Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

An Open Source Chatbot for a Complex Domain

by

Philip Corcoran

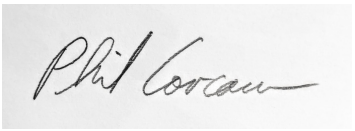
Under supervision of Prof. Vincent Wade

October 5, 2023

A dissertation presented in fulfilment of the requirements for the Degree of Master in Computer Science

Declaration Of Authorship

I, Philip Corcoran , declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

Signed: 

Date: October 5, 2023

Acknowledgements

I would like to sincerely thank Prof. Vincent Wade for his immense support throughout the two years. He maintained a steady heartbeat of regular meetings where his feedback and insights were invaluable. His contributions and recommendations kept the project on track and they were delivered with good humour at all times.

I would also like to thank my wife, Sunita Corcoran, for her support, encouragement and patience.

Finally, thanks to Cobus Greyling for his extensive articles on intents, entities and dialogue management in dialogue systems. I discovered his writing while struggling with the intricacies of implementing those concepts and took much learning and encouragement from him.

Abstract

The call centre industry has grown rapidly due to advancements in information and communication technology that have enabled automation of many customer service tasks. The computerisation of customer relations has become a primary objective for many businesses, with the goals of generating productivity gains and establishing strong customer relationships. However, the high-pressure work environment of call centres often leads to high levels of stress for agents and high staff turnover rates. AI technology is bringing significant changes to call centres with intelligent call routing and analysis of customer behaviour allowing staff to handle calls more efficiently thereby reducing customer wait times. Chatbots are being deployed in call centres in order to improve the customer experience by handling routine queries efficiently and reducing the service agent workload.

This research explores the chatbot built using open source technology to address customer queries for a complex domain, namely health insurance. This domain has technical terminology and products with detailed specifications. A knowledge base is designed and built to extend the existing domain taxonomy to cater for language and concepts used by ordinary users. A dedicated quiz application is developed to get real time feedback from users as they interact with the chatbot.

The experimental results and evaluation presented in this work corroborate the idea that a chatbot can be applied to a complex domain. An initial survey indicates the complexity of the domain both in terms of the vocabulary and also the products. A chatbot built using the Rasa open source framework is able to present complex data from the domain in a manner that users can understand. An iterative process is used to modify the chatbot using open questions posed to participants and analysis of the language used by participants, resulting in improved understanding of user requests and better usability.

Contents

Tables

Figures

Chapter 1 : Introduction

This chapter introduces the research by stating the problem, the research objectives and the methodology. It gives a brief overview of each chapter in the thesis.

1.1 Problem Statement

The call centre industry has grown significantly becoming one of the fastest-growing areas of employment worldwide. The growth of the industry has been enabled by the advancement of information and communication technology, which has led to the automation of many aspects of customer service [1]. The computerisation of customer relations has become a primary objective for many businesses. There are two main goals of this computerisation. Firstly, to use economies of scale to generate productivity gains in taking customer orders and managing after-sales. Secondly, to create value by establishing strong relationships with customers. By regularly contacting and informing customers, companies can create a sense of loyalty, which can lead to repeat business and customer retention [2].

However, the rapid changes in the call centre work environment, combined with the high-pressure manner in which many customer service operations are managed, often lead to high levels of stress for call centre agents. This level of stress, combined with other factors such as lack of career advancement can lead to high staff turnover rates. The cost of replacing a call centre agent can be significant. Call centre providers understand the direct and indirect costs associated with replacing an experienced employee with an inexperienced new hire [3].

A lack of expertise among call centre agents, resulting from vacancies and inexperienced agents, can compound the problem. Work must be completed by other call centre agents working overtime, which can result in further burnout and increased staff turnover rates. To address these issues, call centre management must provide clear career paths for agents and make full use of their skills. The qualification level of call centre agents is on the whole much higher than is often assumed. The majority of agents have a vocational training qualification, and some have university degrees [1]. As AI brings more automation to the call centre these agents are qualified to advance to other roles.

AI automation can assist agents in tasks such as accessing knowledge in reference materials, policies and procedures, which can otherwise be challenging to obtain. Agents in many industries must sift through lengthy documents to find the information they need. As a result, many turn to calling other departments for even basic questions, leading to overloaded call centres and long wait times for assistance [4]. The development of AI technology is bringing significant changes to the industry. AI is first replacing some of the tasks allocated to a service role. This transition stage is seen as augmentation.

The progression of AI task replacement from lower to higher intelligences results in changes over time in the relative importance of the intelligences for service employees.This stage is

already happening in industry, with leading vendors such as Kore.ai offering AgentAssist functions to augment agent response efficiency and reduce Average Handling Time (AHT) [5]. This means that call centre staff can handle calls more efficiently, allowing more time for

expressions of empathy with the end user. This efficiency can also reduce customer wait times and bring cost benefits to the centre. It allows more time for agents to engage with higher value tasks, opening the door towards their career advancement.

Given the benefits of conversational AI for call centres in general, the motivation for this research is to apply this to a complex domain, namely health insurance. Corpora containing the logs of customer queries from call centres exist for domains such as restaurant or flight booking but other industries do not yet have sufficient, publicly available, training data for conversational AI systems. This research will determine whether a chatbot implementation can be effective without the benefit of a large corpus of training data. It will investigate whether open source software can be leveraged to design a dialogue system which can handle domain specific terminology of a technical and medical nature and handle queries from potential customers, giving responses in language which they understand.

1.2 Research Objectives

This thesis examines the use of open source software for conversational agents. It investigates how well a dialogue system, built with open source, can interpret and return appropriate responses for customer care queries regarding products with technical domain specific language.

The objectives of the thesis are as follows;

- establish that the domain in question , health insurance, is complex in the sense that it has technical domain specific language that is not understood by potential customers and to indicate where in the domain confusion would arise for customers and how severe this problem is.
- design and implement a chatbot to address the domain language and product data complexities using open source technology.
- evaluate the chatbot design and implementation with participants who are potential purchasers of the products. In particular to answer the following questions;
    - Is the chatbot implementation able to handle the complexity of the domain in terms of product data and domain language?
    - Do the participants understand the chatbot responses?
    - Are the participants satisfied with the usability of the chatbot?

1.3 Methodology

The methodology adopted to address the objectives in the thesis is a design based approach [6]. The domain is surveyed initially to assess where the complexity lies. The survey then informs the definition of general challenges for such an application domain and the development of a case study in implementing conversational AI for that domain. The

implementation uses open source software, predominantly, for development. A prototype system is designed and implemented to address the challenges identified for the domain. A set of evaluation criteria is designed to appraise the success of the system as used by potential customers of the domain. The evaluation of the design & prototype employs both qualitative and quantitative methods Multiple diverse evaluation criteria are defined in order to enable triangulation of results so that a full appraisal of the system can be produced. An iterative approach is used to design, develop and appraise the system so that the evaluation may drive subsequent changes which may then be assessed using the same evaluation criteria.

1.4 Scope

The scope of the research is to study an open source chatbot implementation in a complex domain, health insurance. There are a number of different aspects to this complexity. There is difficulty in the semantics and terminology of the domain, for example health insurance benefits can have elaborate names such as “Multi occupancy or semi private room in a public hospital and day case”. There is difficulty in comparing health policies(plans) in this domain as constituent parts are not comparable in terms of the items or aspects included. An example is the policy benefit for advanced high tech hospitals where one insurer says the policy covers the “*Beacon Hospital with once-off €300 excess per family per year. Reduced cardiac list covered in Mater Private and Blackrock Clinic”* while another insurer just states there is *“90% cover”* on their plan. The users of the health insurance domain are everyday people, not medical insurance domain experts, so this makes the domain more complex in their eyes. Chatbots can be used to respond to specific questions from the users and to express the answers in simpler language which is more understandable to the users.

The research focuses on two representative health insurance plans and excludes all others. The chosen plans are documented in Appendix Comparison of Two Popular Health Insurance Plans from the HIA. These plans contain a broad representation of typical health insurance benefits - there are over 60 benefits included in each plan. The Rasa chatbot framework was chosen for the implementation and other frameworks were excluded as described in 2.9 Rasa open source software. Study participants were recruited from a single research platform, Prolific. Voice processing was implemented in the beta version of the chatbot to investigate multi-modal channels but it was not part of the experimental studies. The voice channel is detailed in the Appendix Voice Channel Implementation.

1.5 Research Contribution

This research contributes to the field by identifying approaches to chatbot development where progress can be made in domains where there is no publicly available corpus of customer queries. The research proposes methods of open questioning with potential customers to expose their natural language expressions for domain concepts and products and to use this language to improve the natural language understanding (NLU) and knowledge base of a system to respond to their queries on the domain. An application is developed to obtain immediate feedback from potential customers regarding products on an attribute by attribute basis to identify where cognitive load is highest for customers when

assessing a product. A technique is developed to extract entities which have large variance in word count and little semantic similarity from the customer utterances to a chatbot. A method is developed in open source software to initiate call recording on a voice communications platform to record audio dialogue with a chatbot. The research contributions are explained in detail in the subsequent chapters of the thesis and the next section provides a summary outline of each chapter.

1.6 Dissertation Outline

The following is an outline of the chapters in the dissertation.

Chapter 2 outlines the current state of the art in dialogue systems for conversational AI. It distinguishes task oriented systems which are used to assist customers in achieving a goal, for example booking a flight, from non task oriented, generative systems. It highlights the contribution of open source software to the development of the field, in particular the development of transformers, foundational models and chatbot frameworks. It describes recent developments where the line between task oriented and generative systems is blurring and where additional systems such as recommenders are being combined with dialogue systems.

Chapter 3 details the high level requirements for the design of the case study to meet the research objectives. It describes the design of the case study, including the design of a survey and a chatbot to represent the domain to potential customers and handle their queries. This chapter presents the architecture of the chatbot, including a sample conversation to illustrate components of the architecture. It describes the design of questions to be asked of participants in the chatbot experiment and the design of the evaluation methods.

Chapter 4 describes the implementation of the chatbot including the web based front end and the server side components. It describes the natural language understanding (NLU) pipeline and how it is trained with customer intents and domain entities. It documents the dialogue state tracking and dialogue policy and how an open source framework is leveraged with additional custom python code. It shows the implementation of a custom quiz application to ask questions of the participants as they interacted with the chatbot to gain immediate feedback from them. It details the voice implementation which leverages a commercial voice communications platform (Twilio). It explains the deployment on the Google Compute Platform (GCP) to make the system publicly available for external testing.

Chapter 5 contains the objectives, results and setup of both the initial survey and the first experimental study with the chatbot, Study 1. It details the initial survey questions and describes how the results indicate that both the domain language and the product data caused confusion among participants. Nearly three quarters of participants in the survey are confused by the domain in question. Results indicate that it would be challenging to implement a chatbot for the domain. The chapter provides the results of Study 1 where participants compared two products from the domain while interacting with the chatbot implementation. The results of the open questions concerning the product benefits are documented. Feedback from the participants is detailed with their usability scores for the

standard System Usability Scale (SUS). Requests which were misunderstood by the chatbot are detailed along with other errors occurring over one or more dialogue turns. A discussion of the results is provided.

Chapter 6 details the objectives, results and setup of the second chatbot experiment, Study 2. An iterative development process is used so that new participant terms can be readily incorporated to extend the chatbot's range of entities. The results of Study 1 are leveraged to modify the chatbot and to train it on a larger data set. Navigating categories of product benefits can be difficult and this navigation is further developed for the second execution of the chatbot experiment. The knowledge base of the chatbot is developed further after Study 1 to attune it to the language of participants resulting in fewer misunderstood requests by the chatbot. Open questions to participants are used to generate new terms and their utterances are mined for new ways of expressing intents. Product comparison questions are used to highlight deficiencies in product specifications. The second execution of the chatbot uses an improved implementation where GUI affordances such as buttons are used for better presentation and navigation. Usability scores are measured using the standard System Usability Scale (SUS). The results of Study 2 are documented and compared with those of the previous iteration, Study 1.

Chapter 7 contains the conclusions reached from the research. The objectives set out in the introduction were achieved by the project. The discussion of results is summarised and the research achievements are highlighted. Future work along a number of avenues is described.

