scientific reports

Check for updates

OPEN

# Elicitation of security threats and vulnerabilities in Insurance chatbots using STRIDE

Zilungile Bokolo[1] & Olawande Daramola[2]✉

**Although chatbots are used a lot for customer relationship management (CRM), there needs to be more data security and privacy control strategies in chatbots, which has become a security concern for financial services institutions. Chatbots gain access to large amounts of vital company information and clients' personal information, which makes them a target of security attacks. The loss of data stored in chatbots can cause major harm to companies and customers. In this study, STRIDE (viz. Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege) modelling was applied to identify the data security vulnerabilities and threats that pertain to chatbots used in the insurance industry. To do this, we conducted a case study of a South African insurance organisation. The adopted methodology involved data collection from stakeholders in the insurance organisation to identify chatbot use cases and understand chatbot operations. After that, we conducted a STRIDE-based analysis of the chatbot use cases to elicit security threats and vulnerabilities in the insurance chatbots in the organisation. The results reveal that security vulnerabilities associated with Spoofing, Denial of Service, and Elevation of privilege are more relevant to insurance chatbots. The most security threats stem from Tampering, Elevation of privilege, and Spoofing. The study extends the discussion on chatbot security. It fosters an understanding of security threats and vulnerabilities that pertain to insurance chatbots, which is beneficial for security researchers and practitioners working on the security of chatbots and the insurance industry.**

**Keywords** Data security, Chatbots, Chatbot security, STRIDE, Threat modelling, Insurance, Customer relationship management, Artificial Intelligence

The insurance industry provides a wide range of products that allow society to function safely by sharing the risk. These products include short-term (property and vehicle) and long-term (life insurance) insurance, which would then cover both commercial and personal clients for their loss at the claims stage. Insurers have been facing increasing competition from other industries such as banks, mutual funds, and investment advisory firms, which offer the same products[1,2]. For insurance companies to avoid losing business chances to contenders, they need to change how they are associated with their clients by embracing the utilisation of chatbots. Previously, insurers used Customer Relationship Management (CRM) systems to manage their customer relationships. CRM working frameworks typically centre around improving productivity and guaranteeing authority over the client by causing them to feel remembered for the business activities[3]. Chatbots are increasingly adopted for various aspects of CRM in the insurance industry, helping insurance companies to minimise customer service costs, making the response time for clients' requests very fast, and replying to up to 80% of routine inquiries[4]. However, there are issues concerning the security and privacy of data transmitted between customers and chatbots, including customer information and the company's operation information[5]. Chatbots access a lot of company and client information, which makes the chatbot the target of hackers and other security attacks on the web, which raises security concerns regarding the use of chatbots[5].

Companies, including insurance and banks, have taken the exciting next step in technology enhancement by implementing artificial intelligence (AI) for their business activities[6,7]. A chatbot is an AI and customer service platform where clients can access information, help quickly, and provide continuous customer service[7]. Every month, over 3 billion people use chatbots directly or indirectly. Organisations have invested resources into chatbot innovation to ensure competitive advantage by improving client support and reducing expenses by 40%[8]. According to the Grand View research reports, the global chatbot market will reach 1.23 billion US dollars

[1]Department of Information Technology, Cape Peninsula University of Technology, Cape Town, South Africa. [2]Department of Informatics, University of Pretoria, Pretoria, South Africa. ✉email: wande.daramola@up.ac.za

in output value in 2025[9]. In the case of the insurance industry, adopting service innovations built on chatbot technology can provide more benefits like improving the productivity around the insurance value chain, lowering costs and producing more client loyalty and trust[10]. While there is much excitement about conversational AI technologies like chatbots, there is less understanding of how to secure these new technologies[11].

Data security of new technologies like AI and chatbots has become a huge concern[12,13]. The leakage or alteration of data can be intentional, and companies may be punished or held criminally liable for such incidents. Thus, the protection of users' sensitive data is a major information security objective in chatbots[14,15]. According to Ref.[15], the security threats associated with chatbots can stem from malicious input, user profiling, contextual attacks, and data breaches. However, these security threats can be mitigated by using mechanisms such as blockchain technology, end-to-end encryption, and organisational controls. There is also the need to maintain user trust and confidence in chatbots because a security breach or data leak can adversely affect user trust, which can have a damaging impact on a business or an organisation. So far, financial chatbots have been mostly affected by security issues in the areas of insecure authentication, data integrity, system availability, transparency, and privacy. Thus, the management of the security and privacy vulnerabilities of the chatbots is critical[16,17]. Hence, a thorough precautionary analysis of the security and privacy vulnerabilities of a chatbot used in the financial sector should be undertaken prior to deployment.

Most studies on technology adoption in financial services focus on Internet banking customers, with limited research on the insurance industry[18]. According to Ref.[19], trust is important, but other factors, such as privacy concerns and perceived usefulness, are also critical for insurance chatbot usage. Also, Ref.[20] observed that security and integration are challenges for conversational agents in the insurance industry; thus, the issue of privacy and integrity of the data in insurance chatbots should be an active research area[13].

This paper used STRIDE[21,22] as a threat modelling technique to identify the possible threats and vulnerabilities of chatbots in the insurance industry. STRIDE was selected because it is a well-established and prominent threat modelling technique. It enables a focused approach to threat elicitation by covering six specific aspects[23,24]. It is easy to use and the most mature out of all the threat modelling methods[25]. STRIDE, which stands for Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, or Elevation of privilege, is one of the most prominent threat elicitation techniques used in the early phases of software development to identify the security requirements of a system[22]. Our case study is a short-term insurance company based in South Africa that utilises two types of chatbots. These are the iAssist chatbot and the WhatsApp chatbot. The iAssist chatbot is used internally within the organisation for interaction with the Human Resource department, while the WhatsApp chatbot is used for customer engagement. Companies need to improve their level of cybersecurity by successfully adopting a suitable cybersecurity strategy to attain the level of cybersecurity necessary to protect the business, staff, clients, and reputation[26].

So far in the literature, none of the previous studies on the security of chatbots have considered using STRIDE for threat elicitation or focussed on the insurance industry. Thus, more understanding of insurance chatbots' security threats and vulnerabilities is needed. Therefore, this paper makes the following contributions.

- It extends the discussion on the subject of chatbot security, which is an emerging topic.
- It is the first study that has considered threat modelling for the insurance chatbots.
- It demonstrates the application of STRIDE modelling for chatbot data security.
- It provides an understanding of security threats and vulnerabilities of chatbots used in the insurance industry that have not been explored before.

The rest of this paper is organised as follows: "Background and related work" presents the literature review, comprising theoretical background and a review of related work. Section "Methods" describes the methodology employed for this study, while "STRIDE modelling" contains the results. The discussion of results is presented in "Discussion of findings", while the paper is concluded in "Recommendations" with a summary and our perspective on future work.

## Background and related work

This section reviews the literature, covering the theoretical background and previous studies.

### The insurance industry

The insurance industry is understood and known as a subdivision of financial services[27,28]. In South Africa, insurance companies are divided into long-term or life insurance and short-term property or car insurance[29]. Good customer relationship management in the insurance industry is important as it helps to retain existing customers, which can be done effectively by adopting advanced technologies[30]. Although customers can go directly to the insurance company, insurance companies often use brokers or agents as intermediaries between them and customers. Brokers always work closer to the clients to help them understand the products offered by the insurance. When the customer decides on the product, the brokers do a risk assessment and underwriting[27]. Therefore, intermediaries or brokers perform a considerable role in enabling value to insurance clients, and they need support from the insurance companies by providing supportive technologies like a chatbot to improve the client's experience and retail more policies[31]. Chatbots access customer information during interaction with customers. Customer data contained in basic queries such as current claims status, account information, and policy information gets transmitted via chatbots[32]. Hence, the risk of a security breach or data leak as a result of cyberattacks, such as identity theft, spoofing, tampering, loss of confidentiality, information compromise, denial of service, and information disclosure, exists. Chatbots, built on AI, are evolving as a technology[31], which call centres have employed to support brokers. The implemented chatbot aims to improve the call centre agent's

performance by answering client questions, which also helps the organisation deliver a better service than their competitors[33]. Although there is excitement about adopting these new technologies in the insurance industry, security protection is a significant concern, making insurance companies vulnerable to security breaches such as claims fraud[34].

## Chatbots security

Chatbots are mostly accessible through different platforms of messenger apps such as Facebook and Skype, and there is no proper security implementation on these platforms. Electronic Frontier Foundation (EFF) Secure Messaging Scorecard shows that five of seven proven measurements are not secured by Facebook Messenger and eight other messenger platforms. Figure 1 shows the security scores on different messenger apps. The topic was presented at the Privacy Week Conference in Vienna, and the title of the talk was "Privacy and Data Security of Chatbots" and "Why you shouldn't talk to your chatbot about everything"[35]. WhatsApp is the most secure messenger app and provides end-to-end encryption, but should there be any failure, hackers can get the data between users sharing the same network because they can sniff and steal each other's credentials. Messenger also does not confirm the user's identity, e.g. send OTP. Previous chats are not hidden, so if the hackers perform malicious attacks, they can steal the credentials. Lastly, the security design is poorly documented in these messenger apps[36].


**Figure 1.** Electronic frontier foundation (EFF) secure messaging scorecard.

Between February and June 2018, a data breach occurred in Ticketmaster's global customer base, which was discovered on 23 June. Malicious software was in their chatbot, gathering information and sending it to a third party. Due to the compromise of the chatbot, customers' confidential information, including payment information, was stolen[37].

So far, very few studies have focussed on the data security of insurance chatbots. In a study by Ref.[20] that investigated the potential use cases of conversational agents in insurance companies, it was discovered that security and integration issues are among the challenges faced by new conversational agents like chatbots. Also, the need for effective mechanisms that foster trust and address privacy concerns has been emphasised by Ref.[19]. Based on a survey study (involving 215 participants) that examined chatbot usage in the insurance business in Germany, it was discovered that apart from trust and privacy concerns, other factors such as perceived usefulness and perceived ease of use are critical for insurance chatbot usage[19].

## Chatbot architecture

Architecture is a fundamental structure that can be utilised to build a wide range of applications. Now that chatbots are rapidly adopted, organisations must consider the relevant architectural approach when implementing them. Chatbot architecture has been categorised into three[38]: (1) A personal assistant, (2) A customer service bot, and (3) A functional bot. This study focuses more on functional bots since specialist bots are used more in the insurance industry. The typical architecture of a functional bot is shown in Fig. 2.

In Fig. 2, the presentation layer enables user interaction (UI), which must be done in the correct order. After processing, the data has to be structured according to the user association. The service layer gives access to inner and outside information, business usefulness, middleware availability, and different administrations. Chatbot solutions are highly vulnerable because they are exposed to many frameworks, channels, and systems[38]. An insurance chatbot based on the functional bot architecture will facilitate interaction with customers who may access the insurance company's platform on different types of client devices such as mobile phones, web, or social media tools like WhatsApp and WeChat through its presentation layer. User requests will be made from the Presentation layer while corresponding constructed responses will be sent as notifications to the user through the same presentation layer when a user sends a request, which could be any of the following: (1) opening a new account; (2) processing a claim; (3) requesting an insurance a policy quote; (4) changing to a new policy; or (5) reports. The request will be sent to the Business layer, where the appropriate business logic to perform the selected operation will be invoked. The chatbot will initiate the required backend operations, such as retrieving the needed data from the data layer and invocation of the necessary internal and external services from the Service layer to implement the user's request.

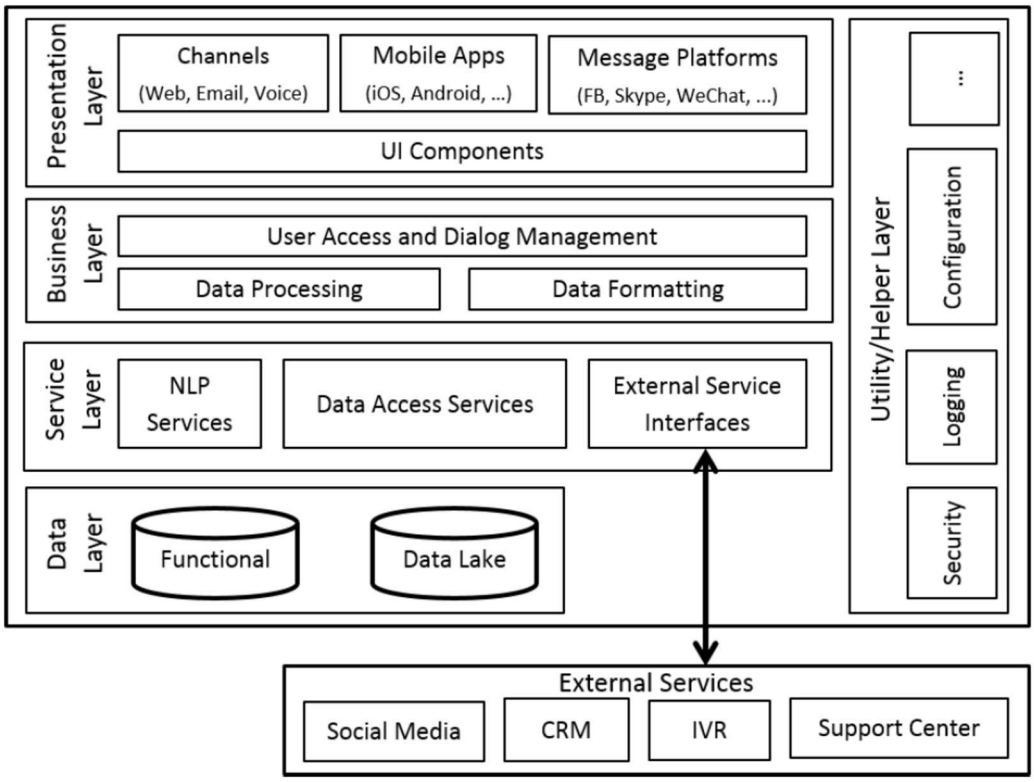


**Figure 2.** The typical architecture of a functional bot[38].

## Threat modelling

The prevalence of cyber attacks on computer systems has made the topic of cybersecurity increasingly relevant[26,39]. No computer system is exempt from cybersecurity attacks, which exist in the form of internal and external security threats. Threat modelling has been proposed as a solution for secure application development and system security evaluations. Threat modelling facilitates secure application development and provides a framework for security assessments. Its goal is to be progressively proactive and make it increasingly hard for aggressors to achieve malicious intent[39].

