introduction and development of chatbot technology within a particular industry, namely the insurance sector. These results are particularly useful for informing decisions in contexts similar to the one studied, as earlier stages of chatbot development and market penetration are not directly comparable to the current landscape.

The fields of artificial intelligence and Insurtech are evolving at a rapid pace, and public perceptions of emerging technologies are highly dynamic. A more comprehensive understanding would require comparable studies conducted at different stages of chatbot evolution. While longitudinal approaches—spanning, for example, a decade—may be well-suited for cultural or ethnographic research, they may be less relevant for managerial decision-making in fast-moving environments where the technology is still undergoing rapid growth and has not yet reached full consolidation.

Author Contributions: Conceptualization: J.d.A.-S. and J.G.-A.; methodology: J.d.A.-S.; validation: J.G.-A.; formal analysis: J.d.A.-S.; investigation: J.d.A.-S. and J.G.-A.; resources: J.d.A.-S.; data curation: J.G.-A.; writing—original draft preparation: J.d.A.-S.; writing—review and editing: J.G.-A.; visualization: J.G.-A.; supervision: J.G.-A.; project administration: J.G.-A.; funding acquisition: J.d.A.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Telefonica and the Telefonica Chair on Smart Cities of the Universitat Rovira i Virgili and Universitat de Barcelona (project number 42.DB.00.18.00).

Institutional Review Board Statement: (1) All participants received detailed written information about the study and procedure; (2) no data directly or indirectly related to the health of the subjects were collected, and therefore the Declaration of Helsinki was not mentioned when informing the subjects; (3) the anonymity of the collected data was ensured at all times; (4) the research received a favorable evaluation from the Ethics Committee of the researchers' institution (CEIPSA-2022-PR-0005).

Informed Consent Statement: All respondents gave permission for the processing of their responses for the content of this publication.

Data Availability Statement: The data supporting the analysis is available at <https://doi.org/10.7910/DVN/LK4LAT> (accessed on 10 August 2025).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DTR	Decision Tree Regression
EE	Effort Expectancy
IPMA	Importance–Performance Map Analysis
PE	Performance Expectancy
RF	Random Forest
SHAP	Shapley Additive Explanations
SI	Social Influence
TAM	Technology Acceptance Model
TR	Trust
UTAUT	Unified Theory of Acceptance and Use of Technology
XGBoost	Extreme Gradient Boosting

Appendix A

Table A1. Items of latent variables assessed in this study.

Items
Intention to Use (IU)
IU1. I intend to be assisted by chatbots.
IU2. I predict that I will use a service managed by chatbots.
IU3. I will opt for management carried out by chatbots.
Performance Expectancy (PE)
PE1. The use of chatbots can be useful for managing my claims.
PE2. Using chatbots will make it easier for me to report my claims.
PE3. Using chatbots is useful and will allow me to receive compensations I am entitled to more quickly.
PE4. Using chatbots is useful and will allow me to manage my claims with less effort and fewer undesired effects (such as errors made by the insurance company’s agent).
PE5. Using chatbots allows the insurance company to offer better service to customers at lower costs.
Effort Expectancy (EE)
EE1. It will be easy for me to adapt to using chatbots in my dealings with my insurer.
EE2. It will be easier to manage my claims with the existence of chatbots.
EE3. It will be easy for me to use the channels provided by the insurer for communication if they are managed by chatbots.
Social influence (SI)
SI1. The people who are important to me believe that using chatbots facilitates the claims process.
SI2. The people who influence me believe that, if I could choose a claims channel, I should opt for one that uses chatbots.
SI3. The people whose opinions I value believe that using chatbots in insurance management by the insured is an advance.
Trust (TR)
TR1. The use of chatbots in my relationship with the insurer gives me trust.
TR2. The use of chatbots makes it easier for the insurer to fulfil its commitments and obligations.
TR3. In managing claims through chatbots, the interests of the insured are taken into account.