Chapter 2 : State of the Art

This chapter describes the state of the art in dialogue systems for conversational AI. Conversational AI is the technology enabling machines to understand, process and respond to human language. Systems built on this technology are commonly called dialogue systems. The application of dialogue systems in industry is discussed. The systems are broadly classified into task-oriented systems which assist the user in completing a goal, for example booking a restaurant, and non task oriented systems which excel at human-like conversation. The cognitive load on users of these systems is considered. The systems generally require significant amounts of training data and there are publicly available corpora of domain specific conversations which can be leveraged. Advances in Automatic Speech Recognition (ASR) are explored for their potential use in generating new corpora. Recently, other systems are being combined with dialogue systems, including recommender systems which can offer personalised product recommendations to users. The contributions of open source software to dialogue systems is explored and the open source chatbot framework used in this research (Rasa) is introduced. The following section discusses the use of AI in call centres.

2.1 Conversational AI in Call Centres

The field of conversational AI has seen rapid advances in the past 15 years and these advances have been most readily applied to call centres. Gartner predicts that by 2026, conversational artificial intelligence (AI) deployments within contact centres will reduce agent labour costs by $80 billion.[7] Chatbots are a significant part of these deployments. A chatbot is a dialogue system which can automatically interact with a user and connect it to business processes such as sales and customer support. It is a new form of customer assistance powered by artificial intelligence via a chat interface. Since it can remove human factors the chatbot is capable of providing 24 hour service which can relieve pressure on overburdened customer service centres. AI chatbot application systems can automate entire business processes. Chatbots use neural networks to extract meaning from the user input. They then use a variety of methods, depending on design, to produce an appropriate response. [8]

The use of chatbots in customer service has evolved beyond simply answering FAQ-type questions or routing calls to agents. Real customer problems are being solved, and the most influential factors when using chatbots for customer service are response relevance and problem resolution. Addressing these factors results in positive customer satisfaction, increased probability for continued chatbot usage, more purchases, and more product recommendations [8]. Chatbots are becoming more sophisticated and can interact with customers in a way that closely resembles human interactions. This can improve the customer experience.

Chatbots or dialogue systems can generally be divided into task oriented and non task oriented systems [9]. Industry deployments consist primarily of task oriented systems and they will be discussed initially.

2.2 Task Oriented Systems

Task oriented systems aim to assist the user in completing a given task such as booking a flight or comparing products for purchasing. They have a common approach to accept a user utterance as input and use a pipeline to produce the best response. Natural language understanding (NLU) forms the first part of this pipeline. Statistical methods and neural networks are used to process the text of the utterance. The text is featurized and the user's intent is detected. Labels are assigned to words in the utterance to indicate predefined entity types which can be used to fill slots in the system's memory. Figure 2.1 illustrates the classification of an intent and the labelling of the entities of a user utterance. The utterance is "show restaurant at New York tomorrow" [9].

Sentence	show	restaurant	at	New	York	tomorrow
Slots	O	O	O	B-desti	I-desti	B-date
Intent	Find Restaurant					
Domain	Order					

Figure 2.1 Intent Classification

The output of the NLU passes on to the next stage of the pipeline - the dialogue manager (DM). Figure 2.2 shows the components of a dialog manager.

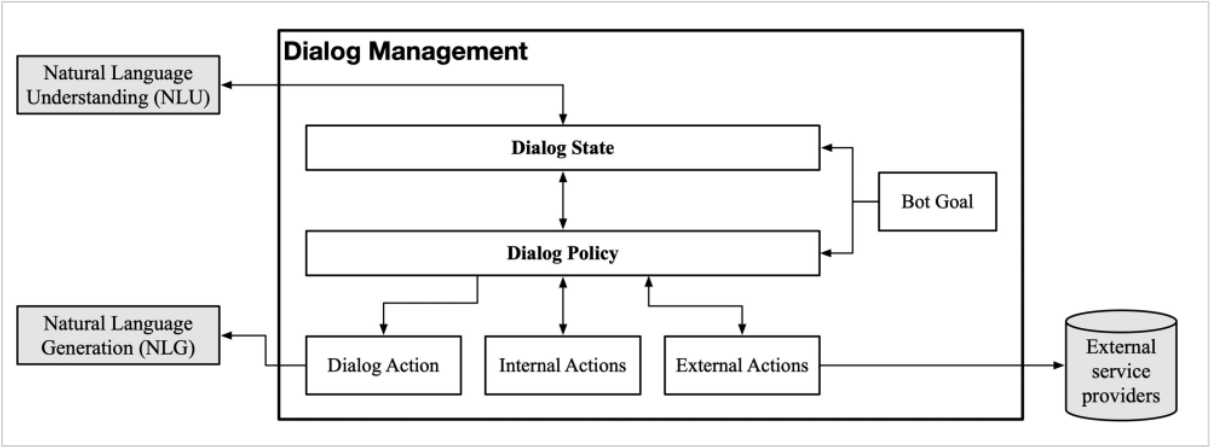


Figure 2.2 Dialog Management

The DM handles dialogue context and decides the next action for the agent to take [10]. The DM can be subdivided into a dialogue state tracker (DST) and a dialogue policy component. The dialogue state represents the dialogue session at a given time. Classically the dialogue state is considered as semantic frames and this aspect has remained in commercial systems since it was introduced by the 1977 Genial Understander System(GUS) [11].Other statistical approaches model the inherent uncertainty in dialogue systems using Partially Observable Markov Decision Processes (POMDPs) [12].

2.2.1 NLU Intents and Entities

Given an utterance, natural language understanding classifies it as a user intent and maps it into semantic slots. The slots are pre-defined according to different scenarios [9].

Slot filling is a challenging problem for language understanding. Unlike intent detection, slot filling has usually been defined as a sequence labelling problem, where words in the sentence are assigned with semantic labels [9]. Some systems support custom validation after entity extraction so associated slots may also be filled. Additionally, entities may be looked up in external data sources or rejected as part of validation [13].

Modern NLU implementations may perform entity recognition and intent classification with the same component. The DIET (Dual Intent and Entity Transformer) component of the Rasa framework has a multi-task architecture for intent classification and entity recognition. The architecture is based on a transformer which is shared for both tasks. A sequence of entity labels is predicted through a Conditional Random Field (CRF) tagging layer on top of the transformer output sequence corresponding to the input sequence of tokens. Both the transformer output for the complete utterance and the intent labels are embedded into a single semantic vector space. The dot-product loss is used to maximise the similarity with the target label and minimise similarities with negative samples [14]. Once intents have been detected and entities extracted the pipeline processing continues.

2.2.2 Dialogue State Management

The intents and entities are used as inputs for the next stage of the pipeline, the dialogue manager (DM). Additional inputs may include the dialogue history and external knowledge bases. The DM has two parts; dialogue state tracking and policy.

Dialogue state represents the dialogue session at a given time. One of the first notable dialogue systems, Eliza, used a finite state transducer [15]. States were well defined and there were fixed transitions between them which occurred under predetermined conditions. Since the arrival of the Genial Understander System(GUS) most task oriented systems have been frame based.

In the context of dialogue systems, a semantic frame refers to a structured representation of the meaning of a user's utterance. It captures the underlying semantic structure and intent of the user's message. It maps the message to a set of concepts, entities, and relations that are relevant to the specific domain and context of the dialogue. The way in which the history of previous utterances is incorporated into the semantic frame will depend on the design and implementation of the dialogue system. Some systems may only consider the most recent user input, while others may maintain a longer history of previous utterances. They may use machine learning or other techniques to model the user's intent and preferences over time.
A frame encapsulates both the context for utterance interpretation and also the context for dialogue progress. It crosses the boundary between dialogue state tracking and policy.

Frame based systems are more flexible than finite state since a data-model is added to the dialogue tree. This means that memory slots can be filled in any sequence or in iterations. In other words, the state transitions are not predetermined. When there is a requirement for a minimal set of data in order for the next action to occur then the system can prompt for outstanding data [15].

Once the dialogue state has been established for the current turn, then the next stage of the pipeline is invoked. The goal of this policy stage is to predict the next action of the system, typically to generate a response to be sent to the user.

Eliza used a simple set of rules for policy and the rules were ranked. Rules evolved into frame based architectures with GUS in which multiple different paths are supported [15]. Modern commercial systems are hybrids incorporating both frames and rules. They often declare rules for situations which should always have the same response e.g. “When the user says goodbye then the bot should always say goodbye”. Systems can use rules to govern the behaviour around forms [16]. Rules become difficult to maintain as the surface area of the dialogue system grows.

Frames allow for more flexibility. Slots can be filled in any sequence. This is the basis of most commercial offerings such as Google Dialog Flow or Amazon Lex [10][17].

In 2017, Google researchers proposed a sequence transduction model based on attention mechanisms instead of the then popular CNN and RNN models [18]. The best performing CNN and RNN models had connected the encoder and decoder through an attention mechanism. The Transformer removed the recurrence and convolutions completely and was based solely on attention.

Recently transformers have been applied to dialogue policy [19]. A self-attention mechanism operates over the sequence of dialogue turns. A single conversation can consist of multiple overlapping discourse segments as speakers interleave multiple topics. A transformer picks which turns to include in its encoding of the current dialogue state, and is naturally suited to selectively ignoring or attending to dialogue history.

Transformers are also being used for the optimization of policy. Learning through reinforcement requires a significant amount of training interactions, making learning with real users time-consuming. To address this issue, some setups incorporate user simulators instead of humans. However, user simulators have their own challenges. Hand-coded, rule-based simulators are effective in small and simple domains. However, these quickly become unwieldy for complex domains because of the large number of rules required. On the other hand, data-driven simulators are currently state-of-the-art but are domain-dependent. They must be retrained and redesigned for each new domain. Researchers have proposed a domain-independent transformer-based user simulator (TUS). The structure of TUS is not tied to a specific domain, enabling domain generalisation and learning of cross-domain user behaviour from data. TUS can compete with rule-based user simulators on pre-defined domains and is able to generalise to unseen domains in a zero-shot fashion [20].

Once the dialogue policy stage is complete, the next action has been chosen. This usually includes generation of a natural language response to the user.

2.2.3 Natural Language Generation

Dialogue systems generate natural language to respond to their users. The various response models can be categorised into four groups: template-based, generative, retrieval-based, and search engines [21]. Many commercial task oriented systems used template-based responses. They infuse static text with variable data. For example, “Your {Premium Plus} account has a balance of {$464,00} and is payable on {4 April}”. This is analogous to the way data is presented to a caller by a telephone system Interactive Voice Response(IVR). Custom actions can be used to call APIs or lookup external databases to complete the response. Frameworks such as Rasa support multiple templates per utterance to enable variety in the responses to make them appear more human-like.

Advanced natural language generation is even more anthropomorphic and fully generates natural language on the fly. Sequence to sequence models like chatGPT produce very human-like responses [22].

2.2.4 Dialogue Breakdown

Users want chatbots to provide adequate responses to simple enquiries. Studies show that even when chatbots fail to supply reasonable answers, users can still have a good experience if the chatbot provides an easy handover to human customer service representatives. Many end users have realistic expectations of chatbots' capabilities. Anthropomorphic qualities of chatbots may play a role in user experience, however the ability of the chatbot to efficiently and adequately handle customer service enquiries is far more important [23].

Current dialogue systems cannot handle all of the intricacies of conversational interactions and breakdowns happen often. If a chatbot fails to provide an answer, users frequently abandon the conversation [24]. Unresolved errors are sufficient to reduce adoption intention i.e. the user’s readiness to continue interacting with a chatbot. However, there is no difference to the end user between a chatbot which seeks clarification and one which never makes a mistake. The ability to resolve miscommunication (clarification) appears as effective as avoiding errors in the first place [25].
The causes of breakdowns can be classified as follows;

- Elaborated
    - Messages are quite long and consist of more than one sentence. The chatbot’s intent detection algorithm fails. It cannot determine what exactly the user wants.
- Specific
    - Here the user makes a very specific query which is not in the scope of the chatbot training or programming.
- Brief
    - These messages have little semantic weight but contain words that are familiar to the chatbot (e.g., greetings, incomplete sentences, insults). The messages can be as short as a single word and therefore they lack context.
- Cryptic

    - These utterances contain a high number of unknown words for the recognition lexicon. The inputs can be short with serious spelling problems and typos or they can be in an unsupported language [26].