The threat modelling process includes identifying security threats in the application and devising mitigation activities. Examples of threat modelling methodologies and techniques include STRIDE, Abuser stories, Stride average model, Attack trees, Fuzzy logic, SDL threat modelling tool, T-map, and CORAS[21]. Microsoft defines threat modelling as a design method that can assist with distinguishing threats, assaults, vulnerabilities, and countermeasures that could influence applications[40]. The process of Microsoft threat modelling is shown in Fig. 3. According to Ref.[14], conducting security analysis to proactively identify security and privacy vulnerabilities of a conversational system such as a chatbot before deployment will help to avoid significant damage. Thus, a threat modelling method like STRIDE modelling is critical for insurance chatbots.

## STRIDE modelling

The chatbot does not present new security concerns; all the security concerns in the chatbot were previously discovered in other AI systems that were built[12,13]. According to Ref.[41], chatbots' security concerns are categorised into threats and vulnerabilities. The authors in Ref.[41] used STRIDE Modelling to give different threats and attack intentions. STRIDE is a widely used threat modelling methodology. STRIDE consists of six categories of threats, as shown in Table 1. In the STRIDE modelling, a data flow diagram (DFD) of the application (system) is implemented, and the STRIDE model is applied at each node of the DFD to identify threats and vulnerabilities that are associated with the six categories of STRIDE[21].

## Related work

So far, the literature has reported some studies on the security of chatbots used in the financial industry. However, very few studies have focused on the insurance industry. Some of the previous research efforts on chatbot security are presented as follows. According to Ref.[42], understanding the underlying issues requires identifying the critical steps in the methods used to design chatbots related to security. The authors discussed all the significant security, privacy, data protection, and social aspects of using chatbots by reviewing the existing literature and

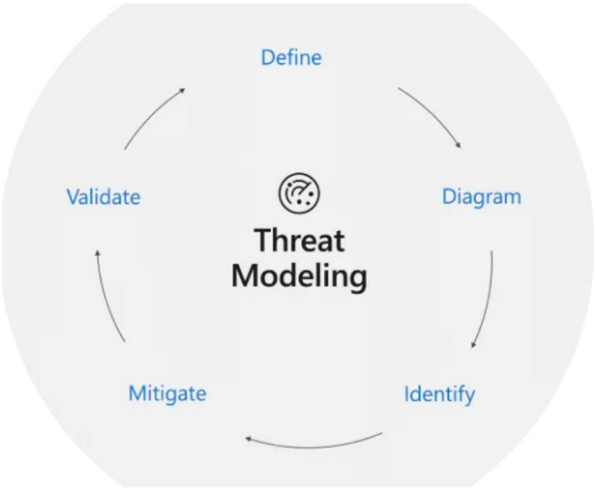


**Figure 3.** Microsoft threat modelling tool[40].

| No# | Category | Property | Attack intention |
|---|---|---|---|
| 1 | Spoofing | Authentication | Illegal access and use of another user's credentials. Impersonating something or someone else |
| 2 | Tampering | Integrity | Aimed to change/modify data maliciously |
| 3 | Repudiation | Non-repudiation | Aimed to perform an illegal operation in a system |
| 4 | Information disclose | Confidentiality | Data theft |
| 5 | Denial of service | Availability | Aimed to deny access to valid users |
| 6 | Elevation of privilege | Authorisation | Aimed to gain privileged access |

**Table 1.** Types of security threats.

producing a complete view of the given problem. The study identified security challenges and suggested ways to reduce the security challenges that are found with chatbots. The authors in Ref.[17] stated that chatbots' security and privacy vulnerabilities in the financial sector must be considered and analysed before the developers do the deployment. Through an analysis of the literature, the researchers identified the security issues but did not provide a framework or methodology for identifying the security threats in chatbots. In Ref.[6], authors used a vignette-style methodology as an induction protocol to investigate the effects of chatbot vignettes with socio-emotional features and without socio-emotional features on intent to utilise the chatbot for financial support purposes. The findings reveal that the social-emotional characteristics of chatbots in the financial industry can indicate a discrepancy between privacy and trust. The authors concluded that suitable precautionary analysis concerning chatbots' security and privacy vulnerabilities in the financial industry must be executed before deployment.

In Ref.[16], the integration of chatbots and blockchain technology was proposed to improve chatbot security issues in the financial sector. Using a blockchain-enabled chatbot, the authors implemented a proof of concept and evaluated the performance based on several security and privacy concerns. After developing threat models for chatbots in the financial sector, the study formulated a list of requirements. The authors in Ref.[14] examined existing chatbots' security and privacy vulnerabilities and proposed that chatbot developers perform a security analysis before deploying to avoid substantial harm. The study analysed potential security and privacy exposures in the chatbot architecture and discovered that the security community has not yet implemented comprehensive requirements for chatbot security. The researchers started by understanding how the existing chatbot architecture works by following the path that a message takes from the client module to the communication module, the response generation module, and the database module. After that, they identified possible attacks in each module. However, they did not provide a framework for identifying chatbot security attacks and mitigations.

In Ref.[8], it was observed that chatbots with AI features might encroach on client security and individual protection. Thus, security has become a significant issue that chatbot developers must address. The study developed the Chatbot Security Control Procedure (CSCP) for banks to monitor chatbots' security and ensure clients' protection. Their research findings show no security in the chatbot, and the AI security software causes the security loophole in chatbots. In Ref.[9], it was stated that security and privacy in chatbots require serious attention. The study investigated the initial set of issues assumed to be factors affecting clients' trust in chatbots for client service. The findings from the study show that the main issue of the clients not trusting chatbots is their poor security and privacy. An exploratory research design, which involves conducting semi-structured interviews, was used to answer the research questions. In Ref.[43], a PreBot was developed that allows privacy within a conversation or chat between the user and the chatbot. The conventional privacy bot was developed because of the concern that the current chatbots are failing to protect users' privacy. The PreBot has an interface that provides the user with privacy settings and service provider privacy policies.

In summary, we observe that none of the previous studies have focused on threat modelling for insurance chatbots. Also, no study has used STRIDE modelling to identify security threats that pertain to insurance chatbots. STRIDE, as the oldest and most mature threat modelling method[25], has the capabilities to afford reliable, proactive security assessment of insurance chatbots. Thus, as a contribution, this paper presents a first attempt at threat modelling for insurance chatbots using STRIDE. It also presents an empirical study from the South African industry, which is a geographical context not yet covered in the literature to date. Thus, this study extends the existing discussion on threat modelling through a case study from a new context. It also demonstrates the application of STRIDE modelling for threat elicitation for data security of insurance chatbots, which has not received sufficient attention in the literature.

## Methods

We used a combination of semi-structured interviews and document reviews to collect data for the study. The semi-structured interview was used for one-on-one interviews with selected participants within the organisation. The participants that were purposively selected to participate in the study are the following: security experts, chatbot developers, chatbot testers, chatbot users, and chatbot managers in an insurance company based in South Africa (see Table 2). The criteria for selection is that the participants must have spent at least one full year as staff in the insurance organisation. Document review was used to collect information from the documentation on the organisation's insurance chatbots. The document review was done before and after the semi-structured interview sessions. The pre-interview document review activity enabled us to gain preliminary insights into the expected functionalities and features of the organisation's insurance chatbots, which helped us formulate appropriate interview questions. The document review after the semi-structured interview enabled us to verify data collected from participants during the semi-structured interview. Based on the data collection, we identified the use cases of chatbots in the organisation. We then applied STRIDE modelling to elicit the data security threats

| S/N | Role | Number | Reference |
|---|---|---|---|
| 1 | Security experts | 2 | A, D |
| 2 | Chatbot developers | 2 | B, E |
| 3 | Chatbot testers | 2 | F, G |
| 4 | Chatbot users | 2 | C, J |
| 5 | Chatbot managers | 2 | H, I |

**Table 2.** Profile of participants in the study.

and vulnerabilities associated with the usage scenarios of the insurance chatbots. An overview of the study's research design is shown in Fig. 4.

## Data collection

The purpose of the data collection, which involves semi-structured interviews of participants and document review, was to (1) identify the potential use cases of chatbots for CRM in a South African insurance organisation and (2) understand the chatbot operations and the issues around data security in chatbots in the organisation. The insurance organisation has two types of chatbots: iAssist chatbots, an AI chatbot, and a WhatsApp chatbot. The summary of participants' responses to questions about the two types of chatbots within the organisation are shown in Tables 3 and 4.

## Chatbots use cases

Based on the findings from the data collected during the participants' interviews, we identified the use cases of chatbots within the organisation. The identified chatbot use cases are shown in Figs. 5 and 6.

Figure 5 shows the contact centre agent, employee, and administrator as the actors for the iAssist use case. The functions within the iAssist use case are personal lines, commercial lines, claims, and human resources.

Figure 6 shows the contact centre agent, client, and administrator as the actors for the WhatsApp use case. The functions within the use case are policy schedule, confirmation of cover, border letter, and the home agent.

Four sample use case narratives that describe the events and interactions of some of the use cases are presented in Tables 5, 6, 7 and 8.

## Ethics approval and consent to participate

All methods used in this study followed relevant guidelines and regulations. Ethical approval and waiver of consent were obtained from the Research Ethics Committee of the Faculty of Informatics and Design of the Cape Peninsula University of Technology (approval number: 215296273/2021/12).

## Human ethics and consent to participate

Informed consent was obtained from all participants that were part of the study.

## STRIDE modelling

This section presents the activities undertaken during the STRIDE modelling to identify how insurance chatbot use cases are vulnerable to each category of STRIDE. We started by performing the data flow analysis of all the use cases and then determining if they are susceptible to any of the components of STRIDE: Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Elevation of privilege.

