Motivations, Classification and Model Trial of Conversational Agents for Insurance Companies

Falko Koetter[1], Matthias Blohm[1], Monika Kochanowski[1], Joscha Goetzer[2], Daniel Graziotin[2], and Stefan Wagner[2]

[1]*Fraunhofer Institute for Industrial Engineering, Nobelstr. 12, 70569 Stuttgart, Germany*

[2]*University of Stuttgart, Universitätsstr. 38, 70569 Stuttgart, Germany*

falko.koetter@iao.fraunhofer.de, matthias.blohm@iao.fraunhofer.de, monika.kochanowski@iao.fraunhofer.de, joscha.goetzer@gmail.com, daniel.graziotin@iste.uni-stuttgart.de, stefan.wagner@iste.uni-stuttgart.de

Abstract: Advances in artificial intelligence have renewed interest in conversational agents. So-called chatbots have reached maturity for industrial applications. German insurance companies are interested in improving their customer service and digitizing their business processes. In this work we investigate the potential use of conversational agents in insurance companies by determining which classes of agents are of interest to insurance companies, finding relevant use cases and requirements, and developing a prototype for an exemplary insurance scenario. Based on this approach, we derive key findings for conversational agent implementation in insurance companies.

1 INTRODUCTION

With the digital transformation changing usage patterns and consumer expectations, many industries need to adapt to new realities. The insurance sector is next in line to grapple with the risks and opportunities of emerging technologies, in particular *Artificial Intelligence* (Nordman et al., 2017).

Fraunhofer IAO as an applied research institution supports digital transformation processes in an ongoing project with multiple insurance companies (Renner and Kochanowski, 2018). The goal of this project is to scout new technologies, investigate them, rate their relevance and evaluate them (e.g. in a model trial or by implementing a prototype). While insurance has traditionally been an industry with very low customer engagement, insurers now face a young generation of consumers with changing attitudes regarding insurance products and services (Pohl et al., 2017).

Traditionally, customer engagement uses channels like mail, telephone and local agents. In 2016, chatbots emerged as a new trend (Guzmn and Pathania, 2016), making it a topic of interest for Fraunhofer IAO and insurance companies.

With the rise of the smartphone, many insurers started offering apps, but success was limited (Power, J. D., 2017), which may stem from app fatigue (Schippers, 2016). App use has plateaued, as users have too many apps and are reluctant to add more (Gartner, 2015). In contrast, conversational agents require no separate installation, as they are accessible via messaging apps, which are likely to be already installed on a user's smartphone. Conversational agents are an alternative to improve customer support and digitize processes like claim handling.

The objective of this work is to facilitate the creation of conversational agents by defining the traits of an agent more clearly using a (1) classification framework, which is based on current literature and research topics, and systematically analyzing (2) use cases and requirements in an industry, shown in the example insurance scenario. The applicability of this approach is shown by implementing a prototype (3) including an evaluation. Furthermore, we derive key findings for conversational agent implementation in insurance companies and open points for research.

2 RELATED WORK

In this section we investigate work in the area of conversational agents, dialog management, and research applications in insurance.

McTear et al. (2016) offer detailed explanations about background and history of conversational in-

terfaces as well as techniques to build and evaluate own agent applications. Another literature review about chatbots was provided by Cahn (2017), where common approaches and design choices are summarized followed by a case study about the functioning of IBM's chatbot Watson, which became famous for winning the popular quiz game *Jeopardy!* against humans.

Many chatbot applications have already been built nowadays with the goal to solve actual problems. One example is PriBot, a conversational agent, which can be asked questions about an application's privacy policy, because users tended to skip reading the often long and difficult to understand privacy notices. Also, the chatbot accepts queries of the user which aim to change his privacy settings or app permissions (Harkous et al., 2016).

In the past there have already been several studies with the goal to evaluate how a conversational agent should behave for being considered as human-like as possible. In one of them, conducted by Kirakowski et al. (2009), fourteen participants were asked to talk to an existing chatbot and to collect key points of convincing and unconvincing characteristics. It turned out that the bot's ability to hold a theme over a longer dialog made it more realistic. On the other hand, not being able to answer to a user's questions was regarded as an unsatisfying characteristic of the artificial conversational partner (Kirakowski et al., 2009).

In another experiment, which was done by Sörensen (2017), eight users had to talk to two different kinds of chatbots, one behaving more human-like and one behaving more robotic. In this context, they had to fulfill certain tasks like ordering an insurance policy or demanding an insurance certification. All of the participants instinctively started to chat by using natural human language. In cases in which the bot did not respond to their queries in a satisfying way, the users' sentences continuously got shorter until they ended up with writing key words only. Thus, according to the results of this survey, conversational agents preferably should be created human-like, because users seem to be more comfortable when feeling like talking to another human being, especially in cases in which the concerns are crucial topics like their insurance policies (Sörensen, 2017).

Dialog management strategies (DM) define the conversational behaviors of a system in response to user message and system state McTear et al. (2016).

In industry applications, DM often consists of a handcrafted set of rules and heuristics, which are tightly coupled to the application domain (McTear et al., 2016) and improved iteratively. One problem with handcrafted approaches to DM is that it is chal-lenging to anticipate every possible user input and react appropriately, making development resource-intensive and error-prone. But if few or no recordings of conversations are available, these *rule-oriented* strategies may be the only option.

As opposed to the rule-oriented strategies, data-oriented architectures work by using machine learning algorithms that are trained with samples of dialogs in order to reproduce the interactions that are observed in the training data. These statistical or heuristical approaches to DM can be classified into three main categories: Dialog modeling based on *reinforcement learning*, *corpus-based* statistical dialog management, and *example-based* dialog management (simply extracting rules from data instead of manually coding them) (McTear et al., 2016; Spierling, 2005). Spierling (2005) highlights neural networks, Hidden-Markov Models, and Partially Observable Markov Decision Processes as possible implementation technologies.