Since the cost of breakdown may be abandonment then it is important to somehow “repair” the conversation. Repair strategies for chatbots have been developed from communication theory [27].

A strategy can be as simple as ignoring low confidence predictions and simply following the highest predicted intent. Alternatively the uncertainty can be acknowledged by repeating the last prompt or by asking the user to confirm their intent as in “Sounds like you want to order pizza. Is that correct?” Options can also be presented to empower the user with some choice, “I’m not sure, Did you want pepperoni or pastrami?”

Design recommendations for chatbots have evolved from communication theory as follows;

- Acknowledge mistakes with honesty and less redundancy. Acknowledge breakdowns.
- Explain models naturally and aesthetically.
- Intelligently repair conversations with user control. Provide resources to assist users to repair and proactively suggest solutions.
- Adapt to individuals and contexts since repair preferences are not universal [27].

2.3 Non Task Oriented Systems

Non task oriented systems excel at human-like conversation. They use sequence to sequence models to process the user utterance and generate a response. The model typically works by taking in the user input and converting it into a numerical representation. Then an encoder-decoder architecture is used to generate a response sequence one word at a time. The output sequence is then post-processed to produce the final response [28]. The first sequence to sequence natural language generators used Long Short-Term Memory (LSTM). More recently, transformers have set the benchmark for language tasks while also improving the computation performance [29].

2.3.1 Foundation Models

Transformers have led to the emergence of foundation models [22][30]. Since 2017, AI has undergone a paradigm shift with the rise of models such as BERT and GPT-3 which were trained on large quantities of unlabelled text using self supervision at scale. This process provides a reasonable parameter initialization for adaptation to a wide range of downstream tasks. The parameters are typically billions of weights or more. In the case of ChatGPT, reinforcement learning was also used. This has shifted the focus of natural language processing research away from the previous paradigm of training specialised supervised models for specific tasks [22].

Foundation model is a term used to describe a broad class of models and their function.

The term “Large Language Models” (LLMs) refers more specifically to language-related models with “large” emphasising their significant scale. Pretrained models are adapted to downstream tasks through fine-tuning. Foundation models are thus a superset of LLMs [22].

Foundation models have entered the mainstream media with the release of ChatGPT, an artificial intelligence chatbot developed by OpenAI and launched in November 2022 [31]. It is built on top of OpenAI's GPT-3 family of LLMs. The GPT architecture is a Generative Pre-trained Transformer. The pretraining process allows the model to learn general patterns and structures of language, which it can then use to generate responses to new input text. ChatGPT grew to 100 million active users within 2 months of its launch. This astonishing rise reveals both its utility in helping with a wide range of tasks and the public’s fascination with a machine which appears to be so human-like [32].

The rise in the capabilities of foundational models has led to them taking on more of the responsibilities of task oriented systems. This is discussed in the next section.

2.4 A Blurred Line between Dialogue Systems

The line between task and non task oriented dialogue systems is blurring. Large language models like GPT and BERT are used by the NLU pipelines of task oriented systems [33]. Generative models like chatGPT are being applied to more and more tasks. ChatGPT even outperforms fine-tuned models on some tasks [34]. However LLMs are not yet ready to replace commercial task oriented systems. In a recent study, developers for a health intervention system which leveraged an LLM were frustrated by some of its responses. They noted that even though the example dialogue corpus did not include cases making infeasible suggestions, the system still generated responses doing so [35].

LLMs are limited in so far as they do not incorporate external knowledge bases once they have been trained. They lack integration for transactional tasks such as ordering products and they suffer from hallucinations [35]. Commercial applications such as chatbots for the insurance industry must deliver factual information about their products to the end consumer. Research is evolving to reduce the effect of hallucinations by LLMs. One approach is to use retrieval augmented generation(RAG). Retrieval can be practically implemented using dense representations where embeddings are learned from a small number of questions and passages by a dual encoder framework. Dense retrieval can outperform and potentially replace the traditional sparse retrieval component in open-domain question answering [36]. RAG has been shown to work in conversational systems beyond simple Q&A. It has been proven to have open-domain conversational capabilities which substantially reduce the problem of knowledge hallucination in chatbots [37].

2.5 Cognitive Load on Chatbot Users

It is important to manage the cognitive load on users of dialogue systems so that they can complete their goals and be satisfied with their interactions. The cognitive load demanded by a given task depends on the proportion of time for which it captures attention. During this time it is impeding other attention-demanding processes. Attention is a sequential and time-based function of working memory [38]. Processing and storage of items in memory both rely on attention. Rearranging or evaluating items while they are being retained interferes with their storage. Cognitive tasks can be completed only with enough ability to hold information as it is processed. Working memory storage capacity is important. In 1956 Miller said there was a magic number of 7 items which could be stored in central memory. More recent research suggests a limit of 3 to 5 meaningful items [39]. This working memory varies depending on the task, therefore it is important to design the system to prevent overloading the user with complexity. When complexity is revealed gradually then the user has time to assimilate it [40]. The mental load on a subject depends upon their level of expertise in the task domain [41]. For the domain studied in this research, health insurance, the purchasers of the products generally have little medical or insurance expertise. Cognitive load is also increased when there is high element interactivity i.e. when the information presented to users is interrelated so that its multiple elements must be processed simultaneously [42]. This occurs in the health insurance domain where a single benefit such as cover for a private hospital room refers to multiple other facets such as excess, shortfall and waiting periods.

2.6 Industry Corpus

Chatbot intents should be ground truthed by a corpus of data. Many industries still do not have a recognised corpus of conversational data. The health insurance industry is an example. Research recommends that companies offering chatbots must thoroughly analyse the collected data to gain more insights into their customers' needs [24]. There is a need for businesses in these industries to access a relevant corpus of data to enable appropriate conversational AI system development. Setup challenges, including training data and maintenance, were among the top reasons for not implementing chatbots in enterprises, according to a recent Deloitte survey [43].

Attempts to transcribe real world audio conversations have failed in the past because Word Error Rate (WER) was not of sufficient quality. The metrics published by vendors of Automatic Speech Recognition (ASR) systems were achieved in laboratory settings and ignored the following interfering factors;

1. Different kinds of noise
2. Crosstalk among conversation participants
3. Accents
4. Rare words
5. Normalisation of transcripts [44]

Leading ASR vendors now provide models optimised for phone calls and meetings with significant cross-talking. They also enable clients to customise their speech-to-text models for industries with significant jargon or for customers with particular accents [44].

Call centres record their incoming audio conversations. The availability of accurate, customizable transcription can turn this audio into high quality conversation transcripts. The transcription process can be done offline to minimise costs. Tools like HumanFirst can automatically create clusters of intents and extract entities. The conversations interface in HumanFirst allows for past dialogs to be interrogated and utterances labelled. This means that long-tail NLU training datasets can be readily created for Contact Centre AI (CCAI) use-cases. Businesses can effectively create their own industry specific or even business specific corpora [44][45].

2.7 Recommender Systems

Conversational user interfaces (CUIs) and recommender systems (RecSys) are beginning to be used in conjunction with each other. CUIs and RecSys have traditionally focused on optimising the means and ends of user interactions, respectively. CUI studies have explored ways to facilitate conversations between users and technology, while RecSys research has focused on presenting users with personalised recommendations. However, recent advancements in both fields have led to the emergence of conversational recommender systems (CRSs). These seek to optimise both the means and ends of user interactions [46].

By integrating conversational interfaces with recommender systems, CRSs offer a unique opportunity to improve the overall user experience. Not only can CRSs provide personalised recommendations to users, but they can also foster engaging conversational interactions. These ultimately lead to more successful recommendations. Moreover, by prioritising conversations as an end in themselves, CRSs can help sustain long-term relationships between users and technology.

Several factors must be considered when designing CRSs. It is important to strike a balance between the efficiency of the conversational interface and the quality of the recommendations being offered. Additionally, it is critical to consider how the conversational aspect of the system can be used to enhance user engagement and satisfaction. By carefully considering these and other factors, designers can create conversational recommender systems that truly optimise both the means and ends of user interactions.

Furthermore, the integration of CRSs with other AI technologies such as natural language processing and machine learning can lead to even more advanced systems capable of understanding user preferences and delivering highly targeted recommendations. As the field of conversational AI continues to evolve, it is likely that CRSs will play an increasingly important role in shaping the future of user interactions with technology [46].

In a complex domain where a product has more attributes than a customer can readily understand, a conversational recommender system may play a useful role in enabling the customer to choose between competing products. Amazon is already selling home and contents insurance. In future it may recommend more complex products like health insurance.

2.8 Open Source's Contribution to Dialogue Systems

The need and desire for open source software was called out in a 2007 research paper, ""The need for open source software in machine learning". Open source tools had advanced to the point where they were suitable for building large-scale systems, and the field of machine learning had developed powerful algorithms for various applications. However, the potential of these methods was not being realised because existing implementations were not openly shared. This led to software with low usability and weak interoperability. It was proposed in that 2007 research paper that challenges associated with publishing algorithmic implementations of M/L methods could be mitigated by creating a resource of peer-reviewed software with accompanying short articles [47].

The research project described in this thesis has benefited immensely from open source. It enabled the M/L training and models to be run on a 10 year old refurbished PC with only 6GB of RAM. This was due to the efficiency of linux, Rasa, sanic (python) framework, Tensorflow, python and other open source projects.

Open source can be used to create transformers which excel at modelling sequential data, such as natural language. Transformers can be written using the Keras deep learning API which sits on top of Tensorflow [48]. This latter tool takes its name from the mathematical object describing physical properties, a "tensor". Velocity is an example of a first rank tensor which has a one-dimensional array of numbers for magnitude and direction. Keras was released as open source in 2015 as a high level machine learning library.

Unlike recurrent neural networks (RNNs), transformers are parallelizable. This makes them efficient on hardware like GPUs and TPUs. Transformers replace recurrence with attention and therefore computations can happen simultaneously. Layer outputs can be computed in parallel, instead of in series like an RNN [48].

TensorFlow allows for the expression and execution of machine learning algorithms. It can be used on a wide range of systems, from mobile devices to large-scale distributed systems. TensorFlow is adaptable and can be used to express various algorithms, such as deep neural network models for both training and inference [49].

Tensorflow was open sourced in 2016 and contributed to the ground breaking transformer research of 2017 [18]. Tensorflow was also a key enabler for the research and development of the NLU core of the Rasa framework [50]. Tensorflow and Keras led to the sensational release of chatGPT in 2022. Figure 2.3 shows a high level view of the timeline of open source contributions to conversational AI.

Call for open source	Keras	Tensorflow	Rasa NLU	Chatgpt
2007	2015	2016	2017	Nov 2022

Figure 2.3 Timeline for Open Source

The call for software to accompany research made in 2007 has now come to fruition. This has created enormous benefits for both the machine learning community and the scientific community at large. The core open source framework used for the research in this paper, Rasa, is described in the following section.

2.9 Rasa Open Source Software

Rasa Open Source is a conversational AI platform built with python. It supplies most of the building blocks for creating virtual assistants or chatbots [51]. It has python modules to connect to messaging channels and third party systems through a set of APIs.

Rasa was chosen for the research as the most complete and flexible framework for developing an end to end chatbot solution. The open source chatbot, Chatterbot was also investigated but it did not have the same capabilities for slot filling, custom actions or custom channels. Amazon Lex was also examined as a potential solution but it did not have the custom entity extraction and validation required for the complex domain of health insurance. Google Dialog Flow was considered but it did not have enough flexibility in the management of intents. Figure 2.4 is an architecture diagram demonstrating the key components of Rasa.