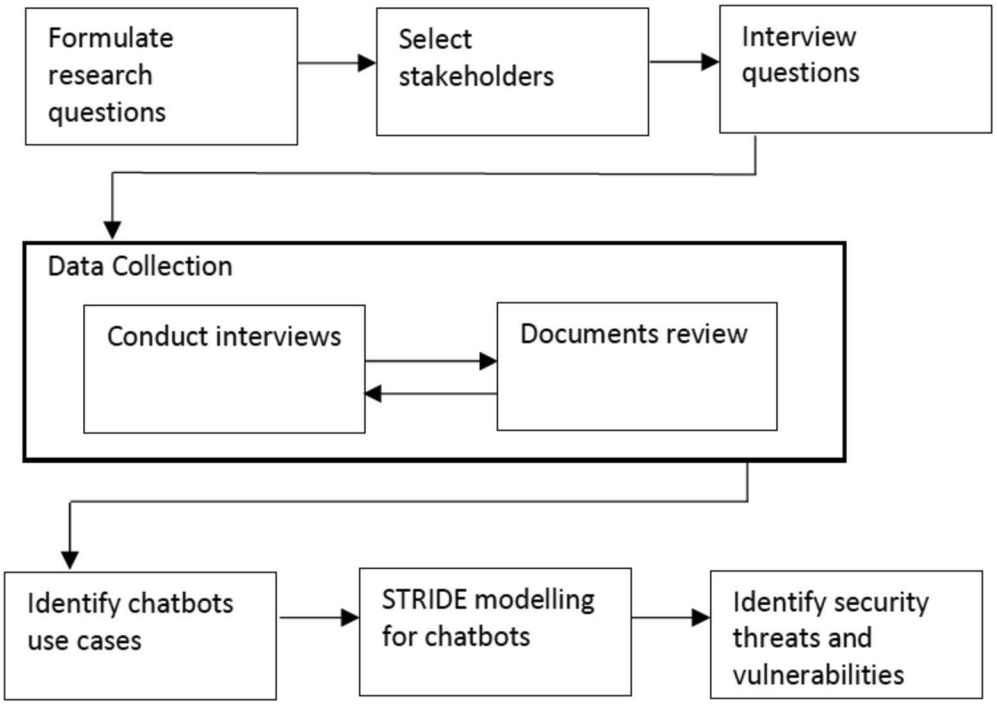


**Figure 4.** Research design.

| S/no | Questions | Finding from participant's responses |
|---|---|---|
| 1 | What purposes does the chatbot fulfil in your organisation? | iAssist bot<br>Manage a relationship with clients effectively<br>Provide a faster and easier way of finding information<br>Improve productivity<br>Reduce training to contact centre agent<br>Less expense<br>Consistency of response<br>Easy to use<br>Allows information search<br>Participants: A, C, E, F, G, H, I & J<br>WhatsApp bot<br>Manage a relationship with clients effectively<br>Cut out the middleman manual intervention of customers having to speak directly with consultants<br>Provide self-service<br>Easy and fast access to information<br>Consistency of response<br>Easy to use<br>Provide 24/7 access<br>Participants: A, B, D, F, G, & J |
| 2 | In what ways are chatbots used in your organisation? | iAssist bot<br>Gives answers to the employees on repetitive questions<br>Guide contact centre agents on how to ensure certain assets<br>Easy and quick access to information<br>Improve productivity<br>Give support to contact centre agents in helping clients<br>Consistency of response<br>Easy to use<br>Improves contact centre agent performance<br>Participants: A, B, C, E, F, G, H, I & J<br>WhatsApp bot<br>Manage a relationship with clients effectively<br>Cut the middleman intervention<br>Easy and fast access to information<br>Consistency of response<br>Easy to use<br>Provide 24/7 access<br>Participants: A, B, D, F, G, & J |
| 3 | List the specific processes/operations where chatbots are used in your organisation | iAssist bot<br>It gives the following functionalities: Personal Lines, Commercial Lines, Claims, and Human Resource<br>HR-related questions asked by staff like annual leave inquiries, Pension Fund, etc<br>HR bot also contains policy documents like the organisation's security policy, social media security policy, etc<br>Personal Lines and Commercial Lines both provide policies offered, underwriting guidelines and rules, details of allowed risk, and items that are allowed to be insured<br>Participants: A, D, F, G, & J<br>WhatsApp bot<br>Provide the following functionalities: Send my policy schedule, Send my confirmation of cover, border letter<br>Participants: A, B,D, F, G, & J |
| 4 | What kind of support and maintenance is available for chatbots in your organisation? | iAssist bot<br>There is a team that always checks how many questions were asked and how many were answered or not answered by the chatbot<br>The service provider provides first-line support, changes existing frequently asked questions and answers, and adds, deletes, or updates documents<br>The service provider provides business and technical support as well as maintenance for the solution, and there are various types of maintenance: preventative maintenance, adaptive maintenance, etc<br>Participants: A, C, D, E, F, G, H, I& J<br>WhatsApp bot<br>Keep and route all the unanswered questions to the available agent<br>If the agent does not know the query, he takes it up with the technical team<br>Participants: B & E |
| 5 | What kind of data is stored in chatbots in your organisation? | iAssist bot<br>HR-related policy documents<br>HR-related information<br>Underwriting rules and information<br>Details of allowed risk<br>Items that are allowed to be insured<br>Client and employee information<br>Participants: A, C, D, E, F, G, H, I & J<br>WhatsApp bot:<br>Policy documents<br>Policy-related information<br>Client information<br>Participants: A, B, D, F, G, & J |
| 6 | Who are the chatbot users? | iAssist bot<br>Employees<br>Call centre agent<br>Participants: A, B, C, D, E, F, G, H, I & J)<br>WhatsApp bot<br>Clients<br>Call centre agent<br>Participants: A, B, C, D, E, F, G, H, I & J |

**Table 3.** Identifying potential use cases of chatbots in the insurance organisation.

## Dataflow analysis of insurance chatbot use cases

As per the distinction of business process operations, the chatbot is divided into various information streams: refining, transmission, and information storage. In threat modelling analysis, the data flow diagram (DFD) is normally used to imitate the data flow interaction association between chatbot external and internal interactors. The signs of the information flow diagram are shown in Table 9. In the sequel section, the data flow diagrams are used to depict the chatbot business process.

## Analysis of chatbot use cases

*iAssist chatbot*

iAssist chatbot assists employees with access to frequently asked questions (FAQ) through cognitive or intelligent search. Beyond just the chat, it is a search bot and document web document viewer. It is exposed to the organisation's intranet (internal internet environment).

The user interacts with the chatbot mainly through:

- User login (1.0): To access the chatbot operations, the user needs to log in to the system and access the information based on the user's role.
- Claim (2.0): anything that has to do with claim information or documents, such as adding basic access and waiver rules regarding the claim.
- Personal Lines (3.0): This can be initiated by agent users; personal line functionality provides the corresponding underwriting information, such as underwriting guidelines for personal lines.
- Commercial Lines (4.0): This can be initiated by agent users; personal line functionality provides the corresponding underwriting information, such as underwriting guidelines for commercial lines.
- Human Resource (5.0): This can be accessed by all employees across the organisation. It provides HR-related information and documents.

The first-level data flow diagram decomposition of these business process operations is shown in Fig. 7.

Figures 8, 9, 10, 11 and 12 give the second level of the data flow diagram decomposition of the iAssist chatbot's five business operations (User Login, Claims, Personal Lines, Commercial Lines, and Human Resources).

Figure 8 depicts the login process. Before the user is given access to a chatbot, the user must log in first with a username and password. Authentication is done through the user account database. Once the user is authorised, the user accesses the chatbot based on the user's role. All the interactions are stored in the log file for auditing purposes.

Figure 9 depicts when the user has already been given rights to access the Claims chatbot. Then, the user requests information and asks FAQ (frequently asked questions) related to the claim. All the interactions, including query processing results, are stored in the log file for auditing purposes.

Figure 10 shows when the user has been given rights to access the Personal Lines chatbot. The user requests information and asks FAQ related to the Personal Lines queries. All the interactions with the chatbot, including query processing results, are stored in the log file for auditing purposes.

Figure 11 shows when the user has been given rights to access the Commercial Lines chatbot. The user requests information and asks FAQ related to the Commercial Lines queries. All the interactions with the chatbot, including query processing results, are stored in the log file for auditing purposes.

Figure 12 shows when the user has been given rights to access the Human Resource chatbot. The user requests information and FAQ related to Human Resource queries. All interactions with the chatbot, including query processing results, are stored in the log file for auditing purposes.

*WhatsApp chatbot*

The WhatsApp insurance chatbot manages a relationship with clients effectively. It cut out the middleman manual intervention of customers having to speak directly with consultants by providing a self-service. It offers easy and fast access to information, consistency of response, ease of use, and 24/7 access. The user interactions with the chatbot mainly include User verification (1.0). To access the chatbot operations, a user must provide an ID or passport number, the OTP is sent to the user's registered mobile number for verification, and all the requests are sent to the client's email.

The first-level data flow diagram decomposition of these above business process operations is shown in Fig. 13.

Figure 14 gives the second level of the WhatsApp data flow diagram decomposition of the above business operations.

Figure 14 shows when the user has been given the right to access the WhatsApp chatbot. The user requests information and FAQ related to the policy queries. All the interactions with the chatbot, including query processing results, are stored in the log file for auditing purposes.

## Identification of security threats and vulnerabilities based on STRIDE

The STRIDE model derives from the following six threat categories: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege. The components of STRIDE are explained further as follows:

1. Spoofing—This threat happens when the attacker illegally accesses and uses another user's credentials. Impersonating something or someone else.
2. Tampering—This threat happens when attackers aim to change/modify data maliciously.

| S/no | Questions | Finding from participant's responses |
|---|---|---|
| 1 | Where are the chatbots used in your organisation hosted? | iAssist bot<br>Third-party cloud storage<br>Participants: A, F, G, & J<br>WhatsApp bot<br>Client's Third-party cloud storage<br>Structured database<br>Participants (A, B & E) |
| 2 | How do the chatbots in your organisation integrate with social media platforms? | iAssist bot<br>It's a web bot, not integrated into the social media platform<br>Participant: (A, D, F, G, & J)<br>WhatsApp bot<br>It's integrated via WhatsApp<br>Participants: A, B & E |
| 3 | How does the chatbot platform in your organisation store data after transactions? | iAssist bot<br>Are stored in a third-party Mongo database<br>Participants: A, D, F, G, & J<br>WhatsApp bot<br>Are stored in a third-party Mongo database<br>Participants: A, B & E |
| 4 | What happens to used data inside chatbots? | iAssist bot<br>Keep for reviews, optimisation purposes, and ongoing maintenance of the solution<br>Participants: A, D, F, G, & J<br>WhatsApp bot<br>Keep for reviews, optimisation purposes, and ongoing maintenance of the solution<br>Participants: A, B & E |
| 5 | What features does your chatbot have? e.g. speech recognition, text-based, or speech-to-text | iAssist bot<br>Text-based<br>Participants: A, B, C, D, E, F, G, H, I & J<br>WhatsApp bot<br>Text-based<br>Participants: A, B, C, D, E, F, G, H, I & J |
| 6 | What kind of security measures does your chatbot have to prevent identity theft? | iAssist bot<br>Chatbot data is subject to privacy legislation, so parts that are hosted here in South Africa are subject to the Protection of Personal Information Act, and any other part that is hosted in Europe is subject to GDPR from a data privacy point of view<br>End-to-end encryption<br>Participants: A, B, C, D, E, F, G, H, I & J<br>WhatsApp bot<br>Chatbot data is subject to privacy legislation, so parts that are hosted here in South Africa are subject to POPOA; any other part that is hosted in Europe is subject to GDPR from a data privacy point of view<br>End-to-end encryption<br>Participant: A, B, C, D, E, F, G, H, I & J |
| 7 | What kind of security does the chatbot have to ensure data privacy? | iAssist bot<br>Chatbot data is subject to privacy legislation, so parts that are hosted here in South Africa are subject to POPOA; any other part that is hosted in Europe is subject to GDPR from a data privacy point of view<br>End-to-end encryption<br>Participant: (A, B, C, D, E, F, G, H, I & J)<br>WhatsApp bot<br>Chatbot data is subject to privacy legislation, so parts that are hosted here in South Africa are subject to POPOA; any other part that is hosted in Europe is subject to GDPR from a data privacy point of view<br>End-to-end encryption<br>Participant: (A, B, C, D, E, F, G, H, I & J) |
| 8 | What kind of security measures does your chatbot have to ensure data integrity? | iAssist bot<br>Chatbot data is subject to privacy legislation, so parts that are hosted here in South Africa are subject to POPOA; any other part that is hosted in Europe is subject to GDPR from a data privacy point of view<br>End-to-end encryption<br>The security measure in the chatbot to ensure data integrity comes under the same controls that are implemented for POPIA and With the GDPR in terms of making sure the person is who they say they are<br>Encryption of information at rest<br>Participant: (A, B, C, D, E, F, G, H, I & J)<br>WhatsApp bot<br>Chatbot data is subject to privacy legislation, so parts that are hosted here in South Africa are subject to POPOA; any other part that is hosted in Europe is subject to GDPR from a data privacy point of view<br>End-to-end encryption<br>Participant: (A, B, C, D, E, F, G, H, I & J) |
| 9 | What kind of security measures does your chatbot have to prevent unauthorised access? | iAssist bot<br>Users can only access it after they have been authenticated<br>A user must sign in with a staff credential<br>End-to-end encryption<br>Participant: (A, B, C, D, E, F, G, H, I & J)<br>WhatsApp bot<br>Users can only access it after they have been authenticated<br>It validates that at least the user's details are those on the organisation's policy<br>End-to-end encryption<br>Participant: (A, B, C, D, E, F, G, H, I & J) |

Continued

| S/no | Questions | Finding from participant's responses |
|---|---|---|
| 10 | What kind of security measure does your chatbot have for user authentication? | iAssist bot<br>The staff signs in with the username and password, which is the security measure for user authentication<br>The authentication is provided by roles stored and implemented in the application. It has 3 roles menu with end-user content, administrator, or manager. Those roles assist in authorisation in terms of which department the user gets access to and which reports are accessed<br>End-to-end encryption<br>Participant: (A, B, C, D, E, F, G, H, I & J)<br>WhatsApp bot<br>It validates, based on something the person knows, and uses that to prove that the user is authentic<br>The user adds an ID number as a user identification method, and the chatbot sends the OTP to the user. After the user has confirmed, it sends the information requested to a user's email<br>If the user enters the OTP number incorrectly, it will just take the user back to the main menu. If the user does not have active policies linked to the ID number, then the bot will tell the user that no active policies are linked to this ID number<br>Also, when the chatbot sends back personal details like phone number, ID number, etc., it is masked, e.g. 083*******92, mil****ab@gmail.com<br>End-to-end encryption<br>Participant (A, B, C, D, E, F, G, H, I & J) |
| 11 | What are the security vulnerabilities that you have found with your chatbot? | No variabilities were found<br>Participants (A, B, C, D, E, F, G, H, I & J) |