The following are common strategies for rule-based dialog management:

- Finite-state-based DM uses a finite state machine with handcrafted rules, and performs well for highly structured, system-directed tasks (McTear et al., 2016).

- Frame-based DM follows no predefined dialog path, but instead allows to gather pieces of information in a frame structure and no specific order. This is done by adding an additional entity-value slot for every piece of information to be collected and by annotating the intents in which they might occur. Using frames, a less restricted, user-directed conversation flow is possible, as data is captured as it comes to the mind of the user (Rudnicky and Xu, 1999).

- Information State Update represents the information known at a given state in a dialog and updates the internal model each time a participant performs a *dialog move*, (e.g. asking or answering). The state includes information about the mental states of the participants (beliefs, desires, intentions, etc.) and about the dialog (utterances, shared information, etc.) in abstract representations. Using so-called update moves, applicable moves are chosen based on the state (Traum and Larsson, 2003).

- Agent-based DM uses an agent that fulfills conversation goals by dynamically using plans for tasks like intent detection and answer generation. The agent has a set of beliefs and goals as well as an information base which is updated throughout the conversation. Within this informa-

tion framework the agent continuously prioritizes goals and autonomously selects plans that maximize the likelihood of goal fulfillment (Nguyen and Wobcke, 2005).

Chu et al. (2005) describes how multiple DM approaches can be combined to use the best strategy for specific circumstances.

A virtual insurance conversational agent is described by Yacoubi and Sabouret (2018), utilizing *TEATIME*, an architecture for agent-based DM. TEATIME uses emotional state as a driver for actions, e.g. when the bot is perceived unhelpful, that emotion leads the bot to apologize. The shown example bot is a proof of concept for TEATIME capable of answering questions regarding insurance and react to customer emotions, but does not implement a full business process.

Kowatsch et al. (2017) describe a text-based healthcare chatbot that acts as a companion for weightloss but also connects a patient with healthcare professionals. The chat interface supports non-textual inputs like scales and pictorials to gather patient feedback. Study results showed a high engagement with the chatbot as a peer and a higher percentage of automated conversation the longer the chatbot is used.

Overall, these examples show potential for conversational agents in the insurance area, but lack support for complete business processes.

3 CLASSIFICATION OF CONVERSATIONAL AGENTS

The idea of conversational agents that are able to communicate with human beings is not new: In 1966, Joseph Weizenbaum introduced *Eliza*, a virtual psychotherapist, which was able to respond to user queries using natural language and which could be considered as the first *chatbot* (Weizenbaum, 1966). Nowadays, the idea of speaking machines has experienced a revival with the emergence of new technologies, especially in the area of artificial intelligence. Novel machine learning algorithms allow developers to create software agents in a much more sophisticated way and in many cases they already outperform previous statistical NLP methods (McTear et al., 2016). Additionally, the importance of messaging apps such as WhatsApp or Telegram has increased over the last years. In 2015, the total number of people using these messaging services outran the total number of active users in social networks for the first time. Today, each of these app has about between 200 million and 1.5 billion users (Inc, 2018).

As a highly popular topic in 2016 (Guzmn and Pathania, 2016), a great variety of different chatbots evolved together with an equally wide range of terminologies. For being able to draw a big picture of the current trends in the area of conversational agents, we divide them into the following four common categories:

- **Chatterbots**: Bots with focus on small talk and realistic conversations, not task-oriented, e.g. Cleverbot (Carpenter, 2018).
- **(Virtual, Intelligent, Cognitive, Digital, Personal) assistants (VPAs)**: Agents fulfilling tasks intelligently based on spoken or written user input and with the help of data bases and personalized user preferences (Cooper et al., 2008) (e.g. Apple's Siri or Amazon's Alexa (Dale, 2016)).
- **Specialized digital assistants (SDAs)**: Focused on a specific domain of expertise, goal-oriented behavior (Dale, 2016).
- **Embodied conversational agents (ECAs)**: Visually animated agents, e.g. in form of avatars or robots (Radziwill and Benton, 2017), where speech is combined with gestures and facial expressions.

Figure 1 shows the results of evaluating these four classes in terms of different characteristics such as *realism* or *task orientation* based on own literature research. Chatterbots provide a high degree of entertainment since they try to imitate the behavior of human beings while chatting, but there is no specific goal to be reached within the scope of these conversations. In contrast, general assistants like Siri or Alexa are usually called by voice in order to fulfill a specific task. Specialized assistants concentrate even more on achieving a specific goal, which often comes at the expense of realism and user amusement because their ability to respond to not goal-oriented conversational inputs like small talk is mostly limited. The best feeling of companionship can be experienced by talking to an embodied agent, since the reactions of these bots are closest to human-like behavior.

When looking at these classification results, a broad spectrum of various possible agents is offered. Therefore, a restriction depending on the specific use case has to be made first, before the realization of a prototypical chatbot can be tackled. Since Fraunhofer IAO aims to investigate solutions supporting processes in the insurance domain, creating a prototype with the properties of a SDA is necessary, because the main purpose in this scenario is to perform and successfully complete a certain task (e.g. reporting a claim). Furthermore, adding additional chatterbot features such as the ability to do small talk in a

limited goal-oriented scope could lead to a more realistic and human-like user experience. However, before considering detailed design and implementation choices, it is helpful to take a general look at the role of chatbots within the special environment of insurance companies for identifying essential issues and needs.
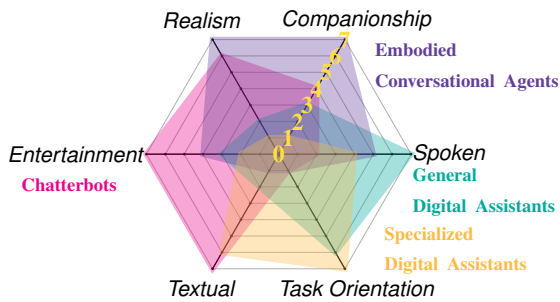

Figure 1: Classification of conversational agents with their characteristics (own presentation). Values between 0 and 7 indicate how strong a characteristic applies for the given type of agent.