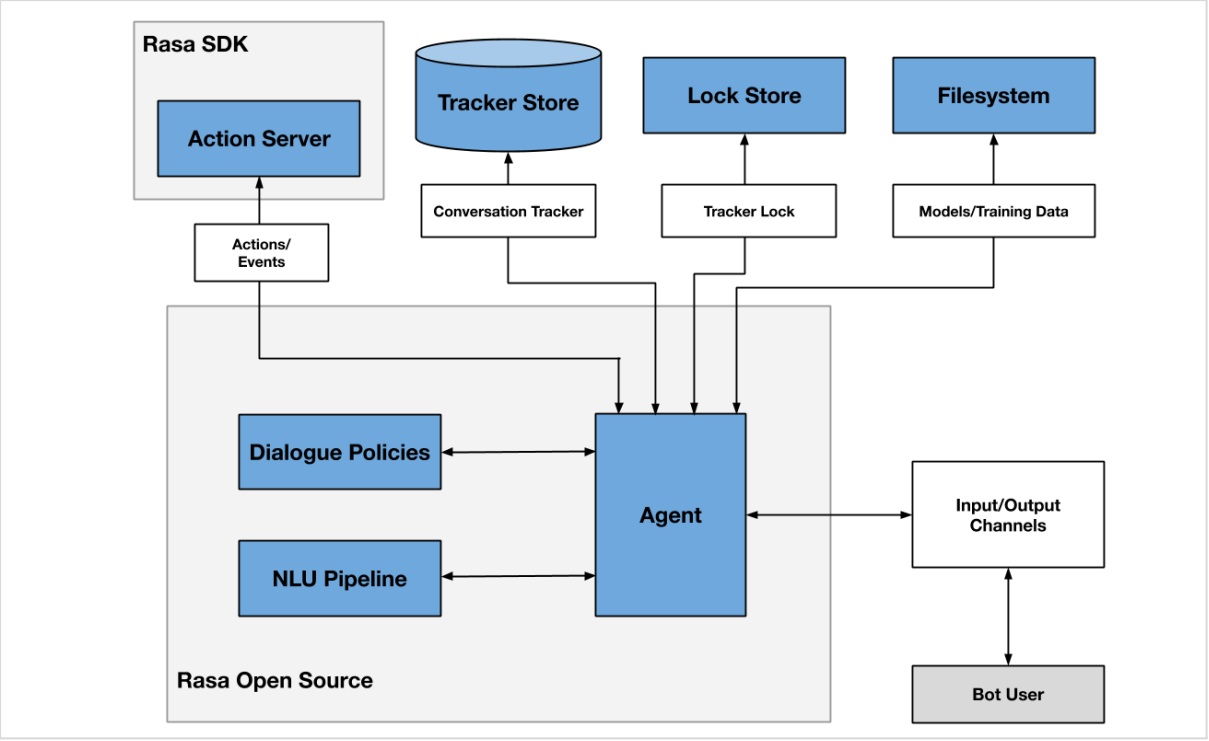


Figure 2.4 Rasa Chatbot Framework Architecture

Rasa has a pipeline architecture for natural language understanding (NLU). This pipeline handles intent classification, entity extraction, and also response generation. Responses are template based and can be augmented by custom actions. The NLU model is generated by the trained pipeline of configurable components which includes tokenizers, classifiers and

other language processing modules. In Rasa 3.0 the pipeline changed from being strictly sequential to be more of a graph [52].

The dialogue management component decides the next action in a conversation based on the context. This appears in the diagram as “Dialogue Policies”. Events from dialogue management are fed to the action server. Custom actions can generate complex responses including dynamic buttons, images and html.

Input / output channels link the framework with external systems using connectors. For this research project the socketio connector was used to connect a front end chat widget. The twilio connector was used to send and receive messages between Rasa and the twilio voice communications platform.

Training data for models is stored in yaml files on the filesystem alongside pipeline and connector configuration. The data includes user utterance examples and stories of sample conversations. The conversation tracker is a python object which maintains data per conversation including slot values and events. The tracker store records conversation events including slot filling and next action prediction. A sqlite database was used as the tracker store for this project.

Rasa components and custom code written by the researcher are discussed in more detail in the following chapters.

2.10 Evaluation of Chatbots

Chatbot developers in healthcare and related industries must consider measurements across a range of aspects in any given study including usability, response quality, and classifier performance [53].The success or failure of chatbots can be evaluated with a combination of qualitative and quantitative measures. Some metrics are subjective, for example user satisfaction and others are objective such as the count of requests which were misunderstood by the chatbot. The objective measures can be calculated from the logs of the conversations while surveys and questionnaires can be used to perform subjective assessments. PARADISE is a well known framework for calculating the performance of dialogue agents. It attempts to optimise a desired quality such as user satisfaction by formulating it as a linear combination of a variety of metrics, such as task success [54]. Task oriented chatbots are evaluated in terms of goal achievability and the ability to satisfy users.

Most developers of both AI-based and rule-based chatbots measure user satisfaction to determine their success [55]. Human-like chatbots lead to greater satisfaction and trust among customers of a business. A chatbot must provide system quality and information quality to ensure good customer experience and usability [56]. The System Usability Scale (SUS) is a standard method to measure this usability. It is a likert scale questionnaire to assess the ease of use of a given system. Empirical results from the analysis of a large number of SUS scores show that the SUS is a highly robust and versatile tool for usability professionals [57].The SUS test has been applied to the analysis of interactions between humans and chatbots in combination with other questionnaires [58]. When building any complex interactive system such as a chatbot an established approach is to design and

improve the system iteratively [59]. The same objective and subjective measures can be applied to each iteration of the system to assess progress.

2.11 Summary

The state of the art in dialogue systems for conversational AI has been described, including the application of dialogue systems in industry. Call centres in particular have benefited from these technologies. The systems are broadly classified into task-oriented systems which assist the user in completing a goal and non task oriented systems, proficient at human-like conversation. The emergence of Large Language Models (LLM's) has meant that generative systems are now capable of performing tasks also. The cognitive load on users of these systems must be carefully managed. The task-oriented systems require significant amounts of training data and there are corpora of domain specific data available. Automatic Speech Recognition (ASR) has advanced rapidly in recent years and shows potential for the generation of new, industry specific corpora. Recommender systems are now being combined with dialogue systems to produce conversational recommender systems (CRSs). The significant contribution of open source software to dialogue systems is described and the open source chatbot framework used in this research (Rasa) is introduced. Evaluation criteria for chatbots which use a combination of qualitative and quantitative measures are assessed.

Chapter 3 : Design

This chapter discusses the design of the research and how it addresses the research objectives. The primary research objective is to investigate the ability of open source to handle a chatbot for a domain with technical language and product data which is not well understood by potential customers. Can a chatbot be designed and implemented to handle this kind of complex language and data? The next objective is to evaluate the chatbot implementation with potential customers of the domain products. Can the participants understand the chatbot responses? And finally, to determine how usable the participants find the chatbot. The high level requirements of the design are derived from these objectives.

3.1 High Level Requirements for the Case Study

A case study approach is seen as the best means for exploring the domain and the chatbot. An initial survey is designed to determine that the domain creates sufficient cognitive difficulty. The survey informs the definition of general challenges for the domain and the development of a case study in implementing conversational AI for that domain. It is desired to select participants for the survey who have some knowledge of the domain so that they can more fully explore the products within the domain. Participants are recruited from the Prolific academic recruitment website where they can be appropriately prequalified.

Two products which accurately represent the domain are chosen - two of the most popular products on the market. The products represent the terminology in the industry as each product is chosen from a different provider. The providers in question are the top two companies in the Irish market. The product attribute values exhibit language of a medical nature and include insurance industry terminology.

A prototype system is designed and implemented to address the challenges identified for the domain and an experiment conducted to evaluate the chatbot prototype. Open questions are included in the chatbot experiment to encourage exploration of the chatbot and to investigate how it would handle unexpected dialogue paths. For example, participants will be asked an open question on which health insurance benefit is of the most importance to them. They will then be required to interact with the chatbot to explore which plan provides better cover for open benefit. The open questions will also be used to discover new participant language for the domain to be used to expand the chatbot intents and entities.

There is a statutory regulator in Ireland for the domain, the Health Insurance Authority (HIA), and the regulator has a stated vision to assist the public in making informed product decisions. The domain model for the chatbot will be derived from the taxonomy provided by the HIA. The open questions will enable comparison between the mental model of participants and the domain model presented in the HIA taxonomy.

It is required to have questions concerning product attributes with complex values across the two providers. This is to require nuanced cognitive decisions from the participants. There will be a cognitive load on participants to rearrange and evaluate sub items within product attributes and also items within categories of product attributes.

In order to fully examine the capabilities of open source software it is required to have a challenging implementation to see if open source can be used to develop a capable task oriented system as described in 2.2 Task Oriented Systems. It should be able to handle the following;

- A complete NLU pipeline with multiple intents and entities.
- Generation of varied chatbot responses with dynamic data.
- Development of a knowledge base to extend the HIA taxonomy in line with participants' language for expressing intents and benefits.
- Various dialogue paths which are more complex than predefined finite state transitions.

The data in the domain is required to be complex enough to necessitate an ontology with multiple entities and relationships.

Given the sensitive nature of the health insurance domain, ethics is given due consideration and an ethics application is prepared and approved by the research institution before each engagement with participants. No questions of a personal nature are asked of participants. In the event of any personal information being inadvertently disclosed by a participant that information is not used and is promptly deleted. The ethical clearance process ensures that the appropriate GDPR safeguards are always in place. Participants are advised of this before commencing the online questionnaire and they are given an option to withhold their consent and exit the process.

3.2 Methodology

A design based methodology is chosen as this is very suitable for systems with a significant user interface component. The approach to the design is to initially provide a survey to identify the complexity challenges for this domain. The survey then informs a series of experimental case studies. Real users of the health insurance domain will interact with a chatbot that answers questions concerning two actual health insurance policies (plans). A prototype chatbot is iteratively developed across two development iterations to evaluate the performance of the design. Both iterations are evaluated using the same criteria so that the evolution of the prototype can be assessed. There is particular focus on the design's ability to deal with complexity of semantics and terminology, with difficulty in comparison on health policies and finally with the ease of use and understanding by end users.

3.2.1 The Case Study, "Frank's Health Insurance"

A fictional health insurer, "Frank's Health Insurance" is created for the case study. "Frank's health insurance" will have two plans available, plan A and plan B. The plans are representations of two of the most popular health insurance plans as presented by the HIA. The data is real. It is taken from a Health Insurance Authority (HIA) website comparison as seen in the **APPENDIX: Comparison of two popular health insurance plans from the HIA.** These particular plans are chosen to represent the market because they are two of the most popular health insurance plans. .Both plans have complex language. The plans overlap on some attributes and differ on others making the comparison task more complex.

A complete chatbot is designed to interact with users via text and voice. The chatbot answers questions regarding benefits covered by both of Frank's plans. Participants will play the role of a user seeking to compare two health insurance products and selecting the better suited one. Comparison questions are posed to the participants to elicit their understanding of the domain and open questions elicit their natural language concerning health insurance as described in 3.5 Question Design and 3.6 Design of the Evaluation Framework

3.3 Chatbot Architecture

The chatbot architecture supports two modes, chat (text) and voice. A GUI is integrated to support chat and a virtual phone number will be set up with a voice communications provider. Figure 3.1 depicts the end to end chatbot architecture, indicating users interacting with the system by typing into the chat GUI or dialling a phone number which is connected via the communications platform(Twilio). The rest of the system is contained in a linux virtual machine which includes a web server, database store, the Rasa instance and connectors to Rasa. The Rasa instance architecture has been shown in more detail in Figure 2.4 Rasa Chatbot Framework Architecture . The core Rasa components and connectors as described in 2.9 Rasa Open Source Software are installed on the linux server.