**Table 4.** Understanding the operations and issues of data security in chatbots in the insurance organisation.

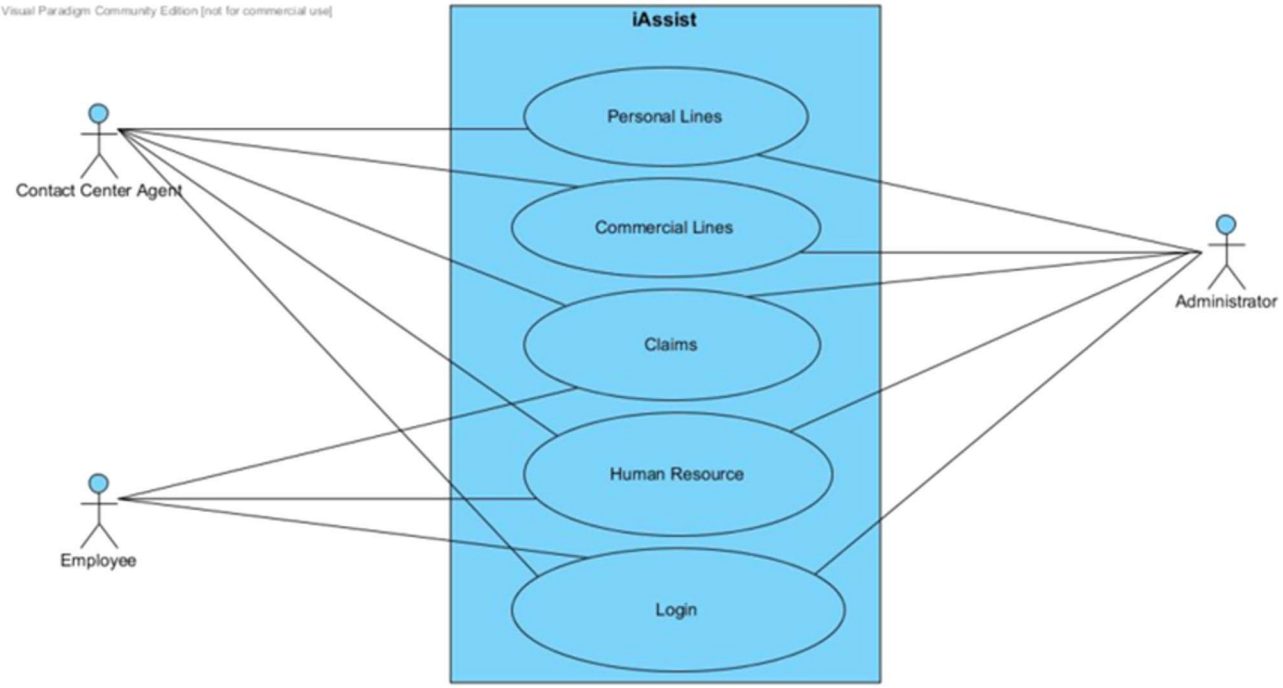


**Figure 5.** iAssist bot use case.

3. Repudiation—This threat happens when an attacker aims to perform an illegal operation in a system
4. Information Disclosure—hackers steal confidential information
5. Denial of Service—This threat happens when an attacker aims to deny access to valid users.
6. Elevation of Privilege—This threat happens when an attacker aims to gain privileged access.

Considering the six types of threats covered by STRIDE, we analysed every element in the Chatbot DFD to ascertain if any of the components of STRIDE pertain to them to identify the security vulnerabilities and threats to the chatbot. The STRIDE analysis identified the security goals, threats, and vulnerabilities for each chatbot in the organisation (see Tables 10 and 11). In Tables 10 and 11, the symbol (√) means that an identified security threat or vulnerability pertains to a specific component of STRIDE, while the symbol (X) means irrelevant. The possible security threats and vulnerabilities (first column in Tables 10 and 11) associated with each component of STRIDE were compiled in consultation with security experts and by using published documentation on Critical Security Controls (CIS Controls V8)[44] and Information Risk Assessment Methodology 2 (IRAM2)[45] as a guide.

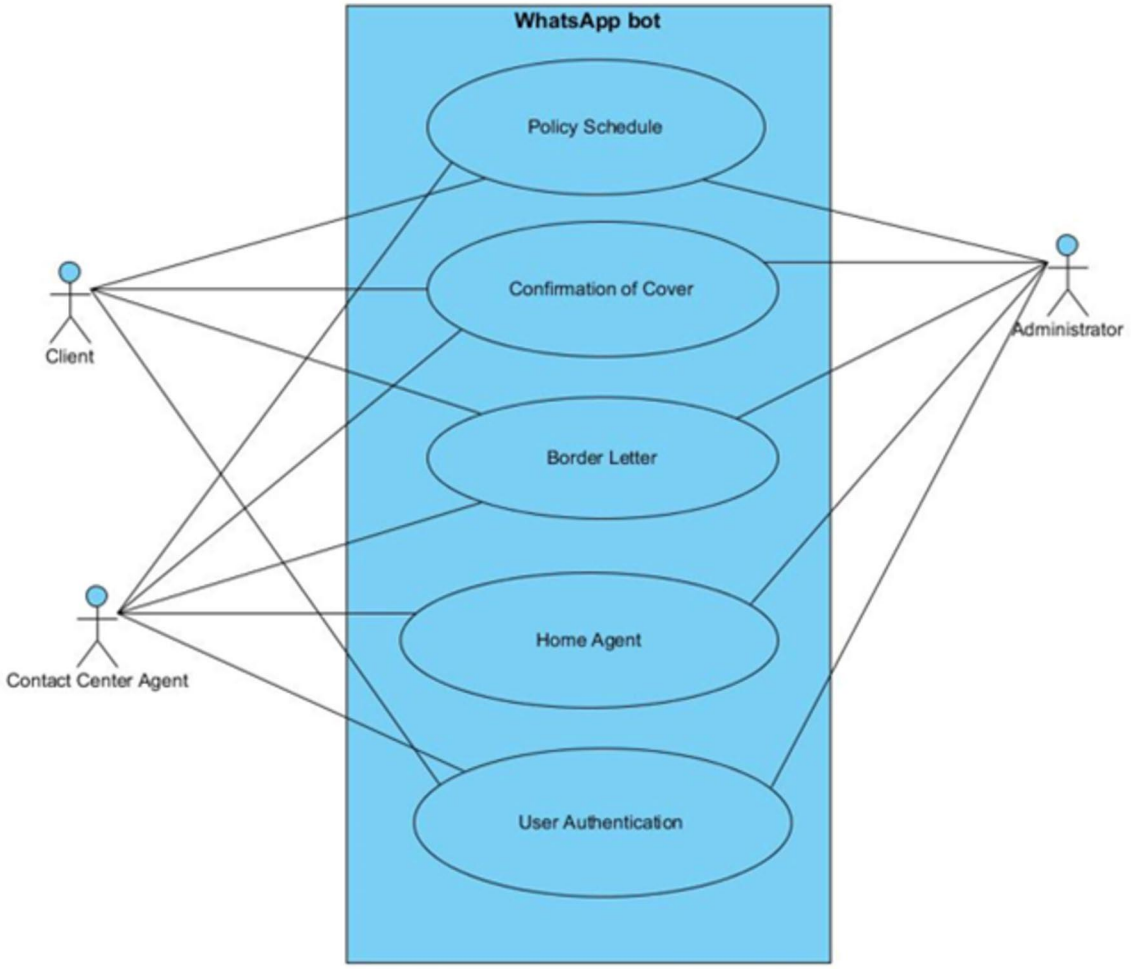


**Figure 6.** WhatsApp bot use case diagram.

| Use Case Name: login | |
|---|---|
| Primary actors | Employee, Contact Centre Agent, Admin |
| Secondary actors | |
| Description | User Login into the iAssist Chatbot |
| User Action | iAssist Response |
| 1. User Enters Username and Password | 2. Validates if text boxes are not empty |
| | 3. Check if the username and password match |
| | 4. Login the User into the chatbot |
| Alternate | |
| Step 2 Text boxes are empty | Notifies the user to fill in mandatory fields |
| Step 3: The Username and Password do not match | Alerts the user that the Username or Password is incorrect. Prompt the user to enter the username and password again |

**Table 5.** Login use case narrative.

For more accessible representation, the different insurance chatbots were represented with the following variables:

1. Claims bot (AI Bot)—A1
2. Personal Lines bot (AI Bot)—A2
3. Commercial Lines bot (AI Bot)—A3
4. Human Resources bot (AI Bot)—A4
5. WhatsApp bot (Declarative/Human Bot)—A5

| Use case name: Human Resource iAssist | |
|---|---|
| Primary actors | Employee |
| Secondary actors | |
| Description | Employee search information |
| User Action | iAssist Response |
| 1. Employee selects HR bot | 2. Chatbot shows an FAQ option for an employee to choose and a text box for a user to search for information |
| 3. Choose FAQ | 4. Chatbot sends back information |
| 5. Search for the annual leave policy | 6. Return the list of annual leave policy documents |
| 7. Search for payslip | 8. Returns a list of payslip information document |
| 9. Search for car and travel allowance policy | 10. Return car and travel allowance policy document |
| 11. Search the Organisation's security policy | 12. Returns a list of the organisation's security policy document |
| 13. Search for information | 14. Chatbot does not contain an answer to the question |
| Alternate | |
| Step 13: No answer that matches the question found | The chatbot does not have a definitive answer to the question. The user can check the search results for what has been able to find<br>Users must also try rephrasing the question using a full sentence, i.e. use more than just a single keyword or two<br>The question has been saved and will be reviewed so that the chatbot can assist better in the future<br>Users can also contact HR directly |

**Table 6.** Human resource iAssist use case narrative.

| Use case name: claims iAssist | |
|---|---|
| Primary actors | Employee, Contact Centre Agent |
| Secondary actors | |
| Description | Employee or Contact Centre Agent search for information |
| User Action | iAssist Response |
| 1. User selects claims bot | 2. Chatbot shows an FAQ option for an employee to choose and a text box for a user to search for information |
| 3. Choose FAQ | 4. Chatbot sends back information |
| 5. User search for claim policy | 6. The bot does not contain a specific answer for this question; it then returns a list of documents that could be related to the search and some options for the related questions the user might want to know |
| 7. The User clicks the displayed information about the claim, e.g., when a claim can be rejected | 8. The bot returns a list of documents about claim rejection and displays some information on the chat |
| 9. Search the purpose of the claim management framework | 10. The bot returns a list of documents about the claim management framework and displays some information on the chat |
| 11. Employee search for who makes business decisions on the claims | 12. The bot returns a list of documents about who made a decision on the claim and displays some information on the chat |
| Alternate | |
| Step 6: No answer that matches the question found | The chatbot does not have a definitive answer to the question. The user can check the search results for what has been able to find |

**Table 7.** Claims iAssist use case narrative.

## Discussion of findings

The study's findings regarding the security vulnerabilities and security threats that pertain to insurance chatbots are outlined in the following sections.

### Identified security vulnerabilities

Table 10 shows that insurance chatbots are vulnerable to all components of STRIDE – Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege in one form or another. The areas of vulnerabilities are:

1. **Spoofing**: The Claims bot (A1), Personal Lines bot (A2), Commercial Lines bot (A3), and Human Resource bot (A4) are vulnerable to *Errors in business logic and business processes*. The WhatsApp bot (A5) is vulnerable to *Social engineering attacks*. All the chatbots (A1—A5) are vulnerable to *Interception of communication*, *Unauthorised access*, *Attack through bad design or mis-configuration*, *Phishing attacks*, and *Manipulation of employees by attackers*.

| Use Case Name: Policy WhatsApp bot | |
|---|---|
| Primary actors | Client |
| Secondary actors | |
| Description | The client starts a chat |
| User Action | WhatsApp Response |
| 1. Client greet | 2. The chatbot shows a welcoming message and the list of services offered by the chatbot (Send my Policy Schedule, Send my Confirmation of Cover, Border Letter, and Home Agent) |
| 3. Choose and send my confirmation of cover | 4. Chatbot prompts the client to enter a preferred method of identification (SA ID number, Namibia ID number, or Passport Number) |
| 5. The client enters the identification | 6. Chatbot sends the OTP number to the client's phone number that is linked to the policy |
| 7. Client enters the OTP number | 8. The chatbot sends the document to the client's email that is linked to the policy and responds to the chat |
| 9. Type the question into the text box | 10. The chatbot does not have the answer to the question |
| 11. Client chooses option 4: Home Agent | Chatbot routes the question to the available agent |
| 11. Client enters 0 to exit | 12. The chatbot responds with the thank you message |
| Alternate | |
| Step 10: Chatbot does not have the answer to the question | Chatbot asks the user to choose the home agent option to route the question to the available agent |

**Table 8.** Policy WhatsApp use case narrative.

| Signs | Sign names | Description |
|---|---|---|
|  | User interaction | User input to the chatbot |
|  | Process | Information manipulation |
|  | Information storage | Permanent and temporal information storage |
|  | Information flow | Shows information flow from information stores, processes, or interactors |
|  | Boundaries | The system, physical, address space, or trust boundary |

**Table 9.** Information flow signs.

2. **Tampering**: All the chatbots (A1—A5) are vulnerable to *Gain access to information assets in transit* (network sniffing) and *Information compromise*.
3. **Repudiation**: All the chatbots (A1—A5) are vulnerable to *Impersonating real users*.
4. **Information Disclosure**: All the chatbots (A1—A5) are vulnerable to *Insecure disposal of assets*.
5. **Denial of Service**: All the chatbots (A1—A5) are vulnerable to *the following: Impair the availability or performance of the organisation's applications*, *Introduce malware into the application*, *Badly designed network architecture*, *Physical damage or tampering with the company's data application*, and *Stealing of physical infrastructure*.
6. **Elevation of Privilege**: All chatbots (A1—A5) are vulnerable to *Loopholes in the authorisation mechanisms*, *Unauthorised scanning or probing of a company's data applications*, *and Unauthorised analysis of publicly available information about an enterprise*.