4 CHATBOTS IN INSURANCE

Insurance is an important industry sector in Germany, with 560 companies that manage about 460 million policies (Schwark and Theis, 2014). However, the insurance sector is under a high cost pressure, which shows in a declining employee count and low margins (Stange and Reich, 2015). The insurance market is saturated and has transitioned from a growth market to a displacement market (Aschenbrenner et al., 2010). For the greater part, German insurance companies have used conservative strategies, caused by risk aversion, long-lived products, hierarchical structures, and profitable capital markets (Zimmermann and Richter, 2015). As these conditions change, so must insurance companies.

Insurance is an industry with low customer engagement, as an insurer traditionally has basically two touch points to interact with customers: *selling a product* and *the claims process*. A study found that consumers interact less with insurers than with any other industry, so the consumer experience with insurers tends to lag behind others (Niddam et al., 2014).

Many insurance companies have heterogeneous IT infrastructures incorporating legacy systems (sometimes from two or more companies as the result of a merger) (Weindelt, 2016). These grown architectures pose challenges when implementing new data-driven or AI solutions, due to issues like data quality, availability and privacy. Nonetheless, the high amount of

available data and complex processes make insurance a prime candidate for machine learning and data mining. The adoption of AI in the insurance sector is in early stages, but accelerating, as insurance companies strive to improve service and remain competitive (Nordman et al., 2017).

Conversational agents are one AI technology at the verge of adoption. In 2017, ARAG launched a travel insurance chatbot, quickly followed by bots from other insurance companies (Gorr, 2018). While these chatbots are still experimental and implement narrow use cases, these first implementations prove public interest and feasibility.

To identify areas of possible chatbot support, we surveyed the core business processes of insurance companies as described in Aschenbrenner et al. (2010) and Horch et al. (2012). Three core areas of insurance companies are customer-facing: *marketing/sales*, *contract management* and *claim management*. Figure 2 shows the main identified processes related to this area.
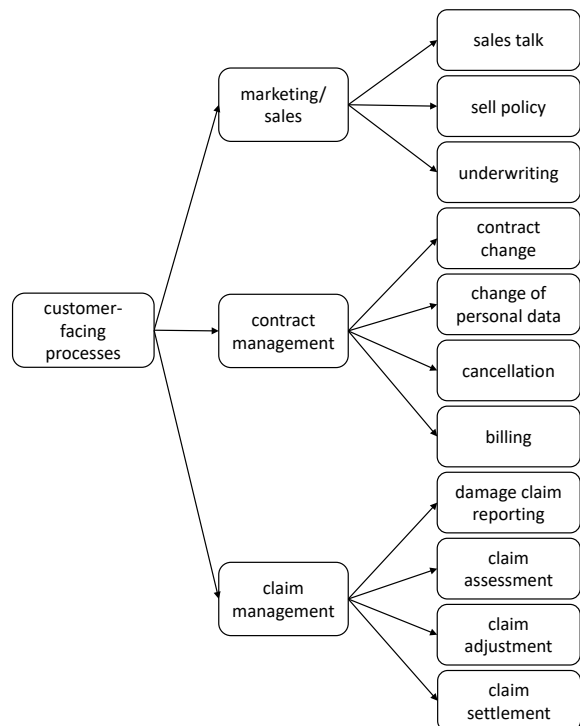

Figure 2: Customer-facing insurance processes (based on Aschenbrenner et al. (2010) and Horch et al. (2012))

We identified all these processes as possible use cases for conversational agent support, in particular support by SDAs.

Furthermore, we investigated general requirements for conversational agents in these processes:

Availability and ease-of-use Conversational agents

are an alternative to both conventional customer support (e.g. phone, mail) as well as conventional applications (e.g. apps and websites). Compared to these conventional solutions, chatbots offer more availability than human agents and have less barriers of use than conventional applications, requiring neither an installation nor the ability to learn a new user interface, as conventional messaging services are used (Derler, 2017).

Guided information flow Compared to websites, which offer users a large amount of information they must filter and prioritize themselves, conversational agents offer information gradually and only after the intent of the user is known. Thus, the search space is narrowed at the beginning of the conversation without the user needing to be aware of all existing options.

Smartphone integration Using messaging services, conversational agents can integrate with other smartphone capabilities, e.g. making a picture, sending a calendar event, setting a reminder or calling a phone number.

Customer call reduction Customer service functions can be measured by reduction of customer calls and average handling time (Guzmn and Pathania, 2016). SDAs can help here by automating conversations, handling standard customer requests and performing parts of conversations (e.g. authentication).

Human handover Customers often use social media channels to escalate an issue in the expectation of a *human* response, instead of an *automated* one. A conversational agent thus must be able to differentiate between standard use cases it can handle and more complicated issues, which need to be handed over to human agents (Newlands, 2017). One possible approach is to use sentiment detection, so customer who are already stressed are not further aggravated by a bot (Guzmn and Pathania, 2016).

Digitize claim handling Damage claim handling in insurance companies is a complex process involving multiple departments and stakeholders (Koetter et al., 2012). Claim handling processes are more and more digitized within the insurance companies (Horch et al., 2012), but paper still dominates communication with claimants, workshops and experts. (Fannin and Brower, 2017) defines maturity levels of insurance processes, defining *virtual handling* as a process where claims are assessed fully digitally based on digital data from the claimant (e.g. a video, a filled digital form), and *touchless handling* as a fully digital process with no human intervention on the insurance side. SDAs help moving towards these maturity levels by providing a guided way to make a claim digitally and communicate with the claimant (e.g. in case additional data is needed).