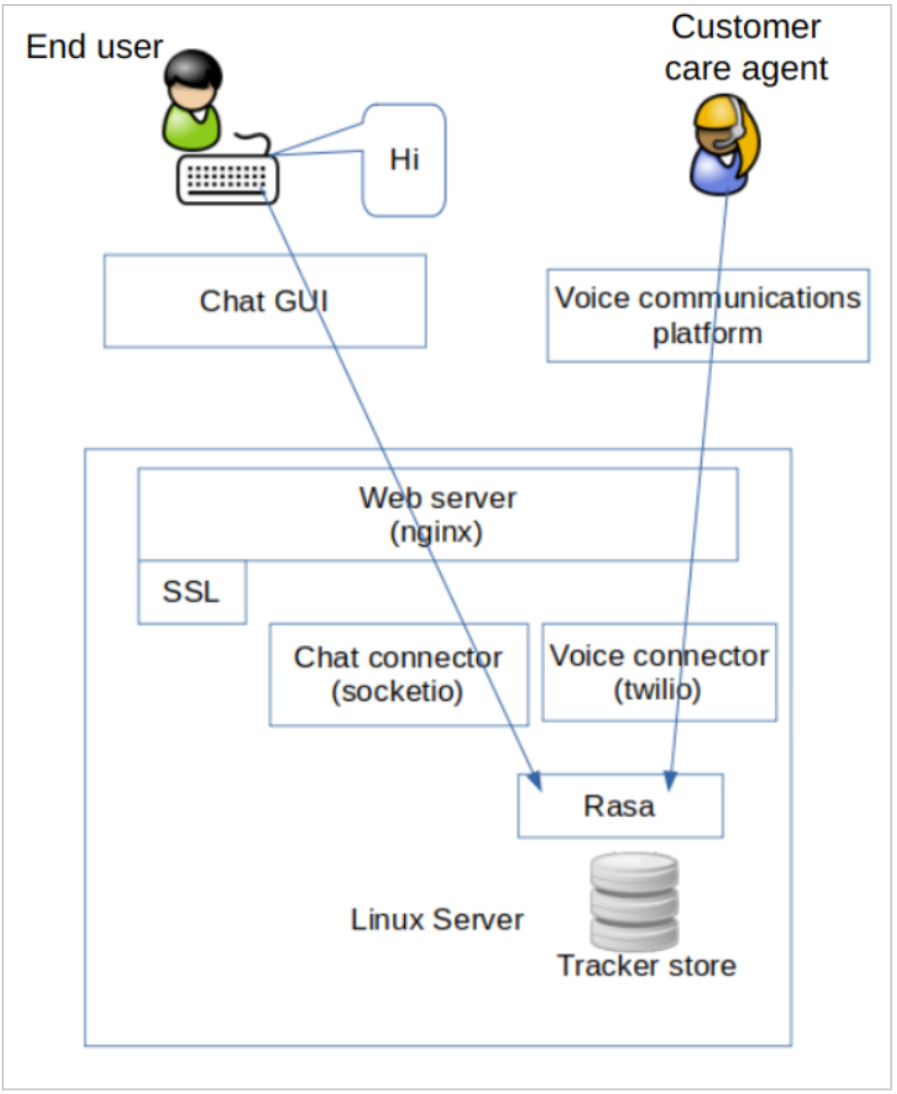


Figure 3.1 End to End Chatbot Architecture

The individual components of the architecture are discussed in more detail in **Chapter 4: Implementation.**

3.4 A Sample Conversation

Table 3.1 illustrates a sample conversation. The participant utterance is the first column, then the response from the chatbot is shown. An explanation is given of each interaction, including which components of the chatbot were involved.

Participant	Chatbot	Explanation
“Hi”		

Participant	Chatbot	Explanation
		NLU detects intent *greet and* rule based policy decides next action is to *utter_greet*
	“Welcome to Frank's Health Insurance. How can I help?”	
“What plans are available?”		
		NLU detects intent *state_interest_in_benefits*
	“I have two different health insurance plans, plan A and plan B. I can tell you about the benefits of these plans. You can say, for example, "tell me about cardiac" “	
		Memo based policy decides next action is *utter_ask_which_benefit*
“Tell me about cancer accommodation support”		
		NLU detects intent *state_pbi_selected*
	“Concerning “cancer accommodation support”, Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment”	
		TED policy decides next action is *action_state_plan_benefit_values*

Table 3.1 A Sample Conversation with a Chatbot

3.5 Question Design

For the initial survey, two questions are asked in an online form to establish confusion and to identify which areas cause confusion. The questions are as follows;

1. “What did you find difficult to understand about your health insurance cover? Give an example of any vocabulary or industry terms (jargon) which you found confusing”
2. “Give a second example of any vocabulary or industry terms (jargon) which you found confusing”

For the experimental study, there are two open questions to elicit the language and the mental model of participants relating to health insurance benefits;

1. “What is the most important benefit you look for in a health insurance policy when purchasing one?”
2. “There are more than 60 benefits associated with each health insurance plan. List as many as you know off the top of your head without searching the internet.”

These are followed by a series of questions which compare one or more benefits between two health insurance plans. The questions are templated as “Which plan is better for <benefit>?” where “<benefit>” is a single health insurance benefit or else a category of benefits. The participant is given a choice between two “<benefit>” options at each stage. The comparison questions to participants are designed in a way to gradually increase their cognitive load. This is in line with the strategy of starting with simplified material and gradually increasing the complexity of the task [38]. The quiz is designed to ask simpler questions initially to allow the participants to become comfortable interacting with the chatbot. Table 3.2 illustrates a simple comparison question. There is only a single line in each plan for the benefit value and the first plan has no coverage at all. It is straightforward to choose “plan B” as the better plan in this scenario.

Which plan is better for cancer accommodation support?	
Plan A	Plan B
No cover	€100 for 1 night's accommodation for each treatment

Table 3.2 A Single Benefit Comparison Question

Later in the experiment questions become more difficult as the number of sub items in a product value increases and different sub items exist on each plan. Table 3.3 shows a sample comparison question with multiple sub items stated within each plan. It can be seen that there is a cognitive load in re-arranging and evaluating the sub-items within this single product attribute in order to choose the better plan.

Which plan has better cover for a semi-private room in a private hospital?	
Plan A	Plan B
coverage with once-off €300 excess per family per year. No excess applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details.	full cover but 90% cover for certain cardiac and special procedures in the Hermitage and the Galway Clinic

Table 3.3 A Benefit Comparison with Sub Items

The question is an example of a situation where the providers have used different parameters to specify the value of the benefit. The values are not directly comparable since

one plan specifies “excess” and the other does not and one plan explicitly names hospitals such as the “Galway Clinic” while the other vaguely references “12 private hospitals".

3.5.1 Sequence of questions

The following list describes the sequence of questions posed to participants in the order in which they are posed.

1. Two open benefit questions.
2. Five single benefit comparison questions.
3. Two category comparison questions.
4. One question to evaluate which plan is better for the benefit which was chosen by the participant as their most important benefit.
5. One question to ascertain the participant’s first impression of the chatbot.
6. One question to ask how the chatbot could be improved.
7. Ten System Usability Scale(SUS) questions

3.6 Design of the Evaluation Framework

The evaluation framework is designed with multiple, varied metrics and open questions so that qualitative and quantitative data can be collected. The utterances typed during the study can be compiled into a dataset to train the chatbot [60]. The results can be triangulated and an overall impression of the chatbot formed. The individual measurements are detailed in the following sections.

3.6.1 Answers to Open Benefit Questions

The open benefit questions ask the participants to list benefits known to them and to state the particular benefit which is of the most importance to them. The answers will enable comparison of the language and domain model of participants with that used by the industry and its regulator, the HIA.

3.6.2 Answers to Single Benefit Comparison Questions

Single benefit questions will compare one individual benefit (product attribute) between both plans. Some of the single benefit questions have determinate answers and it is clear which plan has better cover for the benefit. The answers to these questions will indicate whether participants can interact with the bot on a basic level and understand the bot’s responses.

Other single benefit questions are indeterminate and either plan could be chosen for different reasons. The evaluation here does not concern the answer itself but the reason from a participant explaining why they choose a particular option. The reason will indicate the level of understanding by the participant of the domain and also of the bot’s response.

3.6.3 Answers to Category Comparison Questions

Comparison of categories such as cancer which has four individual benefit items is arbitrary and either plan could be chosen for different reasons. The evaluation here will not concern the answer itself but the reason from a participant explaining why they choose the particular

option. The reason indicates the level of understanding of the domain and the bot's response on the part of the participant. There will be an additional feedback question for categories, "Explain how easy or difficult it was to compare all of the benefits from that category" which will be used to evaluate navigation of categories. Category navigation can also be evaluated from the chatbot message logs.

3.6.4 Evaluation of the Navigation of Categories of Benefits

The chatbot message logs can be used to determine if the participant can find every item in a category. The log of both user and bot messages can be examined to see if every required navigation path is followed.

3.6.5 Chatbot Intent Detection and Entity Recognition

The chatbot message logs can be used to determine whether the chatbot detected the correct user intent and recognized the entities. Human evaluators can mark the messages as "Understood" or "Misunderstood" and the percentages of misunderstood requests can be calculated.

3.6.6 Goal & Task Completion

Every comparison question posed to the participant is a goal. The completion of the goal will occur when an answer is submitted. Category comparison questions can be decomposed into navigation of the individual benefits of the category. Each navigation step can be measured as a completed task.

3.6.7 Time Taken to Achieve Goals

The quiz web application can measure the time taken to answer each question thereby providing the goal time in each case.

3.6.8 User Feedback

Qualitative evaluation of the chatbot by participants can be measured from the feedback questions posed directly after all goals have been completed.

3.6.9 System Usability Scale (SUS)

A standard measure of usability is desired and so the System Usability Scale (SUS) is chosen. SUS is used as a standard evaluation of user interfaces including chatbots [59]. The ten SUS questions will be the final stage of the chatbot experiment and the answers will form the basis of the SUS score to evaluate usability.

3.6.10 The Implementation Experience

The experience of using open source frameworks, libraries and tools will be recorded and then described in **Appendix Reflections on the Implementation**

3.6.11 Iterative Approach

The chatbot is implemented in two stages as an iterative approach is considered best practice in UI design [62]. An experimental study (Study 1) is performed after the first iteration. A quantitative and qualitative assessment is made and the results are used to inform further development of the chatbot. A second experimental study (Study 2) is used to evaluate the second iteration. The same questions are asked and the same measurements applied in both studies to enable comparison.

3.7 Summary

The chapter describes how the design of the research follows the primary objective to evaluate how a chatbot can cope with a domain consisting of complex language and data. An initial survey is designed to inform the breadth of the challenge, exploring the difficulties presented to users who are preparing to purchase the domain's products. A case study approach is used to design and produce a prototype chatbot with an experiment conducted to evaluate the prototype. Open questions are designed to reveal the dissonance or agreement between the mental model of the participants and the domain model. The level of difficulty of questions is planned to increase as the participant becomes comfortable interacting with the chatbot. The evaluation framework is designed with multiple quantitative metrics and open qualitative questions. This enables triangulation of results to form an overall impression of the chatbot. The utterances typed by the participants are used to prepare a dataset for training the chatbot in subsequent development cycles. Navigation by the user through the hierarchy of benefits in the domain is exposed by category questions. A standardised scale, SUS, is chosen to compare multiple executions of the study.

Chapter 4 : Implementation

This chapter discusses the implementation of the end to end chatbot solution and the deployment which made the solution publicly available to participants. It also describes the custom quiz application which was developed to present questions to participants as they interacted with the chatbot. Figure 4.1 presents the Rasa chatbot architecture again for context below.

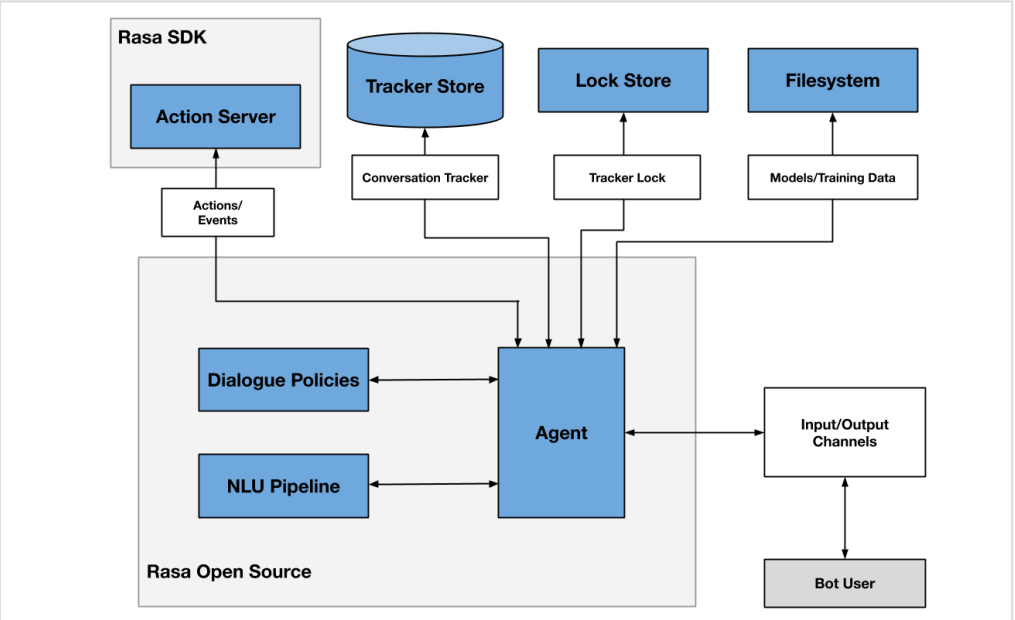


Figure 4.1 Rasa Chatbot Architecture

4.1 Implementation of the Chatbot

Rasa and other open source components were configured and custom Rasa code was written. Rasa was chosen for its breadth and flexibility as described in 2.9 Rasa Open Source .Additional python code was developed as was javascript, html and css for the web components and applications.