References

1. Tamvada, J.P.; Narula, S.; Audretsch, D.; Puppala, H.; Kumar, A. Adopting New Technology Is a Distant Dream? The Risks of Implementing Industry 4.0 in Emerging Economy SMEs. *Technol. Forecast. Soc. Change* **2022**, *185*, 122088. [[CrossRef](#)]
2. He, M. Fintech 4.0 and Financial Systems. In *Innovation, Sustainability, and Technological Megatrends in the Face of Uncertainties: Core Developments and Solutions*; Turi Abeba, N., Lekhi, P., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 55–72. [[CrossRef](#)]
3. Mhlanga, D. Industry 4.0 in Finance: The Impact of Artificial Intelligence (AI) on Digital Financial Inclusion. *Int. J. Financ. Stud.* **2020**, *8*, 45. [[CrossRef](#)]
4. Nicoletti, B. Industry 4.0 and Insurance 4.0. In *Insurance 4.0: Benefits and Challenges of Digital Transformation*; Springer International Publishing: Cham, Switzerland, 2021; pp. 11–40. [[CrossRef](#)]
5. Sosa, I.; Montes, Ó. Understanding the InsurTech Dynamics in the Transformation of the Insurance Sector. *Risk Manag. Insur. Rev.* **2022**, *25*, 35–68. [[CrossRef](#)]
6. Nguyen, D.M.; Chiu, Y.-T.H.; Le, H.D. Determinants of Continuance Intention toward Banks' Chatbot Services in Vietnam: A Necessity for Sustainable Development. *Sustainability* **2021**, *13*, 7625. [[CrossRef](#)]
7. Gebert-Persson, S.; Gidhagen, M.; Sallis, J.E.; Lundberg, H. Online Insurance Claims: When More than Trust Matters. *Int. J. Bank Mark.* **2019**, *37*, 579–594. [[CrossRef](#)]
8. Andrés-Sánchez, J.; Gené-Albesa, J. Explaining Policyholders' Chatbot Acceptance with an Unified Technology Acceptance and Use of Technology-Based Model. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 1217–1237. [[CrossRef](#)]
9. Eckert, C.; Neunsinger, C.; Osterrieder, K. Managing Customer Satisfaction: Digital Applications for Insurance Companies. *Geneva Pap. Risk Insur. Issues Pract.* **2022**, *47*, 569–602. [[CrossRef](#)]
10. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [[CrossRef](#)]
11. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* **2003**, *27*, 425–478. [[CrossRef](#)]
12. Alsharhan, A.; Al-Emran, M.; Shaalan, K. Chatbot Adoption: A Multiperspective Systematic Review and Future Research Agenda. *IEEE Trans. Eng. Manag.* **2023**, *71*, 10232–10244. [[CrossRef](#)]
13. Zarifis, A.; Cheng, X. A Model of Trust in Fintech and Trust in Insurtech: How Artificial Intelligence and the Context Influence It. *J. Behav. Exp. Financ.* **2022**, *36*, 100739. [[CrossRef](#)]
14. Gatzoufa, P.; Saprikis, V. A Literature Review on Users' Behavioral Intention toward Chatbots' Adoption. *Appl. Comput. Inform.* **2022**. [[CrossRef](#)]
15. Loh, W.-Y. Classification and Regression Trees. *WIREs Data Min. Knowl. Discov.* **2011**, *1*, 14–23. [[CrossRef](#)]
16. Imani, M.; Arabnia, H.R. Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis. *Technologies* **2023**, *11*, 167. [[CrossRef](#)]
17. Imani, M.; Beikmohammadi, A.; Arabnia, H.R. Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS under Varying Imbalance Levels. *Technologies* **2025**, *13*, 88. [[CrossRef](#)]
18. Chung, D.; Jeong, P.; Kwon, D.; Han, H. Technology Acceptance Prediction of Robo-Advisors by Machine Learning. *Intell. Syst. Appl.* **2023**, *18*, 200197. [[CrossRef](#)]
19. Richter, N.F.; Tudoran, A.A. Elevating Theoretical Insight and Predictive Accuracy in Business Research: Combining PLS-SEM and Selected Machine Learning Algorithms. *J. Bus. Res.* **2024**, *173*, 114453. [[CrossRef](#)]
20. Cuc, L.D.; Rad, D.; Cilan, T.F.; Gomoi, B.C.; Nicolaescu, C.; Almasi, R.; Pandelica, I. From AI Knowledge to AI Usage Intention in the Managerial Accounting Profession and the Role of Personality Traits—A Decision Tree Regression Approach. *Electronics* **2025**, *14*, 1107. [[CrossRef](#)]
21. Venkatesh, V.; Davis, F.D. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Manag. Sci.* **2000**, *46*, 186–204. [[CrossRef](#)]
22. Han, J.; Conti, D. The Use of UTAUT and Post Acceptance Models to Investigate the Attitude towards a Telepresence Robot in an Educational Setting. *Robotics* **2020**, *9*, 34. [[CrossRef](#)]
23. de Andrés-Sánchez, J.; Gené-Albesa, J. Not with the Bot! The Relevance of Trust to Explain the Acceptance of Chatbots by Insurance Customers. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 110. [[CrossRef](#)]
24. Kasilingam, D.L. Understanding the Attitude and Intention to Use Smartphone Chatbots for Shopping. *Technol. Soc.* **2020**, *62*, 101280. [[CrossRef](#)]
25. Pitardi, V.; Marriott, H.R. Alexa, She's Not Human But... Unveiling the Drivers of Consumers' Trust in Voice-Based Artificial Intelligence. *Psychol. Mark.* **2021**, *38*, 626–642. [[CrossRef](#)]
26. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
27. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]

28. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
29. Abalo, J.; Varela, J.; Manzano, V. Importance Values for Importance–Performance Analysis: A Formula for Spreading out Values Derived from Preference Rankings. *J. Bus. Res.* **2007**, *60*, 115–121. [[CrossRef](#)]
30. Ringle, C.M.; Sarstedt, M. Gain More Insight from Your PLS-SEM Results. *Ind. Manag. Data Syst.* **2016**, *116*, 1865–1886. [[CrossRef](#)]
31. DeAndrade, I.M.; Tumelero, C. Increasing Customer Service Efficiency through Artificial Intelligence Chatbot. *Rev. Gest.* **2022**, *29*, 238–251. [[CrossRef](#)]
32. Standaert, W.; Muylle, S. Framework for Open Insurance Strategy: Insights from a European Study. *Geneva Pap. Risk Insur. Issues Pract.* **2022**, *47*, 643–668. [[CrossRef](#)] [[PubMed](#)]
33. Gené-Albesa, J. Interaction Channel Choice in a Multichannel Environment, An Empirical Study. *Int. J. Bank Mark.* **2007**, *25*, 490–506. [[CrossRef](#)]
34. Alt, M.A.; Vizeli, I.; Säpläcan, Z. Banking with a Chatbot—A Study on Technology Acceptance. *Stud. Univ. Babeş Bolyai Oeconomica* **2021**, *66*, 13–35. [[CrossRef](#)]
35. Huang, S.Y.; Lee, C.-J.; Lee, S.-C. Toward a Unified Theory of Customer Continuance Model for Financial Technology Chatbots. *Sensors* **2021**, *21*, 5687. [[CrossRef](#)] [[PubMed](#)]
36. Shaikh, I.A.K.; Khan, S.; Faisal, S. Determinants Affecting Customer Intention to Use Chatbots in the Banking Sector. *Innov. Mark.* **2023**, *19*, 257–268. [[CrossRef](#)]
37. Toh, T.-J.; Tay, L.-Y. Banking Chatbots: A Study on Technology Acceptance among Millennials in Malaysia. *J. Logist. Inform. Serv. Sci.* **2022**, *9*, 1–15. [[CrossRef](#)]
38. PromTep, S.; Arcand, M.; Rajaobelina, L.; Ricard, L. From What Is Promised to What Is Experienced with Intelligent Bots. In *Advances Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC)*; Springer International Publishing: Cham, Switzerland, 2021; Volume 1, pp. 560–565. [[CrossRef](#)]
39. Rodríguez-Cardona, D.; Janssen, A.; Guhr, N.; Breitner, M.H.; Milde, J. A Matter of Trust? Examination of Chatbot Usage in Insurance Business. In *Proceedings of the 54th Hawaii International Conference on System Sciences, Honolulu, HI, USA, 5–8 January 2021*; pp. 556–565. [[CrossRef](#)]
40. Vassilakopoulou, P.; Haug, A.; Salvesen, L.M.; Pappas, I.O. Developing Human/AI Interactions for Chat-Based Customer Services: Lessons Learned from the Norwegian Government. *Eur. J. Inf. Syst.* **2023**, *32*, 10–22. [[CrossRef](#)]
41. Rajaobelina, L.; PromTep, S.; Arcand, M.; Ricard, L. Creepiness: Its Antecedents and Impact on Loyalty When Interacting with a Chatbot. *Psychol. Mark.* **2021**, *38*, 2339–2356. [[CrossRef](#)]
42. Hasan, S.; Godhuli, E.R.; Rahman, S.; Mamun, A.A. The Adoption of Conversational Assistants in the Banking Industry: Is the Perceived Risk a Moderator? *Heliyon* **2023**, *9*, e20220. [[CrossRef](#)]
43. Andrés-Sánchez, J.; Gené-Albesa, J. Assessing Attitude and Behavioral Intention toward Chatbots in an Insurance Setting: A Mixed Method Approach. *Int. J. Hum. Comput. Interact.* **2023**, *40*, 4918–4933. [[CrossRef](#)]
44. Van Pinxteren, M.M.E.; Pluymaekers, M.; Lemmink, J.G.A.M. Human-like Communication in Conversational Agents: A Literature Review and Research Agenda. *J. Serv. Manag.* **2020**, *31*, 203–225. [[CrossRef](#)]
45. Guiso, L. Trust and Insurance. *Geneva Pap. Risk Insur. Issues Pract.* **2021**, *46*, 509–512. [[CrossRef](#)]
46. Andrade, C. The Inconvenient Truth About Convenience and Purposive Samples. *Indian J. Psychol. Med.* **2020**, *43*, 86–88. [[CrossRef](#)] [[PubMed](#)]
47. Hair, J.F.; Risher, J.J.; Sarstedt, M.; Ringle, C.M. When to Use and How to Report the Results of PLS-SEM. *Eur. Bus. Rev.* **2019**, *31*, 2–24. [[CrossRef](#)]
48. Faul, F.; Erdfelder, E.; Buchner, A.; Lang, A.-G. Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses. *Behav. Res. Methods* **2009**, *41*, 1149–1160. [[CrossRef](#)]
49. Morgan, R.M.; Hunt, S.D. The Commitment-Trust Theory of Relationship Marketing. *J. Mark.* **1994**, *58*, 20–38. [[CrossRef](#)]
50. Probst, P.; Wright, M.N.; Boulesteix, A.-L. Hyperparameters and Tuning Strategies for Random Forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [[CrossRef](#)]
51. Kuhn, M.; Johnson, K. Over-Fitting and Model Tuning. In *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013; pp. 61–92. [[CrossRef](#)]
52. Alhazeem, E.; Alsobeh, A.; Al-Ahmad, B. Enhancing Software Engineering Education through AI: An Empirical Study of Tree-Based Machine Learning for Defect Prediction. In *SIGITE '24: Proceedings of the 25th Annual Conference on Information Technology Education*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 153–156. [[CrossRef](#)]
53. Mostafa, R.B.; Kasamani, T. Antecedents and Consequences of Chatbot Initial Trust. *Eur. J. Mark.* **2022**, *56*, 1748–1771. [[CrossRef](#)]
54. Gansser, O.A.; Reich, C.S. A New Acceptance Model for Artificial Intelligence with Extensions to UTAUT2: An Empirical Study in Three Segments of Application. *Technol. Soc.* **2021**, *65*, 101535. [[CrossRef](#)]
55. Joshi, H. Integrating Trust and Satisfaction into the UTAUT Model to Predict Chatbot Adoption—A Comparison between Gen-Z and Millennials. *Int. J. Inf. Manag. Data Insights* **2025**, *5*, 100332. [[CrossRef](#)]