Conversational commerce is the use of Conversational Agents for marketing and sales related purposes (Eeuwen, 2017). Conversational Agents can perform multiple tasks using a single interface. Examples are using opportunities to sell additional products (*cross-sell*) or better versions of the product the customer already has (*up-sell*) by chiming in with personalized product recommendations in the most appropriate situations. One example would be to note that a person's last name has changed during an address update customer service case and offer appropriate products if the customer has just married.

Internationalization is an important topic for large international insurance companies. However, most frameworks for implementing conversational agents are available in more than one language. To the best of our knowledge today the applied conversational agents in German insurance are optimized only for one language. So this topic is future work in respect to the prototype.

Compliance to privacy (GDPR) is usually guaranteed by the login mechanisms on the insurance sites, therefore the topic is out of scope for our research prototype. For broader scenarios not requiring identification on the insurance site and the usage of the data for non-costumers, this is an open area of research.

5 PROTOTYPE

Based on the work presented in the last sections and our talks with insurance companies, we arrived at the following non-functional requirements that the chatbot prototype ideally should fulfill:

- **Interoperability**: The agent should be able to keep track of the conversational context over several message steps and messengers.

- **Portability**: The agent can be run on different devices and platforms (e.g. Facebook Messenger, Telegram). Therefore it should use a unified, platform-independent messaging format.

- **Extensibility**: The agent should provide a high level of abstraction that allows designers to add new conversational content without having to deal with complicated data structures or code.

Additionally, the following functional requirements should be regarded in the implementation:

- **Report a claim**: The system must provide the possibility for a user to report a damage claim using the conversational agent (prototype scenario).

- **Human language understanding**: The system should be able to understand and process the

user's inputs and intents given in form of written natural language (German).

- **Response generation**: The system should be able to generate an answer sentence in written human language (German) according to the user queries.

For dialog design within the prototype, experimenting with machine learning algorithms was the preferred implementation strategy. For this purpose, discussions with insurance companies were held to assess the feasibility of receiving existing dialogs with customers, for example for online chats, phone logs or similar. However, such logs generally seem to be not available at German insurers, as the industry has self-regulated to only store data needed for claim processing (GDV, 2012). As a research institute represents a third party not directly involved in claims processing, data protection laws forbid sharing of data this way without steps to secure personal data. During our talks we have identified a need for automated or assisted anonymization of written texts as a precondition for most customer-facing machine learning use cases, at least when operating in Europe (Kamarinou et al., 2016). However, these issues go beyond the scope of our current project, but provide many opportunities for future research.

To still build a demonstrator in face of these challenges, dialogs for the prototype were manually designed without using real-life customer conversations and fine-tuned by user testing with fictional damage claims. As this approach entails higher manual effort for dialog design, a narrower scenario was chosen to still allow for the full realization of a customer-facing process. The chosen scenario was a special case of the damage claim process: *The user has a damaged smartphone or tablet and wants to make an insurance claim.*

Figure 5 shows the main components of the prototype and their operating sequence when processing a user message. To provide extensibility prototype architecture strictly separates service integration, internal logic and domain logic.

The user can interact with the bot over different communication channels which are integrated with different *bot API clients*. To integrate a different messaging service, a new bot API client needs to be written. The remainder of the prototype can be reused.

Once a user has written a message, a lookup of *user context* is performed to determine if a conversation with that user is already in progress. User context is stored in a database so no state is kept within external messaging services. Afterwards, a *typing* notification is given to the user, indicating the bot has received the message and is working on it. This prevents multiple messages by a user who thinks the bot is not responsive.

In the next step, the message has to be understood by the bot. In case of a voice message, it is transcribed to text using a *Google* speech recognition web service.

For natural language understanding, we compared four possible frameworks (Microsoft's LUIS, Google's Dialogflow, Facebook's wit.ai and IBM's Watson) regarding important criteria for prototype implementation. A comparison table for these criteria is shown in Table 1. As a result of the comparison, Dialogflow was chosen as a basic framework.

Table 1: Comparison of Microsoft's LUIS, Google's Dialogflow, Facebook's wit.ai, and IBM's Watson (based on Davydova (2017))

	LUIS	Dialogflow	Wit.ai	Watson
Python bindings	no	yes	yes	yes
German language	yes	yes	in Beta	yes
Free service	no	yes	yes	no
Remember state	yes	yes	yes	yes
Service bound	yes	yes	yes	yes
Simple training	with effort	yes	yes	yes

Dialogflow is used for intent identification, which determines the function of a message and based on that a set of possible parameters (McTear et al., 2016). For example, the intent of the message "the display of my smartphone broke" may have the intent `phone_broken` with the parameter `damage_type` as `display_damage`, while the parameter `phone_type` is not given. Together, this information given by Dialogflow is a `MessageUnderstanding`

As soon as the message is understood, the user context is updated. Afterwards, a response needs to be generated. This process, which was labeled with *Plan and Realize Response* in Figure 5, is shown in detail in Figure 6.

In the prototype, an agent-based strategy was chosen in order to combine the capabilities of the frame-based entities and parameters in Dialogflow with a custom dialog controller based on predefined rules in a finite state machine. This machine allows to define rules that trigger *handlers* and *state transitions* when a specific intent or entity-parameter combination is encountered. That way, both intent and frame processing happen in the same logically encapsulated unit, enabling better maintainability and extensibility. The rules are instances of a set of `*Handler` classes such as an `IntentHandler` for the aforementioned intent and parameter matching, supplemented by other handlers, e.g. an `AffirmationHandler`, which consolidates different intents that all express a confirmation along the lines of "yes", "okay",

Figure 3: Dialog excerpt of the prototype, showing the possibility to clarify the phone model via multiple-choice input)

"good" and "correct", as well as a `NegationHandler`, a `MediaHandler` and an `EmojiSentimentHandler` (to analyze positive, neutral, or negative sentiment of a message with emojis). Each implements their own `matches(MessageUnderstanding)` method.