## Identified security threats

From Table 11, we deduced security threats associated with the components of STRIDE that pertain to each chatbot within the insurance organisations. These are outlined as follows:

1. **Spoofing**: The four possible security threats to all chatbots (A1 – A5) are: An attacker can *exploit insecure default configuration*, *access the system from an unauthorised network*, *execute administrator functions*, and *gain unauthorised access*.
2. **Tampering:** We found ten possible security threats that pertain to all the chatbots (A1 – A5). These are: an attacker can *exploit insecure default configuration*, *penetrate an enterprise infrastructure*, *make organisations*

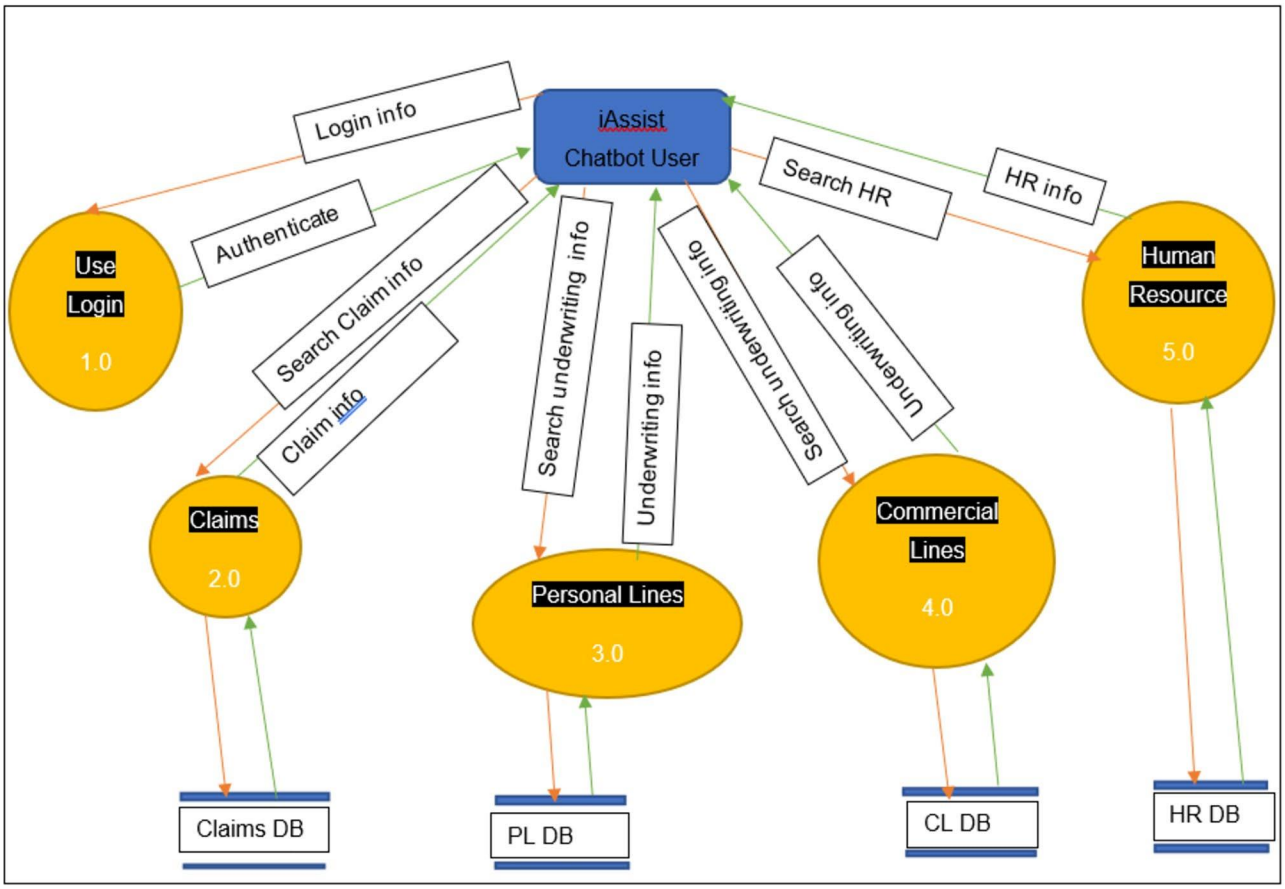


**Figure 7.** First-level data flow diagram of business process operations.

*lose control over protecting sensitive data, and bypass security controls.* Other security threats are *Insecure network infrastructure*, *Hackers scanning for vulnerabilities to exploit the remote software*, *Scanning internet address space*, *Hacking new insecure configurations and unpatched assets*, *Weak security on unidentified assets*, *and Lack of adequate logging and regular log review in applications*.

3. **Repudiation:** There are five possible security threats to all the chatbots (A1 – A5): An attacker can *exploit insecure default configuration*, *make organisations lose control over protecting sensitive data, penetrate an enterprise infrastructure*; *Insecure network infrastructure*, *Lack of adequate logging and regular log review in applications*.
4. **Information Disclosure:** The four possible security threats to all the chatbots (A1—A5) are: an attacker can *exploit insecure default configuration*, *make organisations lose control over protecting sensitive data*; *Hackers scan for vulnerabilities to exploit the remote software*, and *Ransomware attacker from the third party*.
5. **Denial of Service:** We found three possible security threats for chatbots (A1—A5): *Insecure network infrastructure*, *Lack of adequate logging and regular log review in applications Lack of adequate logging and regular log review in applications*, and *an attacker can make changes to configurations*.
6. **Elevation of Privilege:** The seven possible security threats to all the chatbots (A1—A5) are: An attacker can *exploit insecure default configuration*, *gain unauthorised access*, *bypass security controls*, *exploit a specific vulnerability through insecure design and infrastructure*, *exploit a specific vulnerability through poor design and incorrect testing*, *target unknown vulnerabilities*; *Proactive attack on newly published vulnerabilities*; *Hackers scan for vulnerabilities to exploit the remote software*

Table 12 provides an overview of the number of vulnerabilities and threats per STRIDE component based on our analysis.

The study's findings show that each component of STRIDE can affect the insurance chatbot. Several security threats and vulnerabilities are associated with Spoofing, while most security threats can stem from Tampering, Elevation of privilege and Spoofing (See Table 12). Significant security vulnerabilities also exist in insurance chatbots through Spoofing, Denial of Service, and Elevation of privilege. Specifically, the highest number of security vulnerabilities in insurance chatbots relate to spoofing, Denial of Service, and Elevation of Privilege. The most security attacks are associated with Tampering, Elevation of Privilege, Spoofing, Repudiation, and Information Disclosure.

According to Ref.[16], financial chatbots have been mostly affected by security issues in terms of insecure authentication, data integrity, system availability, transparency, and privacy concerns. The findings of this study

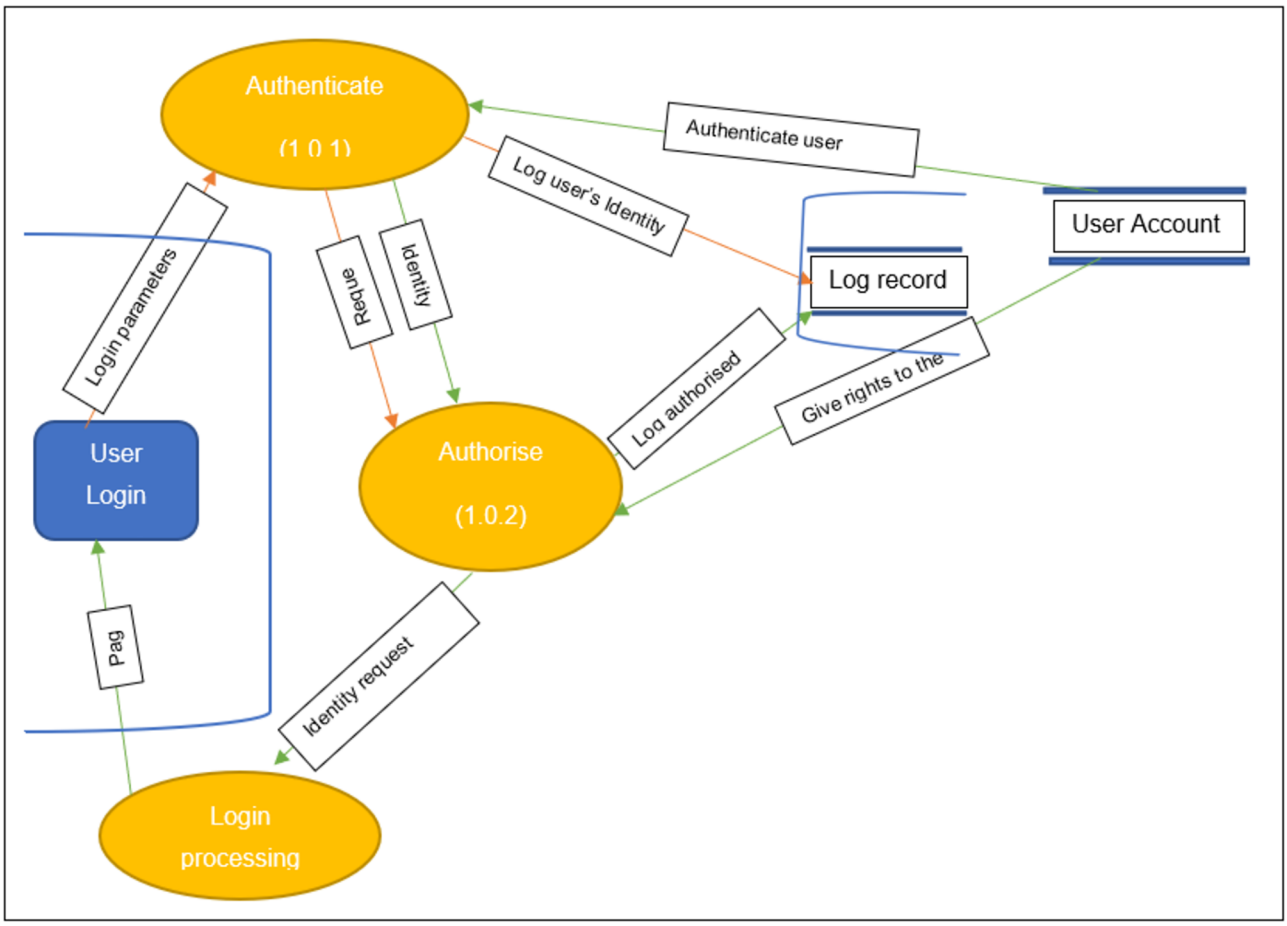


**Figure 8.** Login process.

buttress this perspective because security threats like Spoofing, Tampering, and Elevation of privilege can only occur when a chatbot has insecure user authentication, which will lead to data compromise and breach of privacy. Also, the findings of this study reveal the privacy vulnerabilities of insurance chatbots, particularly with respect to Spoofing, Denial of Service, and Elevation of Privilege, which have the highest number of security vulnerabilities for insurance chatbots. This finding confirms the observation of Ref.[17] that the management of the security and privacy vulnerabilities of the chatbots is critical. The authors recommended that a precautionary analysis of the security and privacy vulnerabilities of a financial chatbot should be conducted before it is deployed. In addition, the finding of the SLR by Ref.[15] identified malicious input, user profiling, contextual attacks, and data breaches as the main sources of security attacks for chatbots, which can be mitigated by applying blockchain technology, end-to-end encryption, and organisational controls. This study provides an empirical basis for understanding the nature of threats and vulnerabilities associated with insurance chatbots in that it reveals that specific security threats such as Spoofing, Tampering, Elevation of privilege, Repudiation that can stem from malicious input, and user profiling can lead to data breaches in chatbots.

Thus, this study makes a theoretical contribution by deepening the understanding of threat modelling and data security in insurance chatbots, which has not received sufficient attention in the literature. So far, most studies on financial chatbots are focused on banking instead of insurance. The few studies on insurance chatbots have investigated issues of adoption, design and development, and the imperatives for trust and privacy. However, they have not focussed on threat modelling as a form of precautionary analysis of financial chatbots before deployment[17]. Regarding the topic of data security in chatbots, the study also makes an empirical contribution by offering new evidence from the geographical context of the South African insurance industry, which is not common in the literature. Practically, it also provides a good guide for designers and developers (particularly insurance chatbots) on critical aspects where financial chatbots can be vulnerable and susceptible to security attacks.