4.1.1 Chat Widget

A front end open source javascript application was required to take user input and present the chatbot framework responses. The final solution used an open source chat widget available on github [60]. The solution was developed as open source by a company called Botfront which was acquired by a Canadian company named Dialogue. Figure 4.2 shows a sample screen from the chatbot widget where the user has typed a question, “What benefits do you have for cancer?”

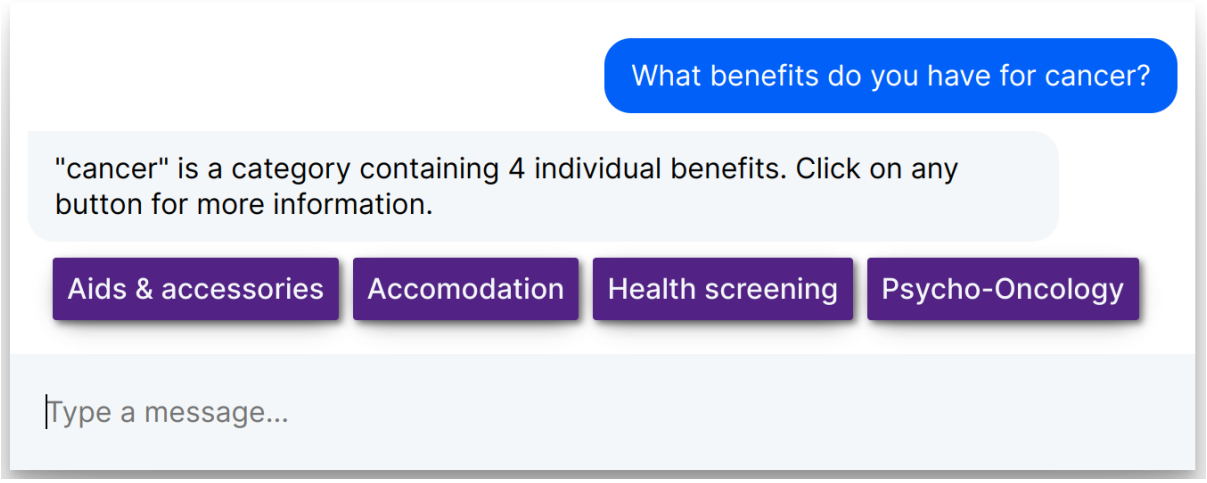


Figure 4.2 Chatbot Widget Screen

Initially the Rasa widget was trialled and this was later abandoned when it exhibited problems such as clipping longer responses. It also lacked functionality such as html buttons.

4.1.2 Knowledge Base

An external knowledge base was developed which could be queried from custom actions running in the Rasa action server. The knowledge base was implemented with structured json files and python code which loaded these structures into memory for querying. A domain taxonomy with a fixed set of 8 categories from the Health Insurance Authority (HIA) was extended to be a broader ontology in the knowledge base as seen in the following diagram. Figure 4.3 illustrates how the HIA taxonomy was extended with aliases and benefit_tags to become an ontology for the chatbot.

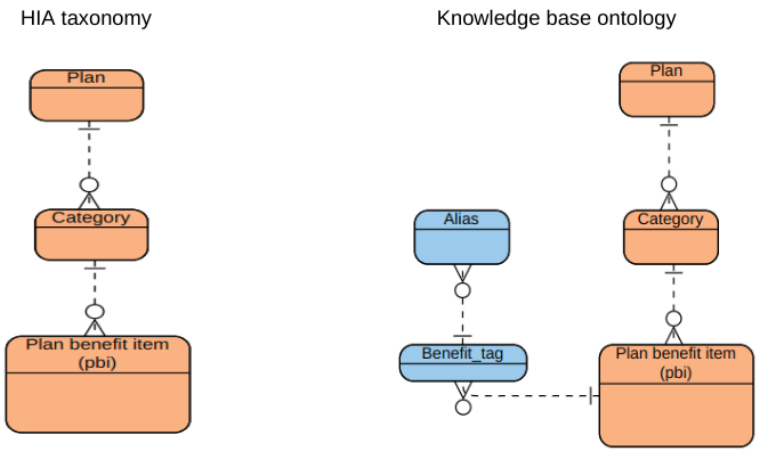


Figure 4.3 The Chatbot Ontology Compared to the HIA Taxonomy

In the HIA taxonomy a plan benefit item (*pbi*) can exist in a single category only. This is very limiting since most benefits could fit in multiple categories, e.g. a “delivery consultant” belongs to the categories of “consultants” and “maternity”. In the chatbot ontology a *pbi* can exist in multiple *benefit_tag* categories. Moreover, multiple *aliases* can be mapped to each *benefit_tag* to translate participant terminology into the ontology. The entities in the knowledge base are outlined in the following sections.

4.1.2.1 Plan Benefit Item (pbi)

The HIA taxonomy had more than 60 line items for benefits. Each of these was termed a *plan benefit item* (*pbi*) in the chatbot ontology. Every *pbi* had a unique key in the ontology e.g. “Child Home Nursing” was “child_nurse”.

4.1.2.2 Category & Benefit_tag

The HIA taxonomy has only 8 categories to contain the benefit line items and each line item could be in a single category only. The participants in the chatbot experiment categorised benefits in very different ways. The chatbot ontology had 61 benefit_tags to reflect this and each pbi could be in multiple categories. Table 4.1 presents the HIA taxonomy in the leftmost column to illustrate how few categories it contains relative to the chatbot ontology which covers much more of the natural language of the domain users.

HIA taxonomy categories	**Chatbot ontology benefit_tags**		
Prices Per Annum In Patient Maternity Out-patient Benefits Out-patient Radiology Overseas Psychiatric Cover Fertility Benefits	a and e alternative medicine ask insurer breastfeeding cancer cardiac child healthcare chiropody chiropractor consultant convalescence day case dental dietitian emergency employee assistance excess family excess fast access fertility	gender gender reassignment surgery gp health screening healthcoach hearing high tech hospital home nursing hospital overseas hrt for gender dysphoria individual excess inpatient liver disease maternity occupational therapy optical orthoptistry osteopath outpatient overseas	physio podiatry policy limit post natal pre natal price private hospital private room psychiatry psychology public hospital radiology reflexology repatriation routine dental semi-private room special procedures speech therapy surgery vaccination

		parent with child	

Table 4.1 Comparing Scope of HIA Categories with Scope of Chatbot Benefit Tags

4.1.2.3 Sample PBI Definition

Figure 4.4 shows a sample definition of a plan benefit item (*pbi*). The key in the json file is “private_hospital_semi” in this case and the relevant category in the HIA taxonomy is “inpatient”.

```
        "private_hospital_semi": {
            "button": "Semi-private in private",
            "hiaCategory": "inpatient",
            "hiaDescription": "Private Hospital Semi-Private Room",
            "faqExplanation": "This means a bed in a private hospital which
is in a room shared with no more than 5 other patients",
            "benefit_tags": [
                "hospital",
                "private hospital",
                "semi-private room"
            ],
            "comment": "Does not include day case!!"
        }
```

Figure 4.4 A Sample Definition of a Plan Benefit Item (pbi)

4.1.2.4 Alias

Aliases were used to connect health insurance benefits to the natural language of domain users As testing with real users progressed, terms emerged in the chatbot logs of the utterances of participants. The participants used a much more varied language than had originally been specified in the chatbot ontology. These newly discovered terms were implemented as *aliases* which pointed to *benefit_tag* categories. Post processing was done to remove common words which lacked semantic value in the context of the chatbot including “cover” and “benefit”. Table 4.2 Presents the complete list of aliases which were mapped to health insurance benefits in alphabetical order.

Aliases		
A and E a&e access to a gp accident accident and emergency birth cardio cardiovascular charge cheap	emergency everyday excess for one person eyes free glasses general practitioner good price having a baby hearing aid heart	natal oncology online doctor outpatient radiology phone assistance overseas podiatry post operative post-op postnatal Pre Existing condition

cheaper chiropody clinical psychologist conceiving conception cost counselling day to day deafness digital doctor doctor	heart disease high-tech hospitals hospital stay hrt illness infertility inpatient accommodation inpatient stay language mental health motherhood	premium prenatal Prices per annum psycho therapy quick access radiology recovery return of a body return of remains technologically advanced hospital urgent

Table 4.2 The List of Aliases for Benefits

4.1.2.5 API to Retrieve Values of Plan Benefit Items (pbis)

Each plan benefit item (*pbi*) had a value in both of the plans in the study. The API to get the value of a *pbi* in a particular plan consisted of python code to read the structured json files containing the values. Table 4.3 shows some sample *pbi* values. These values were manually extracted from the HIA comparison documented in Appendix: Comparison of Two Popular Health Insurance Plans from the HIA and populated in the json files which were independently validated.

Plan A	Plan B
"child_nurse": "€80 x 28 days", "pre_post_natal": "Not covered",	"child_nurse": "€100 x 14 days", "pre_post_natal": "€255",

Table 4.3 Sample Values for Plan Benefit Item(pbi)s

4.1.3 NLU Pipeline

An NLU pipeline is a key component of a task oriented dialogue system. Figure 4.5 shows a typical pipeline as a sequence of components designed to process the user utterance in stages.

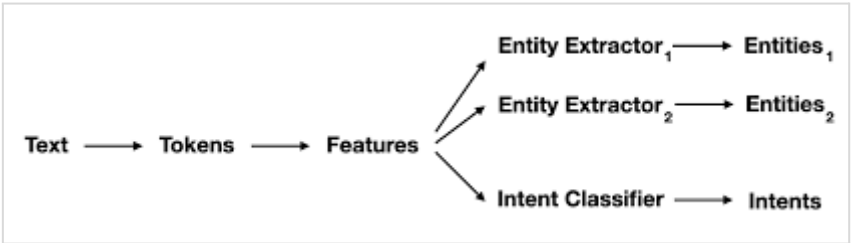


Figure 4.5 A Typical NLU Pipeline

Table 4.4 explains the NLU pipeline as it was configured in Rasa. Explanations of the pipeline configuration are in the right hand column of the table, alongside the relevant configuration item.

`pipeline:` `  - name: WhitespaceTokenizer`	User utterances are tokenized using the WhitespaceTokenizer.
`  - name: RegexFeaturizer` `  -name: LexicalSyntacticFeaturizer`	The RegexFeaturizer provides features for the intent classifier The LexicalSyntacticFeaturizer creates lexical and syntactic features for a user message to support entity extraction

`  - name: CountVectorsFeaturizer`
`  - name: CountVectorsFeaturizer`
`    analyzer: char_wb`
`    min_ngram: 1`
`    max_ngram: 4`
	The CountVectorsFeaturizer creates a sequence of token counts features based on sklearn's CountVectorizer

`  - name: DIETClassifier`
`    epochs: 100`
`    constrain_similarities: true`
	The DIETCLassifier is a key component developed by the Rasa team. It is a transformer-based model that acts as both an entity extractor and an intent classifier.

`  - name: EntitySynonymMapper`	The EntitySynonymMapper is used to map between extracted entities and their synonyms

`  - name: ResponseSelector`
`    epochs: 100`
`    constrain_similarities: true`
	The ResponseSelector embeds user inputs and candidate responses into the same space. Supervised embeddings are trained by maximising similarity between them. It is based on the star space model

`  - name: FallbackClassifier`
`    threshold: 0.3`
`    ambiguity_threshold: 0.1`
	The FallbackClassifier ensures that an intent named *nlu_fallback* is predicted when all other intent predictions fall below the configured threshold value. This is configured in *config.yml* to invoke the action, *action_default_fallback* when NLU prediction is very uncertain. For this project the bot will state the response “I'm sorry. I didn't quite understand what you said.”