Rules (handlers) are used within the dialog state machine:

1. *Stateless* handlers are checked independently of the current state. For example, a `RegexHandler` rule determines whether the formality of the address towards the user should be changed (German differentiates the informal "du" and the formal "Sie")

2. *Dialog States* map each possible state to a list of handlers that are applicable in that state. For instance, when the user has given an answer and the system asks for *explicit confirmation* in a state `USER_CONFIRMING_ANSWER`, then an `AffirmationHandler` and a `NegationHandler` capture "yes" and "no" answers.

3. *Fallback* handlers are checked if none of the applicable state handlers have yielded a match for an incoming `MessageUnderstanding`. These `fallbacks` include static, predefined responses with lowest priority (e.g. small talk), as well as handlers to repair the conversation by bringing the user back on track or changing the topic.

At first, the system had only allowed a single state to be declared at the same time in the router. However, this had quickly proven to be insufficient as users are likely to want to respond or refer not only to the most recent message, but also to previous ones in the chat. With only a single contemporaneous state, the user's next utterance is always interpreted only in that state. In order to make this model resilient, every state would need to incorporate every utterance that the user is likely to say in that context. As this is not feasible, the prototype has state handlers that allow layering transitions on top of each other, allowing multiple simultaneous states which may advance individually.

To avoid an explosion of active states, the system has *state lifetimes*: new states returned by callbacks may have a lifetime that determines the number of dialog moves this state is valid for. On receiving a new message, the planning agent decreases the lifetimes of all current dialog states by one, except for the case of utter non-understanding ("fallback" intent). If a state has exceeded its lifetime, it is removed from the priority queue of current dialog states.

Figure 6 contains details about how the system creates responses to user queries. Based on the applicable rule, the conversational agent performs chat actions (e.g. sending a message), which are generated from response templates, taking into account dialog state, intent parameters, and information like a user's name, mood and preferred level of formality.

RuleHandlers, states and other dialog specific implementations are encapsulated, so a new type of dialog can be implemented without needing to change the other parts of the system.

Generated chat actions are stored in the user context and performed for the user's specific messenger using the *bot API*. As the user context has been updated, the next message by the user continues the conversation. Before the user utters an intent to make a damage claim, the prototype explains its functionality and of-

fers limited small talk. As soon as the user wants to make a damage claim, the conversational agent gathers the required information using a predetermined questionnaire. Questions concern type of damage, damaged phone, phone number, IMEI, damage time, damage event details, etc. Answers are interpreted using dialog flow (e.g. determining a point in time). Interpretation results have to be confirmed by the user. In case the answer is not understood, not correct or not confirmed, the question is repeated. Alternatively, for specific questions domain specific actions for clarification are implemented. For example, a choice for specific phone model is shown in Figure 3. For each question, users may ask for details or for an example answer. A skip intent is recognized and causes the dialog to advance to the next question if the current question is optional.

After the questionnaire is concluded, the bot thanks the user and stores the data. In a real-life application, claim management systems would be integrated to automatically trigger subsequent processes.

6 EVALUATION

To evaluate the produced prototype's quality and performance, we conducted a model trial with the goal to report a claim by using the chatbot without having any further instructions available.

Of the 14 participants (who all had some technical background), 35.7% claimed to regularly use chatbots, 57.1% to use them occasionally, and only 7.1% stated that they had never talked to a chatbot before. However, all participants were able to report a claim within a range of about four minutes, resulting in an overall task completion rate of 100%.

Additionally, the users had to rate the quality of their experiences with the conversational agent by filling out a questionnaire. For each question they could assign points between 0 (did not apply at all) and 10 (did apply to the full extent). The most important quality criteria, whose choice was oriented on the work of Radziwill and Benton (2017), are listed with their average ratings in Figure 4 and are discussed in detail. With an average of 8 points for *Ease of Use*, the users had no problems with using the bot to solve the task. In the same way, 8.3 points for *Appropriate Formality* indicate that the participants were comfortable with the formal and informal language the bot talked to them. Only one user stated that he felt worried about permanently being called by his first name after he told it. Fewer points were given for the bot's degree of human-like behavior: The rating for convincing *Natural Interaction* with 7.9 points may be

due to the fact that the conversation was designed in a strongly questionnaire-oriented way, which might have restricted the feeling of having a free user conversation. Also, the satisfaction with given answers to users' domain specific questions was considered quite (but not totally) convincing with 7.6 points. The least convincing experience was that chatbot's *Personality*, which was rated with only 5.2 points on average. This is not surprising, since during this work we put comparatively less efforts in strengthening the agent's personal skills as it does not even introduce itself with a name, but instead mainly acts on a professional level, always concentrating on the fulfillment of its task. With 7.2 points, talking to the chatbot was experienced as quite *Funny & Interesting*, but still with a lot of room for further improvement. Similarly, the agent's *Entertainment* capabilities, which are at 7.7 points on average at the moment, could be upgraded by extending the conversational contents with additional enjoyable features not related to the questionnaire. For the future we plan to do another larger evaluation on a bigger and more heterogeneous group of participants.
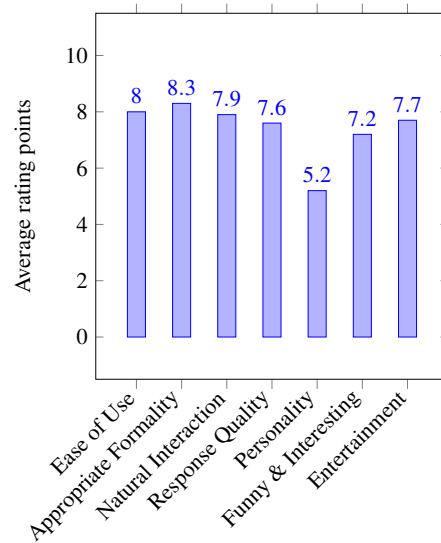

Figure 4: Survey results: average user experience ratings (fourteen participants, 0..10 points).