## Limitations of the study

The findings of this study are credible and will add value to the existing body of knowledge since minimal exploration of issues of data security of insurance chatbots has been reported so far in the literature. However, a limitation of this study is that our findings are based on a single case study, which limits the scope of generalisation of our results. A study that is conducted on a larger scale using a multiple case study involving more insurance companies with diversity in their modes of operation, culture, and context could provide findings that would

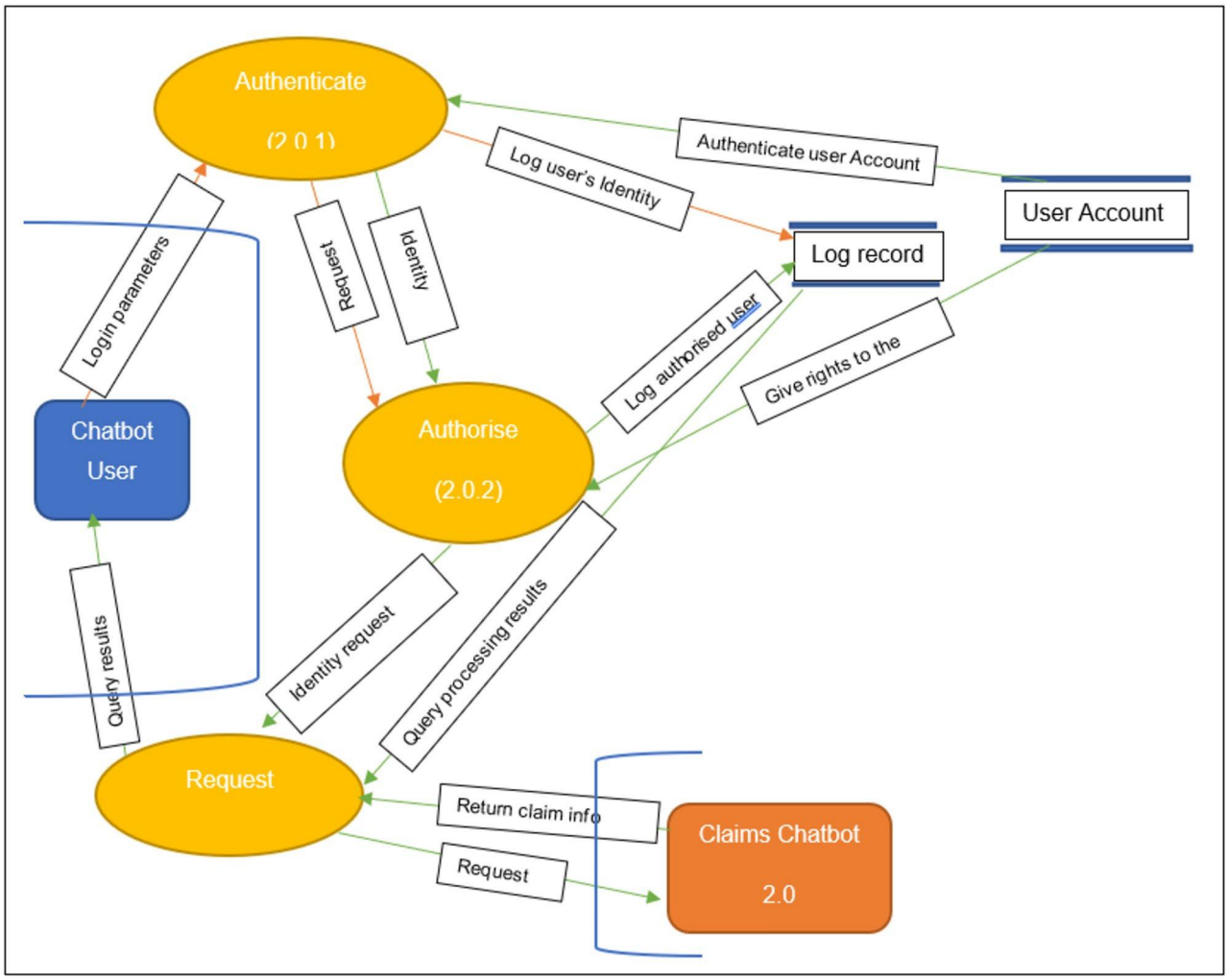


**Figure 9.** Claim request.

lead to a different conclusion compared to our study. Nevertheless, this study is a worthy exploration of the data security of insurance chatbots, which has received very little attention thus far in the literature.

## Recommendations

In this section, we present some recommendations based on the key findings of this study.

### More studies on data security in insurance chatbots

More empirical studies are needed that focus on the data security of insurance chatbots. Although studies that focus on the design and development of insurance chatbots exist[18,19], there are no studies that focus on data security in insurance chatbots. More studies that focus on data security in chatbots will enable a deeper understanding of the relationship between chatbot adoption and usage in the insurance industry and data security. It will also enable an understanding of the security challenges associated with the design and development of insurance chatbots.

### Triangulation of threat modelling methods in chatbot design and development

There are several types of threat modelling methods. Each of these methods has its strengths and weaknesses[25]. Although STRIDE is the oldest, most mature and established of all threat modelling methods, it cannot be the most suitable for use in all circumstances. Hence, there is a need for insurance chatbot developers to be knowledgeable of other threat modelling techniques and be ready to use them when appropriate. It is essential to have comparative studies that assess the suitability/effectiveness of these threat modelling methods for precautionary security analysis of insurance chatbots. Such studies will offer a credible guide for chatbot development in the insurance industry.

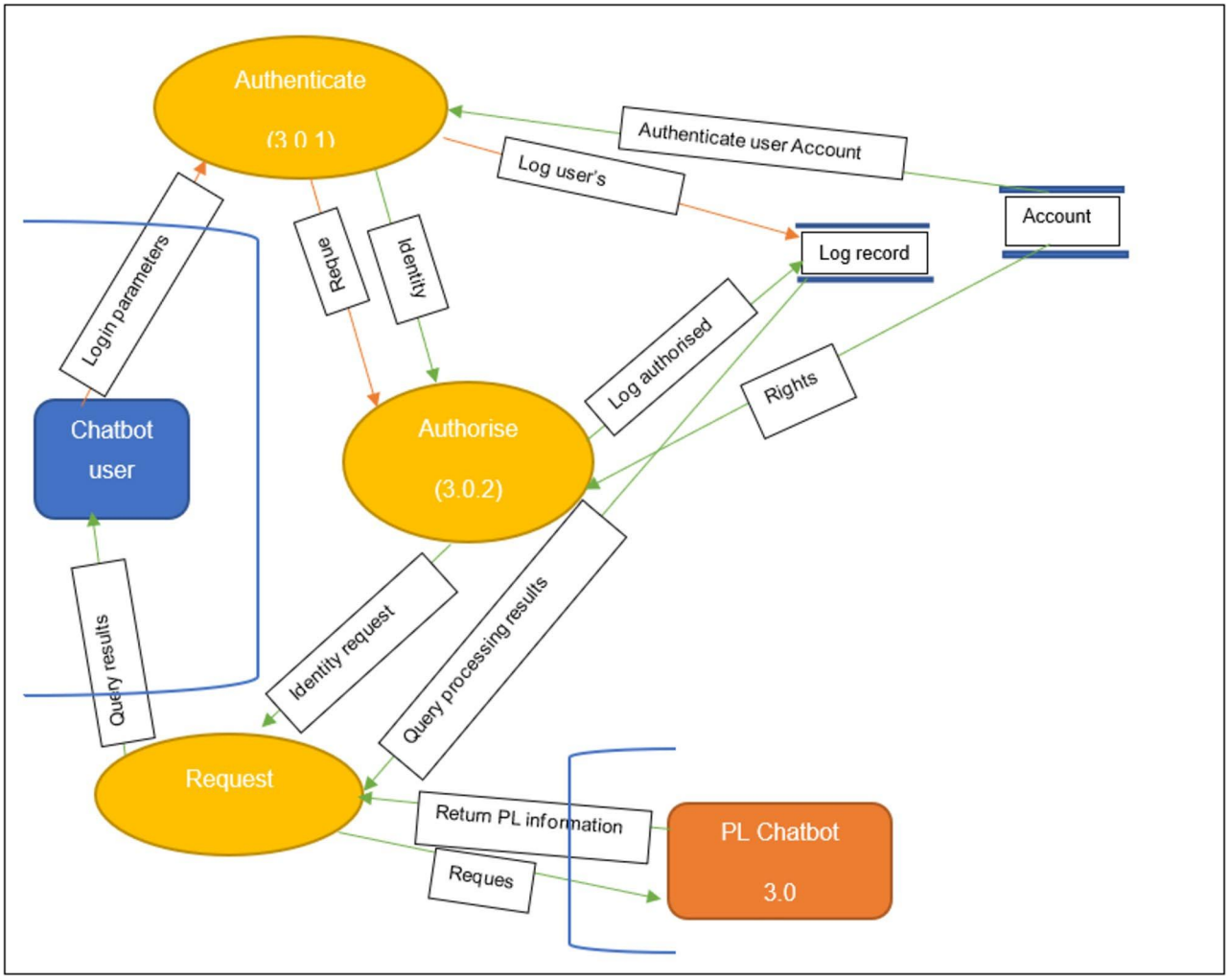


**Figure 10.** Personal lines request.

## Secure sofware development practices for insurance chatbots

Evidence from the literature suggests that secure software development is not the defacto standard or guideline for building insurance chatbots. This observation aligns with the fact that security-by-design[46] is not yet an established practice in the software industry as a whole. Most developers do not make building security into the software right from the beginning a top priority[46,47]. A good reason for this is the lack of security experts or security requirements engineers in many organisations. Also, few requirements engineers/analysts have expertise in the areas of identifying, analysing, specifying, and managing security requirements[47,48].

Several security threats and vulnerabilities are associated with insurance chatbots. Since data leaks and security breaches will negatively affect the customer and insurance company, software developers of insurance chatbots need to embrace secure software development as a standard practice. Secure software development, which has security-by-design as its core philosophy, will ensure that activities like threat modelling, security risk

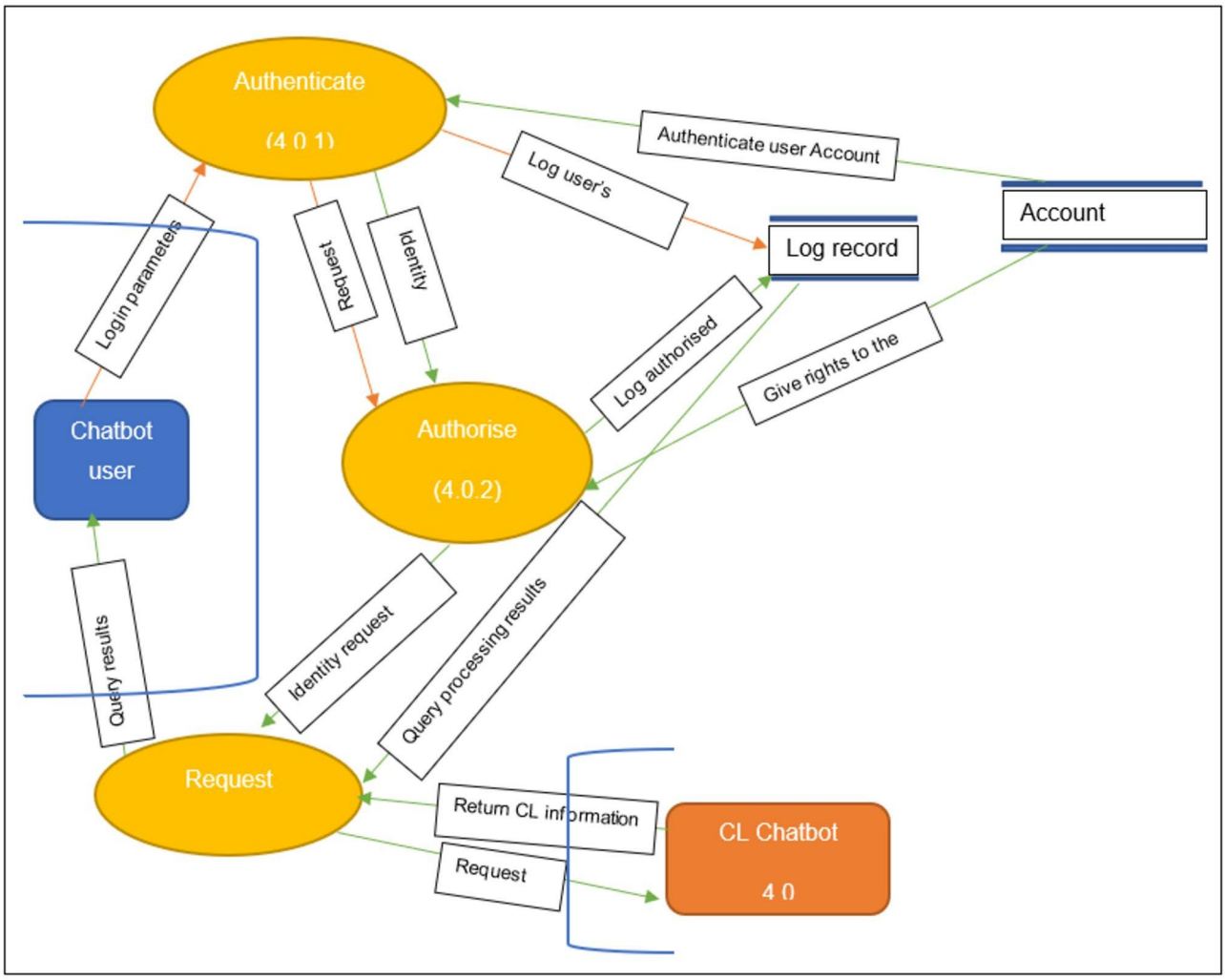


**Figure 11.** Commercial lines request.

assessment, secure coding practices, security testing, and secure configuration management take centre stage during the development of conversational AI systems like chatbots.

## Conclusion

In this study, we used STRIDE modelling to elicit the security threats and vulnerabilities that pertain to insurance chatbots by using a case study of a South African insurance organisation. We adopted a research approach that involved data collection from qualified participants selected from within the organisation, document review, use case modelling, and STRIDE modelling to identify security vulnerabilities and threats pertaining to chatbots used within the organisation. Our results show that insurance chatbots can be prone to security vulnerabilities that stem from Spoofing, Denial of Service and Elevation of privilege, while most security threats are associated with Tampering, Elevation of privilege and Spoofing.

As a contribution, this study deepens understanding of the application of STRIDE modelling. It also offers a case study on chatbot security regarding the insurance industry, which is a first attempt to the best of our knowledge. The fact that the case study is also from the South African context constitutes an empirical contribution because case studies on chatbot security from developing countries, particularly Africa, are uncommon in the literature.