56. Goli, M.; Sahu, A.K.; Bag, S.; Dhamija, P. Users' Acceptance of Artificial Intelligence-Based Chatbots: An Empirical Study. *Int. J. Technol. Hum. Interact.* **2023**, *19*, 18. [[CrossRef](#)]
57. Liu, M.; Yang, Y.; Ren, Y.; Jia, Y.; Ma, H.; Luo, J.; Fang, S.; Qi, M.; Zhang, L. What Influences Consumer AI Chatbot Use Intention? An Application of the Extended Technology Acceptance Model. *J. Hosp. Tour. Technol.* **2024**, *15*, 667–689. [[CrossRef](#)]
58. Akram, S.; Buono, P.; Lanzilotti, R. Recruitment Chatbot Acceptance in a Company: A Mixed Method Study on Human-Centered Technology Acceptance Model. *Pers. Ubiquitous Comput.* **2024**, *28*, 961–984. [[CrossRef](#)]
59. Biloš, A.; Budimir, B. Understanding the Adoption Dynamics of ChatGPT among Generation Z: Insights from a Modified UTAUT2 Model. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 863–879. [[CrossRef](#)]
60. Xie, C.; Wang, Y.; Cheng, Y. Does Artificial Intelligence Satisfy You? A Meta-Analysis of User Gratification and User Satisfaction with AI-Powered Chatbots. *Int. J. Hum. Comput. Interact.* **2024**, *40*, 613–623. [[CrossRef](#)]
61. Iancu, I.; Iancu, B. Interacting with Chatbots Later in Life: A Technology Acceptance Perspective in COVID-19 Pandemic Situation. *Front. Psychol.* **2023**, *13*, 1111003. [[CrossRef](#)] [[PubMed](#)]
62. de Andrés-Sánchez, J.; Gené-Albesa, J. Drivers and Necessary Conditions for Chatbot Acceptance in the Insurance Industry. Analysis of Policyholders' and Professionals' Perspectives. *J. Organ. Comput. Electron. Commer.* **2024**, 1–28. [[CrossRef](#)]
63. Ngai, E.W.T.; Wu, Y. Machine Learning in Marketing: A Literature Review, Conceptual Framework, and Research Agenda. *J. Bus. Res.* **2022**, *145*, 35–48. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.