7 CONCLUSIONS AND OUTLOOK

In this work we have investigated the potential of using conversational agents in insurance companies by general research and implementing a prototype. We determined which classes of conversational agents are

useful for insurance companies, which insurance processes can be supported, and what the requirements and motivations for using conversational agents in insurance are. These findings can be used to facilitate the development of conversational agents. We found a need for *Specialized Digital Assistants* in customer facing processes.

Based on these findings we formulated requirements for conversational agents in insurance and selected the smartphone damage claim as an example scenario. We implemented this scenario in a prototype, using machine learning for intent recognition but relying on manual dialog design. Instead of a single dialog state, we implemented a system of multiple conversational states enabling more flexible conversations. We evaluated our prototype with real users and gathered their reactions with a questionnaire. Overall, we found that the prototype is able to handle the example scenario to the user's satisfaction. Possible improvements in the prototype scenario are a better determination of the desired degree of formality as well as defining a consistent persona for the agent (a first step would be to provide a name).

The findings indicate technology is ready to implement conversational agents for insurance customer service scenarios. However, as real-life scenarios are broader than the example scenario, considerably more effort is necessary to design the dialogs. One area not covered by the prototype is human handover in case the conversational agent cannot complete and interaction to the user's satisfaction.

Data protection and privacy remain open areas for research. Legal and practical questions regarding data collection, storage and processing must be worked on alongside technical requirements, as they tend to be complex and have limited precedent (Smart-Data-Begleitforschung, 2018).

In future research, we would like to extend the prototype to different scenarios as well as perform a real-life evaluation with an insurance partner to quantify the benefits of agent use, e.g. call reduction, success rate, and customer satisfaction.

REFERENCES

Aschenbrenner, M., Dicke, R., Karnarski, B., and Schweiggert, F. (2010). *Informationsverarbeitung in Versicherungsunternehmen*. Springer.

Cahn, J. (2017). Chatbot: Architecture, design, & development.

Carpenter, R. (2018). Cleverbot. https://www.cleverbot.com/.

Chu, S.-W., O'Neill, I., Hanna, P., and McTear, M. (2005). An approach to multi-strategy dialogue management. In *Ninth European Conference on Speech Communication and Technology*, pages 865–868.

Cooper, R. S., McElroy, J. F., Rolandi, W., Sanders, D., Ulmer, R. M., and Peebles, E. (2008). Personal virtual assistant. US Patent 7,415,100.

Dale, R. (2016). Industry watch: The return of the chatbots. *Natural Language Engineering*, 22(5):811–817.

Davydova, O. (2017). 25 chatbot platforms: A comparative table. https://chatbotsjournal.com/25-chatbot-platforms-a-comparative-table-aeefc932eaff.

Derler, R. (2017). Chatbot vs. app vs. website chatbots magazine. https://chatbotsmagazine.com/chatbot-vs-app-vs-website-en-e0027e46c983.

Eeuwen, M. (2017). Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers. Master's thesis, University of Twente.

Fannin, T. and Brower, B. (2017). 2017 future of claims study. Technical report, LexisNexis.

Gartner (2015). Market trends: Mobile app adoption matures as usage mellows. https://www.gartner.com/newsroom/id/3018618.

GDV (2012). Verhaltensregeln für den Umgang mit personenbezogenen Daten durch die deutsche Versicherungswirtschaft. http://www.gdv.de/wp-content/uploads/2013/03/GDV_Code-of-Conduct_Datenschutz_2012.pdf. Datum des Aufrufes des Dokumentes: 11.02.2015.

Gorr, D. (2018). Ein Versicherungsroboter für gewisse Stunden. *Versicherungswirtschaft Heute*. http://versicherungswirtschaft-heute.de/schlaglicht/ein-versicherungsroboter-fur-gewisse-stunden/.

Guzmn, I. and Pathania, A. (2016). Chatbots in customer service. Technical report, Accenture.

Harkous, H., Fawaz, K., Shin, K. G., and Aberer, K. (2016). Pribots: Conversational privacy with chatbots. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, Denver, CO. USENIX Association.

Horch, A., Kintz, M., Koetter, F., Renner, T., Weidmann, M., and Ziegler, C. (2012). *Projekt openXchange: Servicenetzwerk zur effizienten Abwicklung und Optimierung von Regulierungsprozessen bei Sachschäden*. Fraunhofer Verlag, Stuttgart.

Inc, S. (2018). Most popular messaging apps 2018. https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/.

Kamarinou, D., Millard, C., and Singh, J. (2016). Machine learning with personal data. *Queen Mary School of Law Legal Studies Research Paper*, (247).

Kirakowski, J., O'Donnell, P., and Yiu, A. (2009). Establishing the hallmarks of a convincing chatbot-human dialogue. In Maurtua, I., editor, *Human-Computer Interaction*, chapter 09. InTech, Rijeka.

Koetter, F., Weisbecker, A., and Renner, T. (2012). Business process optimization in cross-company service networks: architecture and maturity model. In *SRII Global Conference (SRII), 2012 Annual*, pages 715–724. IEEE.