Our future work will focus on developing threat models that contain the identified security threats and vulnerabilities in chatbots and mitigation strategies, culminating in formulating security requirements. Another area we will explore is to attempt solution triangulation by comparing results obtained from STRIDE modelling with other threat modelling approaches such as DREAD (Damage, Reproducibility, Exploitability, Affected Users, Discoverability), Attack Trees and CORAS(Comprehensive, Risk-driven, and Systematic Approach to System Security)[42]. In addition, we will conduct a security risk analysis of insurance chatbots using the STRIDE/DREAD combo.

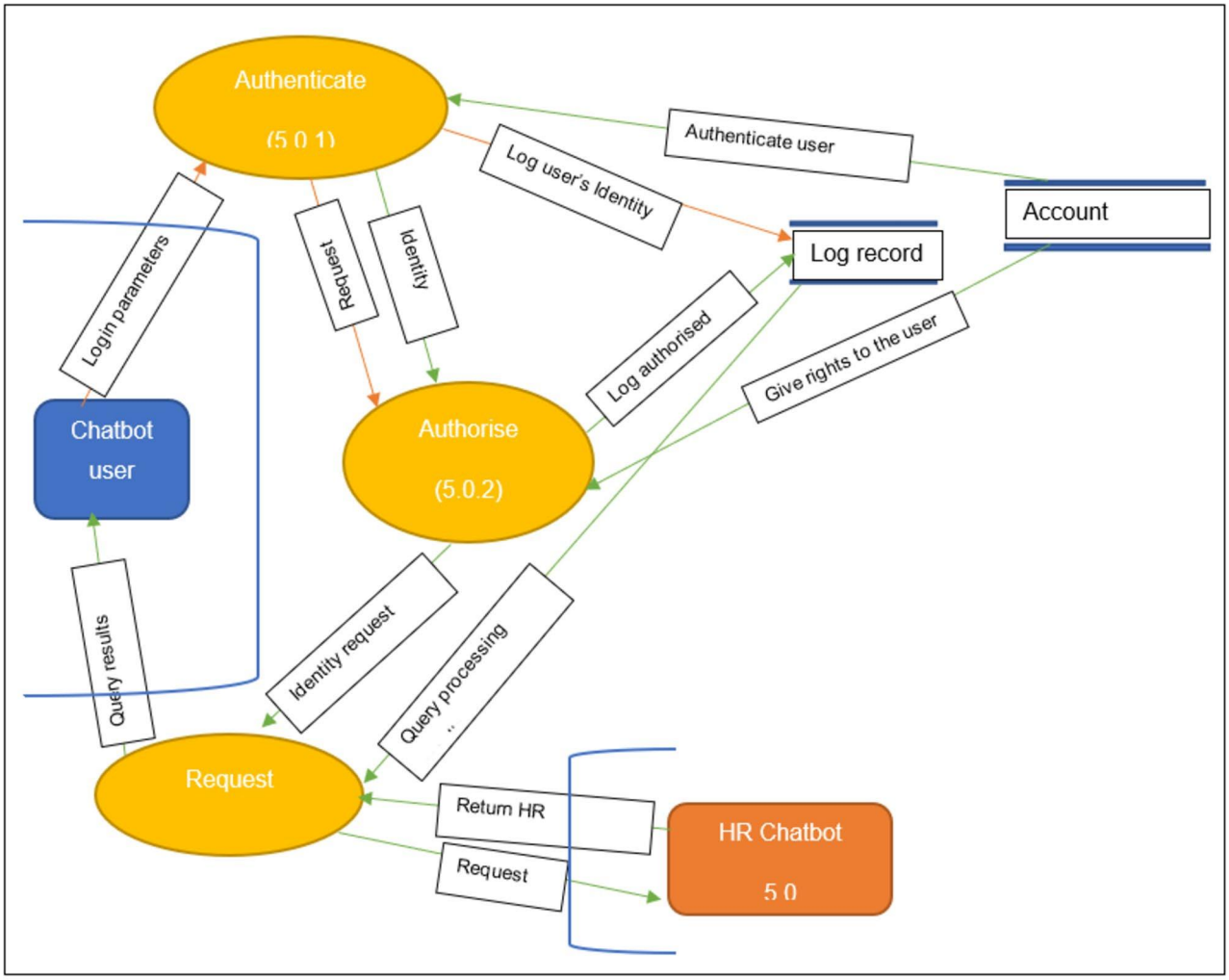


**Figure 12.** Human resource request.

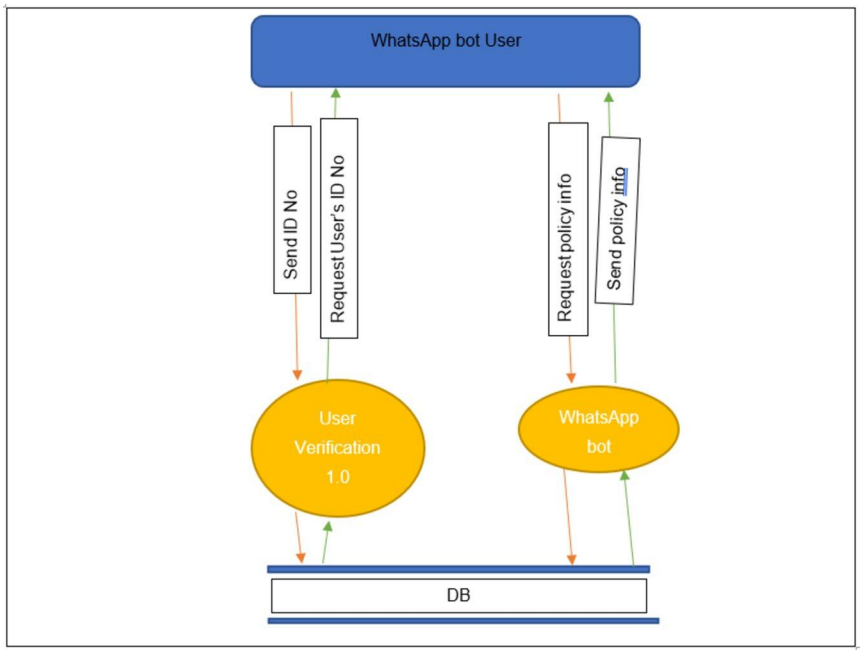


**Figure 13.** WhatsApp business data flow.

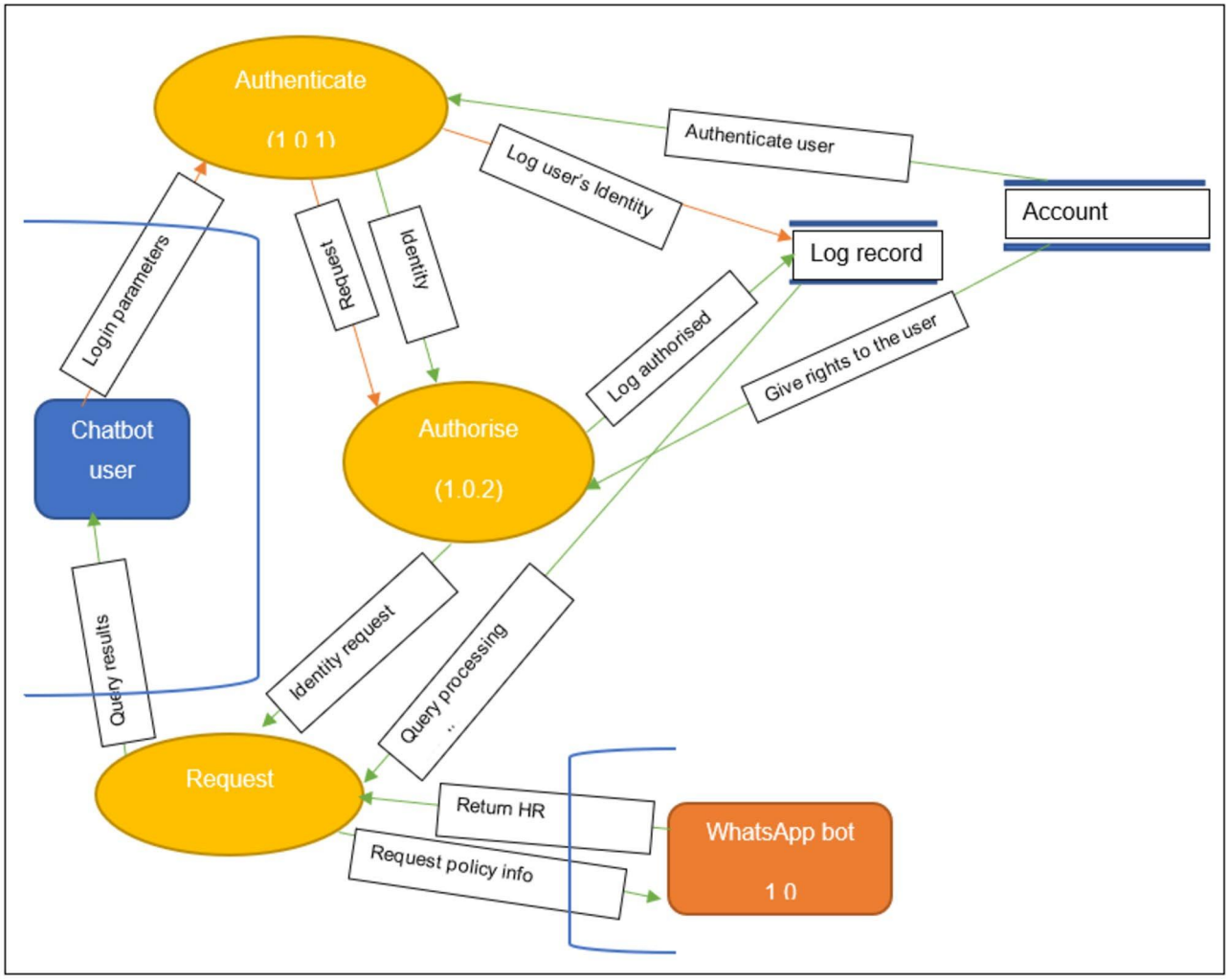


**Figure 14.** WhatsApp bot user interaction.

| Security goal<br>To provide for the confidentiality, integrity, and availability of the organisation's information assets by ensuring that users are authenticated and authorised to access these information assets | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sources of security concern<br>Common Adversarial Threats such as Disgruntled/disaffected employees, Dissatisfied customers, Extremist groups/terrorists, Hackers/hacking groups, Hacktivists, Investigative journalists, Nation-states, Organised criminal groups, Rogue suppliers/vendors/partners, Unscrupulous competitors | | | | | | | |
| **Security vulnerabilities** | **S** | **T** | **R** | **I** | **D** | **E** | **Chatbots** |
| Errors in business logic and business processes | √ | X | X | X | X | X | A1, A2, A3, A4 |
| Social engineering attacks | √ | X | X | X | X | X | A5 |
| Interception of communication | √ | X | X | X | X | X | A1, A2, A3, A4, A5 |
| Unauthorised access | √ | X | X | X | X | X | A1, A2, A3, A4, A5 |
| Attack through bad design or mis-configuration | √ | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| Phishing attacks | √ | X | X | X | X | X | A1, A2, A3, A4, A5 |
| Manipulation of employees by attackers | √ | X | X | X | X | X | A1, A2, A3, A4, A5 |
| Gain access to information assets in transit (network sniffing) | X | √ | X | X | X | X | A1, A2, A3, A4, A5 |
| Information compromise | X | √ | X | X | X | X | A1, A2, A3, A4, A5 |
| Impersonating real users | X | X | √ | X | X | X | A1, A2, A3, A4, A5 |
| The insecure disposal of assets | X | X | X | √ | X | X | A1, A2, A3, A4, A5 |
| Impair the availability or performance of the organisation's applications | X | X | X | X | √ | X | A1, A2, A3, A4, A5 |
| Introduce malware into the application | X | X | X | X | √ | X | A1, A2, A3, A4, A5 |
| Badly designed network architecture | X | X | X | X | √ | X | A1, A2, A3, A4, A5 |
| Physical damage or tampering with the company's data application | X | X | X | X | √ | X | A1, A2, A3, A4, A5 |
| Stealing of physical infrastructure | X | X | X | X | √ | X | A1, A2, A3, A4, A5 |
| Loopholes in the authorisation mechanisms | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| Unauthorised scanning or probing of a company's data applications | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| Unauthorised analysis of publicly available information about an enterprise | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |

**Table 10.** Result of analysis of security vulnerabilities based on the components of STRIDE.

| Security goal<br>To provide for the confidentiality, integrity, and availability of the organisation's information assets by ensuring that users are authenticated and authorised to access these information assets | | | | | | | |
|---|---|---|---|---|---|---|---|
| Sources of security concern<br>Common Adversarial Threats such as Disgruntled/disaffected employees, Dissatisfied customers, Extremist groups/terrorists, Hackers/hacking groups, Hacktivists, Investigative journalists, Nation-states, Organised criminal groups, Rogue suppliers/vendors/partners, Unscrupulous competitors | | | | | | | |
| **Possible security threats** | **S** | **T** | **R** | **I** | **D** | **E** | **Chatbots** |
| An attacker can exploit insecure default configuration | √ | √ | √ | √ | **X** | √ | A1, A2, A3, A4, A5 |
| An attacker can access the system from an unauthorised network | √ | X | X | X | X | X | A1, A2, A3, A4, A5 |
| An attacker can execute administrator functions | √ | X | X | X | X | X | A1, A2, A3, A4, A5 |
| An attacker can gain unauthorised access | √ | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| Insecure network infrastructure | √ | √ | √ | X | √ | X | A1, A2, A3, A4, A5 |
| Attacker penetrating an enterprise infrastructure | √ | √ | √ | X | X | X | A1, A2, A3, A4, A5 |
| An attacker can make organisations lose control over protecting sensitive data | X | √ | √ | √ | X | X | A1, A2, A3, A4, A5 |
| Hackers scan for vulnerabilities to exploit remote software | X | √ | X | √ | X | √ | A1, A2, A3, A4, A5 |
| Scanning internet address space | X | √ | X | X | X | X | A1, A2, A3, A4, A5 |
| Hacking new insecure configurations and unpatched assets | X | √ | X | X | X | X | A1, A2, A3, A4, A5 |
| Weak security on unidentified assets | X | √ | X | X | X | X | A1, A2, A3, A4, A5 |
| An attacker can bypass security controls | X | √ | X | X | X | √ | A1, A2, A3, A4, A5 |
| Lack of adequate logging and regular log review in applications | X | √ | √ | X | √ | X | A1, A2, A3, A4, A5 |
| Poor authentication and incorrect testing | X | √ | X | X | X | X | A1, A2, A3, A4, A5 |
| Ransomware attacker from the third party | X | X | X | √ | X | X | A1, A2, A3, A4, A5 |
| An attacker can make changes to configurations | X | X | X | X | √ | X | A1, A2, A3, A4, A5 |
| Proactive attack on newly published vulnerabilities | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| An attacker can target unknown vulnerabilities | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| Malicious attack | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| An attacker can exploit a specific vulnerability through insecure design and infrastructure | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |
| An attacker can exploit a specific vulnerability through poor design and incorrect testing | X | X | X | X | X | √ | A1, A2, A3, A4, A5 |

**Table 11.** Result of analysis of security vulnerabilities based on the elements of STRIDE.

| S/n | STRIDE components | No. of vulnerabilities | No of threats |
|---|---|---|---|
| 1 | Spoofing | 7 | 6 |
| 2 | Tampering | 2 | 10 |
| 3 | Repudiation | 1 | 5 |
| 4 | Information Disclosure | 1 | 4 |
| 5 | Denial of Service | 5 | 3 |
| 6 | Elevation of Privilege | 4 | 7 |

**Table 12.** Summary of associated security vulnerabilities and threats.

## Data availability

The data used for this study are not publicly available due to confidentiality, but are available from the corresponding author Olawande Daramola (wande.daramola@up.ac.za) on reasonable request.

Received: 23 January 2024; Accepted: 29 July 2024
Published online: 02 August 2024

## References

1. Cummins, J. D., Tennyson, S. & Weiss, M. A.: Efficiency, Scale Economies, and Consolidation in the US Life Insurance Industry. Financial Intitution Center (1998).
2. Alavudeen, R. & Rosa, K. D. Growing role of bancassurance in the banking sector. *Bonfring Int. J. Ind. Eng. Manag. Sci.* **5**(2), 10–16 (2015).
3. Ditizio, A. A. & Smith, A. D. Transformation of CRM and supply chain management techniques in a new venture (2017). https://www.igi-global.com/chapter/transformation-of-crm-and-supply-chain-management-techniques-in-a-new-venture/166517
4. IBM: how-chatbots-reduce-customer-service-costs-by-30-percent (2017). https://www.ibm.com/blogs/watson/2017/10/how-chatbots-reduce-customer-service-costs-by-30-percent/.
5. Tok, Y. C., Chattopadhyay, S.: Identifying threats, cybercrime and digital forensic opportunities in smart city infrastructure via threat modeling (2022).

6. Murugesan, S.: The cybersecurity renaissance: Security threats, risks, and safeguards (2019)
7. Zhang, Z., Li, B. & Liu, L. The impact of AI-based conversational agent on the firms' operational performance: Empirical evidence from a call center. *Appl. Artif. Intel.* **37**(1), 2157592. https://doi.org/10.1080/08839514.2022.2157592 (2023).
8. Lai, S., Leu, F. & Lin, J. A banking chatbot security control procedure for protecting user data security and privacy. BDET 2018, Chengdu, China (2019).
9. Følstad, A., Nordheim, C.B., Bjørkli C.A. What makes users trust a chatbot for customer service? An Exploratory Interview Study. International Conference on Internet Science. St. Petersburg, Russia (2018).
10. Hristidis, V. Chatbot Technologies and Challenges, First International Conference on Artificial Intelligence for Industries (AI4I). https://doi.org/10.1109/AI4I.2018.8665692, pp. 126–126 (2018).
11. Cardona, D. R., Werth, O., Schönborn, S., Breitner, M. H. A mixed-methods analysis of the adoption and diffusion of chatbot technology in the German Insurance Sector. Proceedings of the 25th Americas Conference on Information Systems (AMCIS). Cancun, Mexico (2019).
12. Shabbir, J., Anwer, T. Artificial Intelligence and its Role in Near Future. Cornell University (2018).
13. Bozic, J., Wotawa, F. Planning-based Security Testing for Chatbots. 30th IFIP International Conference on Testing Software and Systems. Spain (2018).
14. Ye, W., Li, Q. Chatbot Security and Privacy in the Age of Personal Assistants. 2020 IEEE/ACMSymposium on Edge Computing (SEC) 2020. 388–393 (2020) https://doi.org/10.1109/SEC50012.2020.00057
15. Yang, J., Chen, Y. L., Por, L. Y. & Ku, C. S. A systematic literature review of information security in chatbots. *Appl. Sci.* **13**(11), 6355 (2023).
16. Bhuiyan, M.S.I., Razzak, A., Ferdous, M. S., Chowdhury, M. J. M., Hoque, M. A., Tarkoma, S. BONIK: A Blockchain-Empowered Chatbot for Financial Transactions. IEEE 19th International Conference on Trust, Security, and Privacy in Computing and Communications 1079–1088 (2020). https://doi.org/10.1109/trustcom50675.2020.00143
17. Wube, H. D., Esubalew, S. Z., Weldesellasie, F. F. & Debelee, T. G. Text-based chatbot in financial sector: A systematic literature review. *Data Sci. Financ. Econ* **2**(3), 232–259 (2022).
18. Gebert-Persson, S., Gidhagen, M., Sallis, J. E. & Lundberg, H. Online insurance claims: When more than trust matters. *Int. J. Bank Marketing* **37**(2), 579–594 (2019).
19. Cardona, D. R., Janssen, A., Guhr, N., Breitner, M. H., & Milde, J. (2021). A matter of trust? Examination of chatbot usage in insurance business. in Proceedings of the Annual Hawaii International Conference on System Sciences (Vol. 2020-January, pp. 556–565). IEEE Computer Society. https://doi.org/10.24251/hicss.2021.068.
20. Koetter, F., Blohm, M., Drawehn, J., Kochanowski, M., Goetzer, J., Graziotin, D., & Wagner, S. (2019). Conversational agents for insurance companies: from theory to practice. In Agents and Artificial Intelligence: 11th International Conference, ICAART 2019, Prague, Czech Republic, February 19–21, 2019, Revised Selected Papers 11 (pp. 338–362). Springer International Publishing, (2019).
21. Hussain. S., Kamal, A., Ahmad, S., Rasool, G., Iqbal, S. Threat modelling methodologies: A survey (2014).
22. Wilhjelm, C., Younis, A. A. A Threat analysis methodology for security requirements elicitation in machine learning based systems. IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C) (2020)
23. Lohmann, P. A., Albuquerque, C., & Machado, R. Systematic Literature Review of Threat Modeling Concepts. ICISSP, 163–173, (2023).
24. Yeng, P.K., Wulthusen, S.D. & Bian, Y. Comparative Analysis of Threat Modeling Methods for Cloud Computing towards Healthcare Security Practice. International Journal of Advanced Computer Science and Applications, vol. 11, no. 11 (2020).
25. Shevchenko, N., Chick, T. A., O'Riordan, P., Scanlon, T. P., & Woody, C. Threat modeling: a summary of available methods. Software Engineering Institute| Carnegie Mellon University (2018).
26. Rowley, S., Slack, F.: Conducting a literature review. The American journal of maternal child nursing 27(6) (2004)
27. Cummins, J. D., Doherty, N. A.: The Economics of Insurance Intermediaries (2006).
28. Sibindi, A. & Godi, N. J. Insurance sector development and economic growth: Evidence from South Africa. *Corporate Ownership Control* **11**(4), 530–538 (2014).
29. Roberts-Lombard, M. Exploring the relationship between trust, commitment and customer loyalty through the intervening role of customer relationship management (CRM). *Afr. J. Business Manag.* **6**(10) (2012).
30. Kanchinadam, T., Qazi, M., Bockhorst, J., Morell, M., Meissner K. & Fung, G. Using discriminative graphical models for insurance recommender systems. in *17th Proceedings IEEE International Conference on Machine Learning and Applications, ICMLA 2018*. IEEE, (5), pp. 421–428 (2019).
31. Riikkinen, M., Saarijärvi, H., Sarlin, P. & Lähteenmäki, I. Using artificial intelligence to create value in insurance. *Int. J. Bank Marketing* **36**(6), 1145–1168 (2018).
32. Singh, A., Ramasubramanian, K., Shivam, S., Singh, A., Ramasubramanian, K., & Shivam, S. Processes in the Banking and Insurance Industries. Building an Enterprise Chatbot: Work with Protected Enterprise Data Using Open Source Frameworks. 1–18 (2019).
33. Meltzer, M. A customer relationship management approach: Integrating the call centre with customer information. *J. Database Marketing* **8**(3), 232–243 (2001).
34. Raikwar, M., Mazumdar, S., Ruj, S. Gupta, S. S., Chattopadhyay, A., Lam, K.: A Blockchain Framework for Insurance Processes. 2018 9th IFIP International Conference on New Technologies, Mobility, and Security (NTMS). 26–28 Feb. 2018 (2019). https://ieeexplore.ieee.org/abstract/document/8328731
35. Ondrisek, B.: Why You Shouldn't Talk to Your Chatbot about Everything (2016) http://venturebeat.com/2016/11/17/why-you-shouldnt-talk-to-your-chatbot-about-everything/
36. Cahn, J.: CHATBOT: Architecture, Design, & Development. Senior Thesis (EAS499) University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science (2017).
37. Mott, N.: Ticketmaster Blames Malware-Plagued Chatbot for Data Breach (2018). https://www.tomshardware.com/news/ticketmaster-data-breach-uk-international,37383.html.
38. Khan, R. Standardised architecture for conversational agents aka chatbots. *Int. J. Computer Trends Technol.* **50**(2), 114–121 (2017).
39. Xiong, W. & Lagerström, R. Threat modelling—A systematic literature review. *Comput. Security* **84**, 53–69 (2019).
40. Microsoft (2017) https://www.microsoft.com/en-us/securityengineering/sdl/threatmodeling
41. Hasal, M. *et al.* Chatbots: Security, privacy, data protection, and social aspects. *Concurrency Comput. Practice Experience* **33**(19), e6426 (2021).
42. Ng, M., Coopamootoo, K. P. L., Ehsan, T., Aitken, M., Elliott, K., Moorsel, A. V.: Simulating the Effects of Social Presence on Trust, Privacy Concerns & Usage Intentions in Automated Bots for Finance. 2020 IEEE European Symposium on Security and Privacy Workshops 2021: 190–199 (2021) https://doi.org/10.1109/EuroSPW51379.2020.00034
43. Harkous, H., Shin, K. G., Fawaz, K. & Aberer, K.: PriBots: Conversational Privacy with Chatbots. Workshop on the Future of Privacy Indicators, at the Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016, June 22–24 (2016).
44. CIS Controls Version 8. https://paper.bobylive.com/Security/CIS/CIS_Controls_v8_Guide.pdf (2021).
45. IRAM2 The next generation of assessing information risk (2014). https://www.securityforum.org/solutions-and-insights/information-risk-assessment-methodology-iram2/
46. Casola, V., De Benedictis, A., Rak, M. & Villano, U. A novel Security-by-Design methodology: Modeling and assessing security by SLAs with a quantitative approach. *J. Syst. Softw.* **163**, 110537 (2020).

47. Daramola, O., Sindre, G. & Moser, T. A tool-based semantic framework for security requirements specification. *J. Universal Comput. Sci.* **19**(13), 1940–1962 (2013).
48. Salini, P. & Kanmani, S. Survey and analysis on Security Requirements Engineering. *Comput. Electr. Eng.* **38**(6), 1785–1797. https://doi.org/10.1016/j.compeleceng.2012.08.008 (2012).

## Acknowledgements
The authors acknowledge the support provided for the study by the Cape Peninsula University of Technology (CPUT), South Africa, and the University of Pretoria, South Africa.

## Author contributions
The article was conceived and structured by all the authors. Z.B — Investigation, Methodology, Data curation, Data analysis, Validation, Writing – original draft. O. D —Conceptualization, Methodology, Data analysis, Validation, Supervision, Writing – final draft. All the authors reviewed the manuscript.

## Competing interests
The authors declare no competing interests.

## Additional information
**Correspondence** and requests for materials should be addressed to O.D.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.