Table 4.4 The NLU Pipeline Explained

4.1.4 NLU Training

Supervised learning from NLU examples was the primary method used to train the NLU. The examples were derived from participant utterances captured during beta testing and experimentation.

4.1.4.1 Supervised Learning

Supervised learning from NLU examples proved to be very effective with Rasa. These examples were classified by intent and annotated with entities With few examples, Rasa correctly predicted intents. Figure 4.6 shows a sample intent named *ask_pbi_or_benefit_tag* The intent lists example utterances such as “What does this plan cost?” with labelled entities such as the “cost” *benefit_tag*.

```
- intent: ask_pbi_or_benefit_tag
  examples: |
    - what does this plan [cost](benefit_tag)
    - how is [cardiac](benefit_tag) [covered](junk)
    - how much is [excess](benefit_tag) for [plan a and b](plan)
```

Figure 4.6 An Intent with Example Utterances

The NLU was tested and a confusion matrix generated using the command , `Rasa test` Figure 4.7 shows the confusion matrix produced for the research project.The diagonal shows that the NLU examples are well designed. For each intent in the chatbot domain, the confusion matrix shows how often the intent was correctly predicted and how often an incorrect intent was predicted instead. The incorrect predictions would lie in an area of the chart outside of the diagonal. The deeper blue colours indicate intents with more training examples than others.

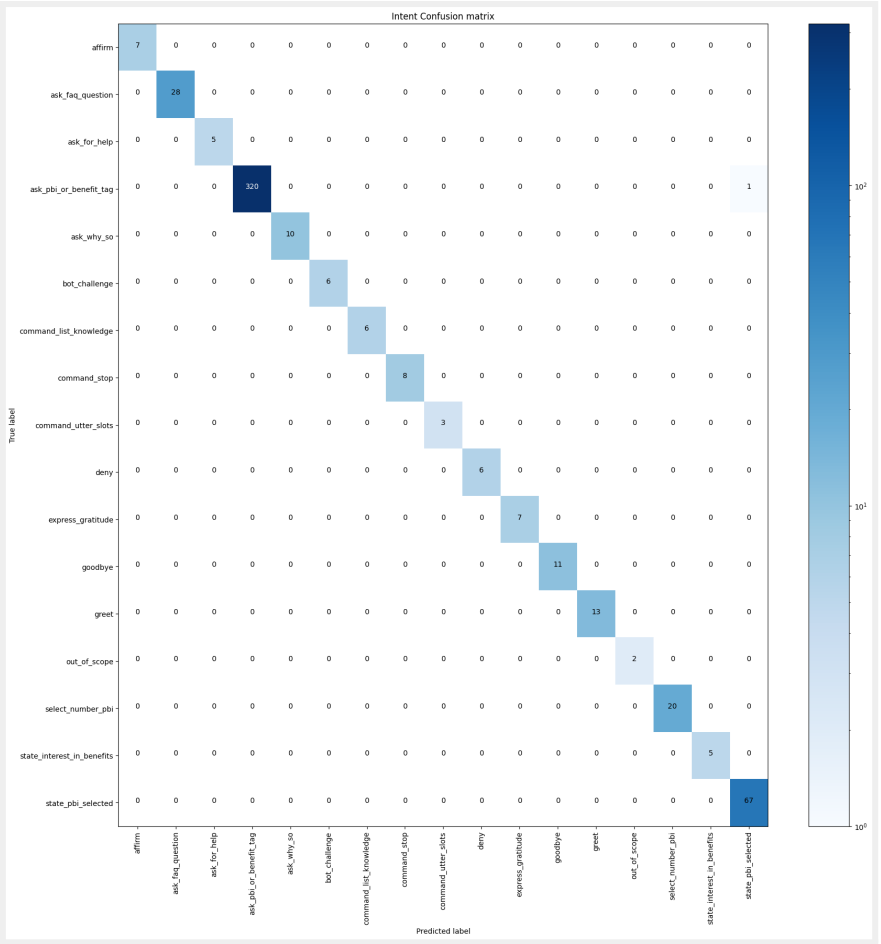


Figure 4.7 The Confusion Matrix

4.1.4.2 Reinforcement Learning

Rasa offers an interactive tool for reinforcement learning. The Rasa blog describes this as follows, *“In interactive learning mode, you provide step-by-step feedback on what your bot decided to do. It's kind of like reinforcement learning, but with feedback on every single step (rather than just at the end of the conversation). “*
This tool generated elaborate data in the yaml file, *story.yml* which was difficult to reason about and harder to maintain. It was decided not to use this tool to develop the chatbot for the project.

4.1.4.3 NLU Python Scripts

As the ontology grew the NLU training data could no longer be hand crafted. Python scripts were used to extract data from the ontology into the format of Rasa NLU examples. Figure 4.8 contains python code using python templates to create NLU examples for the intent *ask_pbi_benefit_tag*

```
def nlu_ask_pbi_benefit_tag_alias():
```

```
    templates=[
        Template("    - what is the [$ben]($ben_type) in [plan A](plan)"),
        Template("    - is [$ben]($ben_type) included"),
        Template("    - tell me about [$ben]($ben_type) "),
        Template("    - what [$ben]($ben_type) can I get"),
        Template("    - what is the [$ben]($ben_type) "),
        Template("    - what does [plan b](plan) have for [$ben]($ben_type)
"),
        Template("    - which plan is better for [$ben]($ben_type) "),
                Template("       - what [cover](junk) do you have for
[$ben]($ben_type) "),
        Template("    - [$ben]($ben_type) [care](junk)"),
        Template("    - [$ben]($ben_type) [cover](junk)"),
        Template("    - [$ben]($ben_type) [benefit](junk)")
    ]
     # many users just state the name of the benefit_tag standalone with no
surrounding words
    standalone_templ=Template("    - [$ben]($ben_type) ")
    position=0
    for name,key in sorted(benefit_tags.items()):
        templ=templates[position]
        print(templ.substitute(ben=name, ben_type="benefit_tag"))
                                print(standalone_templ.substitute(ben=name,
ben_type="benefit_tag"))
        position=position+1
        if position>=len(templates):
            position=0
```

Figure 4.8 Sample Python Code for Generating NLU Examples

The python method was executed to produce Rasa *nlu.yml* examples. Figure 4.9 demonstrates NLU utterances such as “what cover do you have for child healthcare?” which were labelled with appropriate entities.

```
- what is the [cancer](benefit_tag)
- [cancer](benefit_tag)
- what does [plan b](plan) have for [cancer accommodation](benefit_tag)
- [cancer accommodation](benefit_tag)
- which plan is better for [cardiac](benefit_tag)
- [cardiac](benefit_tag)
- what [cover](junk) do you have for [child healthcare](benefit_tag)
- [child healthcare](benefit_tag)
- [chiropody](benefit_tag) [care](junk)
- [chiropody](benefit_tag)
```

Figure 4.9 Example Rasa NLU Utterances

Through the use of these scripts, the templates could be easily changed as new user testing revealed new intent expressions The ontology grew and the NLU examples could be regenerated in seconds. The longest delay in the process was in retraining the model which

took five minutes or more as the number of examples grew. This time could be reduced by partial training using the command, `rasa train –finetune`

4.1.5 Intents

A convention for naming intents was used which followed the appropriate speech acts, defined by Searle [61]. This was done by the researcher to organise the code in a systematic way for readability. Commissive and declarative speech acts were not required for this project. Table 4.5 shows the speech acts with corresponding intents where those acts were applied in the research project.

Speech act	Description	Example	Intents
Commissive	Promising	*I promise to pay the bill on Tuesday*	N/A
Representative (Assertive)	Stating facts	*Dublin is the capital of Ireland*	state_interest_in_benefits state_pbi_selected
Declarative	Making something happen by declaring it	*I now pronounce you man and wife*	N/A
Directive	Asking or commanding	*Can you pass the butter?*	ask_for_help ask_faq_question ask_pbi_or_benefit_tag ask_why_so command_utter_slots command_stop command_list_knowledge select_number_pbi
Expressive	Expressing emotion	*I'm so appreciative of all your help!*	express_gratitude
Other	Acts which are standard Rasa examples or acts which do not easily fit into the categories		greet goodbye bot_challenge

Table 4.5 Speech Acts & Intents

Some problems were discovered concerning intent classification. Table 4.6 presents a sample of the problems discovered alongside the corresponding intent and training examples. More problems are listed in Appendix Chatbot Single Turn Errors and Appendix Chatbot Multi-turn Errors

Intent	Examples	Problems
ask_faq_question	explain [agreed charges](faq_term) what is [alternative medicine](faq_term) what do you mean by [cardiac procedures](faq_term)	Participants often used the same expression to ask a question about the value of a benefit on a plan e.g. "*What is the cancer accommodation?"* In this case we would wish a different intent to be predicted, *ask_pbi_benefit_tag*
ask_faq_question	What is the limit of the visits?	The NLU predicted ask_faq_question but the participant was not asking an FAQ type question. They were referring to a previous benefit, 'GP ' The NLU could not predict that this is referring to the benefit already stated. This functionality could be implemented with custom action code to look at the dialog history

Table 4.6 Problems with Intents

4.1.6 Dialogue State Tracking

Table 4.7 shows the slots used to track the dialog state. They are configured in *domain.yml* in Rasa.

benefit_tag_slot text_pbi_slot number_pbi_slot pbi_object_slot pbi_slot pbi_list_slot aspect_of_cover_slot	alias_slot is_alias_slot faq_term_slot knowledge_item_slot faq_term_object_slot plan_slot junk_slot

Table 4.7 Dialog Slots

Rasa extracts entities and passes them through validator functions which set the slots. The validator functions are useful for setting related slots or rejecting extracted entities. Table 4.8 lists the custom validator functions developed.

Global validators	Validators in number_pbi_form

validate_knowledge_item_slot validate_faq_term_slot validate_benefit_tag_slot validate_pbi_slot Knowledge_item_slot	validate_number_pbi_slot validate_text_pbi_slot

Table 4.8 Custom Validators

4.1.7 Dialogue Policies

A hybrid dialogue system was implemented to cover as many user journeys as possible with the minimum coding. For example, a rule based policy could easily be configured for situations where the response is always the same - every time the user says hello the bot should say hello . Three dialogue policies were used in conjunction with each other; rule based, memoization (story based) and transformer based(TED). For each user utterance the policy which predicts the next action with the greatest confidence is the winner. Initially the UnexpecTEDIntentPolicy was trialled but this is an experimental feature and it was found that this policy predicted the action, *action_unlikely_intent,* too often to be useful. Table 4.9 shows the final policy configuration applied to Rasa.

```
policies:
  - name: MemoizationPolicy
  - name: RulePolicy
    core_fallback_threshold: 0.4
    core_fallback_action_name: "action_default_fallback"
    enable_fallback_prediction: True
  - name: TEDPolicy
    max_history: 5
    epochs: 100
    constrain_similarities: true
```

Table 4.9 Policy Configuration

In the case when the policy predictions are a tie then the following order of priority applies;

1. RulePolicy
2. MemoizationPolicy
3. TEDPolicy

4.1.7.1 Rule Policy

The RulePolicy was used to hard code rules for situations where the next action should always be the same. For example, every time the user says “goodbye” the bot should say “goodbye” also. The RulePolicy is also required for Rasa forms. Figure 4.10 is an extract from the policy configuration including a rule for the chatbot to always say “goodbye” whenever the user says “goodbye”.

```
- rule: Say goodbye anytime the user says goodbye
  steps:
  - intent: goodbye
  - action: utter_goodbye

- rule: Submit number_pbi_form
  condition:
  - active_loop: number_pbi_form # given that this form is active
  steps:
  - action: number_pbi_form
  - active_loop: null
  - slot_was_set:
     - requested_slot: null # we are no longer in need of requesting extra
slots
  - action: action_state_plan_benefit_values
```

Figure 4.10 Rule Policy Configuration

4.1.7.2 Memoization Policy

The Memoization policy follows stories exactly how they are written, similar to a finite state machine. The MemoizationPolicy remembers the stories from your training data. It checks if the current conversation matches the stories in your stories.yml file. If that is the case then it will predict the next action from the matching stories of your training data with a confidence of 1.0. If no matching conversation is found, the policy predicts *None* with confidence 0.0. Table 4.10 shows a sample story, with an explanation in the second column.