Kowatsch, T., Nißen, M., Shih, C.-H. I., Rüegger, D., Volland, D., Filler, A., Künzler, F., Barata, F., Hung, S., Büchter, D., et al. (2017). Text-based healthcare chatbots supporting patient and health professional teams: Preliminary results of a randomized controlled trial on childhood obesity. In *Persuasive Embodied Agents for Behavior Change (PEACH2017)*. ETH Zurich.

McTear, M., Callejas, Z., and Griol, D. (2016). *The Conversational Interface*, volume 6. Springer.

Newlands, M. (2017). 10 ways ai and chatbots reduce business risks. https://www.entrepreneur.com/article/305073.

Nguyen, A. and Wobcke, W. (2005). An agent-based approach to dialogue management in personal assistants. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 137–144. ACM.

Niddam, M., Barsley, N., Gard, J.-C., and Cotroneo, U. (2014). Evolution and revolution: How insurers stay relevant in a digital future. https://www.bcg.com/publications/2014/insurance-technology-strategy-evolution-revolution-how-insurers-stay-relevant-digital-world.aspx.

Nordman, E., DeFrain, K., Hall, S. N., Karapiperis, D., and Obersteadt, A. (2017). How artificial intelligence is changing the insurance industry.

Pohl, V., Kasper, H., Kochanowski, M., and Renner, T. (2017). Zukunftsstudie 2027 #ichinzehnjahren (in german). http://s.fhg.de/zukunft2027.

Power, J. D. (2017). 2017 u.s. auto claims satisfaction study. http://www.jdpower.com/resource/jd-power-us-auto-claims-satisfaction-study.

Radziwill, N. and Benton, M. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*.

Renner, T. and Kochanowski, M. (2018). Innovationsnetzwerk digitalisierung für versicherungen (in german). http://s.fhg.de/innonetz.

Rudnicky, A. and Xu, W. (1999). An agenda-based dialog management architecture for spoken language systems. In *IEEE Automatic Speech Recognition and Understanding Workshop*, volume 13.

Schippers, B. (2016). App fatigue. https://techcrunch.com/2016/02/03/app-fatigue/.

Schwark, P. and Theis, A. (2014). *Statistisches Taschenbuch der Versicherungswirtschaft*. Gesamtverband der Deutschen Versicherungswirtschaft (GDV), Berlin.

Smart-Data-Begleitforschung (2018). Smart Data - Smart Solutions. https://www.digitale-technologien.de/DT/Redaktion/DE/Downloads/-Publikation/2018_06_smartdata_smart_solutions.pdf.

Sörensen, I. (2017). Expectations on chatbots among novice users during the onboarding process. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-202710.

Spierling, U. (2005). Interactive digital storytelling: towards a hybrid conceptual approach. *Worlds in Play: International Perspectives on Digital Games Research*.

Stange, A. and Reich, N. (2015). Die zukunft der deutschen assekuranz: chancenreich und doch ungewiss. In *Change Management in Versicherungsunternehmen*, pages 3–9. Springer Gabler, Wiesbaden.

Traum, D. R. and Larsson, S. (2003). The information state approach to dialogue management. In *Current and new directions in discourse and dialogue*, pages 325–353. Springer.

Weindelt, B. (2016). Digital transformation of industries. Technical report, World Economic Forum and Accenture.

Weizenbaum, J. (1966). Eliza - a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45.

Yacoubi, A. and Sabouret, N. (2018). Teatime: A formal model of action tendencies in conversational agents. In *ICAART (2)*, pages 143–153.

Zimmermann, G. and Richter, S.-L. (2015). Gründe für die veränderungsaversion deutscher versicherungsunternehmen. In *Change Management in Versicherungsunternehmen*, pages 11–35. Springer.
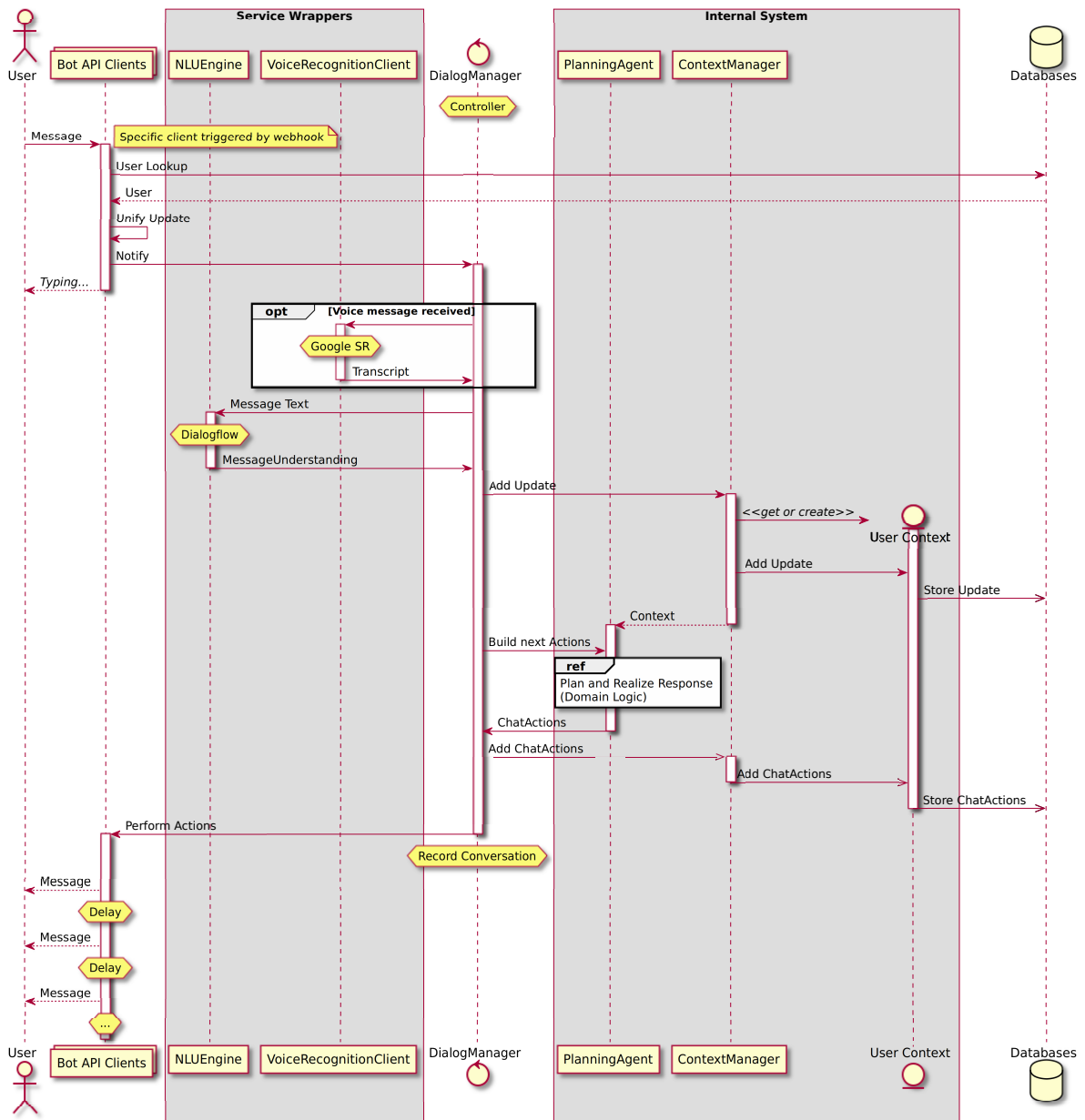
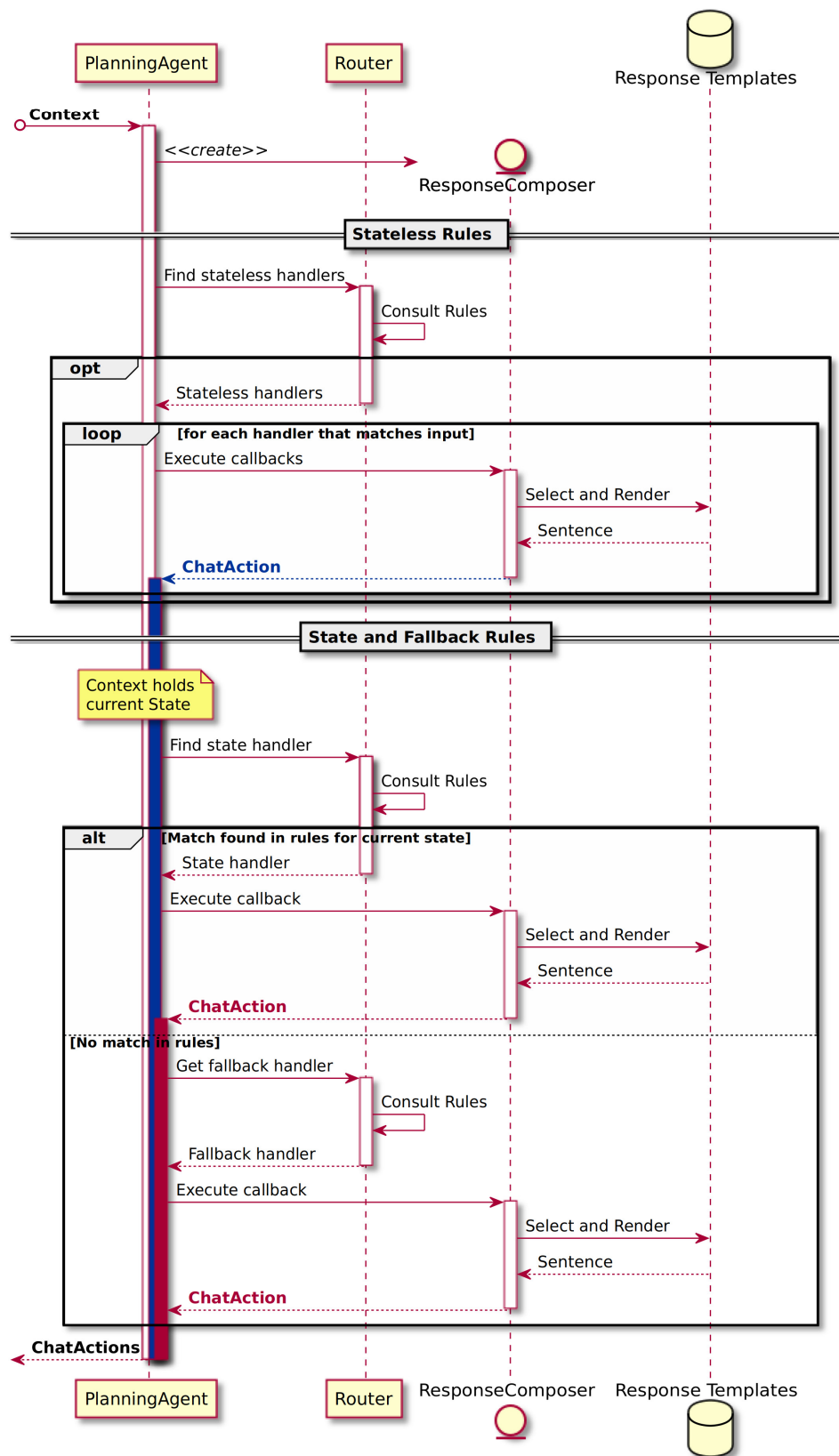Figure 5: Sequence diagram of the conversational agent prototype

Figure 6: Detailed sequence diagram of the response generation in the conversational agent prototype