`- story: answer faq question`
`  steps:`
`  - intent: ask_faq_question`
`  - action: action_state_faq_answer`
	This is a simple story. When the intent of the user is to ask an faq question then the Memoization policy should predict the next action to be *action_state_faq_answer*

Table 4.10 A Sample Story

4.1.7.3 TED Policy

This is a dialogue policy based on a transformer architecture, where the self-attention mechanism operates over the sequence of dialogue turns. This is described in a research paper from the Rasa team [19]. This policy enables users to follow diverse conversation paths not explicitly elaborated in *stories.yml*. There was no additional configuration required for this policy.

The Memoization policy predicts the next action only when there is an exact match with a story whereas the TEDPolicy can predict the correct next action regardless of the user's journey in and out of stories. The Rasa events table from the sqlite tracker store database

shows the prediction in good detail. Figure 4.11 shows a sample prediction where the next action is to present the form, “number_pbi_form”.

```
{
    "confidence": 0.9903851747512817,
    "event": "action",
    "name": "number_pbi_form",
    "policy": "TEDPolicy"
}
```

Figure 4.11 A Sample Prediction of the Next Action

Note that the events table only shows the winning prediction. It would be very useful for debugging if the losing predictions of the other policies were also shown.

4.1.8 Conversation Flow

Following experimentation with beta chatbot users, it was thought the typical conversation would follow a story as documented in Table 4.11 below. The user utterance is in the left hand column with the chatbot response on the right.

User	Chatbot
Hi	Hello
I am interested in the benefits on your plans	I have two different health insurance plans, plan A and plan B. I can tell you about the benefits on these plans. You can say, for example, "tell me about cardiac"
Tell me about cancer	“Cancer” is a category with 4 individual benefits Say the benefit name or click any button to hear more
“Cancer accommodation”	Concerning “cancer accommodation”, plan A has €100 for 1 night's accommodation for each treatment and Plan B has no coverage

Table 4.11 A Typical Conversation

Figure 4.12 shows how the same conversation story is expressed in the corresponding Rasa *stories.yml* syntax below.

stories.yml

```
- story: benefit inquiry for single pbi or multiple
  steps:
```

```
  - intent: state_interest_in_benefits
  - action: utter_ask_which_benefit
  - intent: ask_pbi_or_benefit_tag
  - action: action_state_pbis_for_choosing
  - action: number_pbi_form
  - action: action_state_plan_benefit_values
```

Figure 4.12 Rasa Stories Configuration

During beta testing it was discovered that users often did not follow this expected dialogue path. Some users jumped straight into asking about a particular plan benefit item (*pbi*) with utterances such as “What cover do you have for convalescence?” Figure 4.13 shows the flow of the original story and a flow which was discovered to be necessary after user testing.

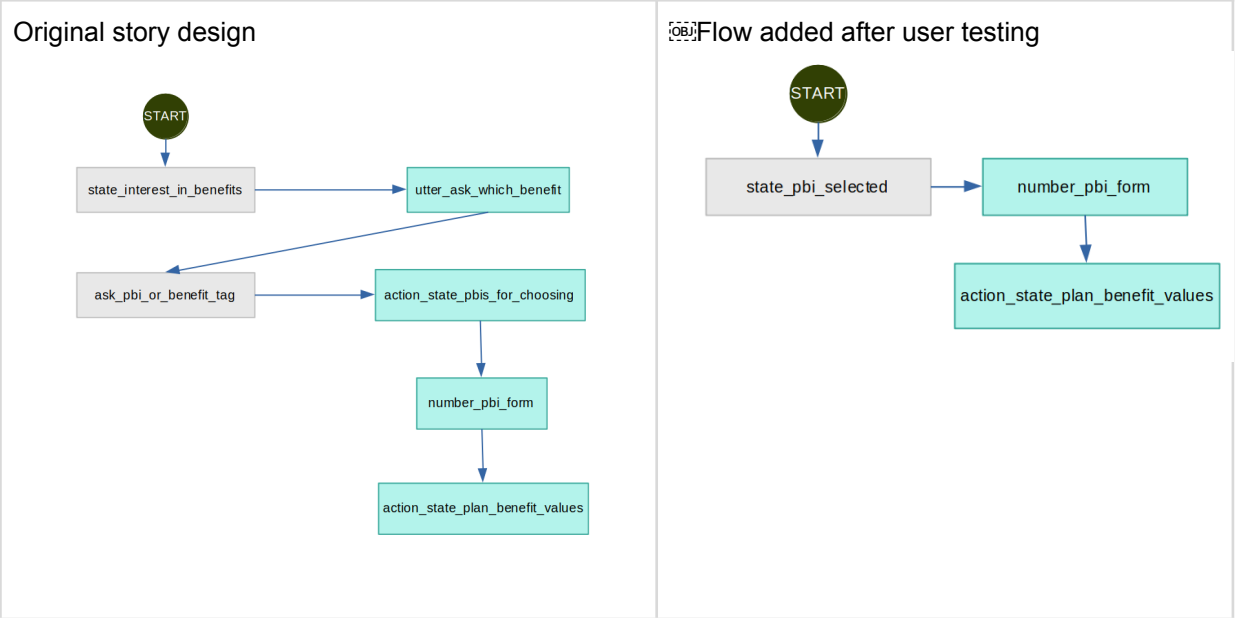


Figure 4.13 Conversation Flows

When the chatbot logs were examined following user testing it was seen that diverse conversation paths were attempted by participants and new user stories were developed to train the policies on handling these paths. Table 4.12 shows stories from the Rasa config which enable the user to jump around to different parts of the flow.

stories.yml	

`- story: benefit inquiry for single pbi or multiple`
`  steps:`
`  - intent: state_interest_in_benefits`
`  - action: utter_ask_which_benefit`
`  - intent: ask_pbi_or_benefit_tag`
	These stories enabled the participant to jump in and out of the following contexts
	• navigating a category of benefits
	• simply stating a particular plan benefit item
	• proceeding to a completely different

- action: action_state_pbis_for_choosing - action: number_pbi_form - action: action_state_plan_benefit_values - intent: select_number_pbi # added to select another pbi - action: number_pbi_form - action: action_state_plan_benefit_values - story: selecting a pbi by its name steps: - intent: state_pbi_selected - action: number_pbi_form - action: action_state_plan_benefit_values	category.

Table 4.12 Stories Enabling Participants to Jump Around in the Flow

4.1.9 Action Server

The Rasa action server is a separate sanic (python) server from the main Rasa core server. Custom actions were very useful for looking up the knowledge base and generating appropriate dynamic responses as in the sample code below. The code reads the plan benefit item (*pbi*) from a slot and retrieves its value in both plans. Figure 4.14 shows sample python code for an action which checks that the value of a slot exists before stating the value in a response to the user.

```
pbi = tracker.get_slot("pbi_object_slot")
if pbi is not None:
   utils.state_pbi_values(tracker,dispatcher,pbi)
```

Figure 4.14 Sample Action Code

Custom slot validator code was also deployed to the action server. These validators were described in a previous section on dialogue state tracking.

4.2 Voice Channel Implementation

A voice channel was implemented and beta tested but not used for the studies. This is described in the Appendix Voice Channel Implementation

4.3 Deployment of the Chatbot

The end to end architecture of the chatbot was presented in the previous design chapter. The individual components are described in the following sections. Figure 4.15 repeats the end to end architecture diagram for context

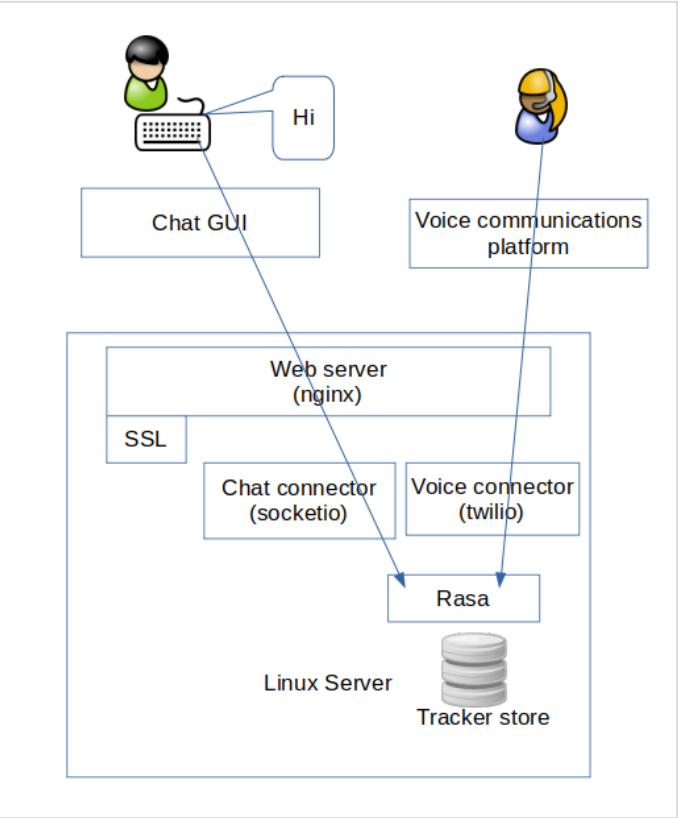


Figure 4.15 End to End Chatbot Implementation

4.3.1 DNS

The chatbot needed to be available on the public internet so that users recruited via the prolific website could access the bot for testing. A domain name was registered, *frankshealthinsurance.xyz*, and a DNS type A record was maintained. Table 4.13 explains the DNS record.

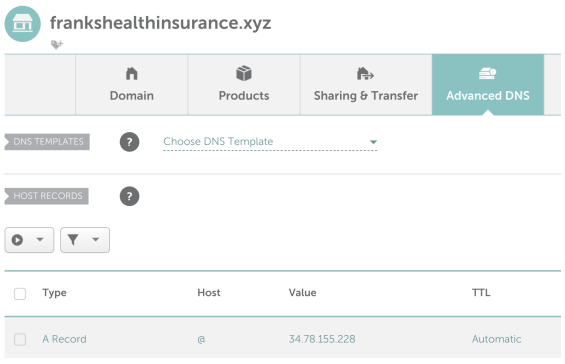
	Maintenance of the DNS record was required because every time the Google Cloud Platform(GCP) virtual machine instance is restarted it acquires a new ip address. This necessitates updating the A record after every restart and waiting a few minutes for DNS propagation. DNS enabled a consistent, named url to be used for testing.

Table 4.13 The DNS Record

4.3.2 SSL

SSL from “Lets encrypt” was used. The certificate needed to be renewed every 3 months but the automated tools provided by “Let’s Encrypt” worked well to keep the certificate current.

4.3.3 Nginx

Nginx was used as the web server to serve up the html and javascript for the chat widget and also to proxy the API requests to the Rasa server. During test and development ngrok was used instead of nginx to tunnel traffic from localhost for speedier development cycles.

4.3.4 Ubuntu Linux Server

A linux server running Ubuntu 20 was set up as a vm instance in Google Compute Engine (GCP). The server was run in the environmentally friendly Belgium data centre , europe-west1-b . The server only operated for the duration of the chatbot experiment and could be started and stopped at will. Table 4.14 lists the linux scripts required.

`cleanup_logs.sh`
`kill_Rasa.sh`
`start_Rasa.sh`
`zip_results.sh`
	Custom linux scripts were written to start and kill the Rasa processes and also to zip the survey results. The results included the sqlite database of answers and events, the Rasa server log and the Rasa action server log.

Table 4.14 List of Linux Scripts

4.3.5 Input / Output Channels

Channels are the means by which the Rasa server communicates with the outside world. Two channels were supported for this project, one for chat and the other for voice. Table 4.15 shows the Rasa channel configuration.

Chat channel (using websockets)	Voice channel (using Twilio’s communications platform)

`socketio:`
`  user_message_evt: user_uttered`
`  bot_message_evt: bot_uttered`
`  session_persistence: true/false`
	`twilio_voice:`
	`  initial_prompt: "Hi"`
	`  assistant_voice: "Polly.Matthew-Neural"`
	`  reprompt_fallback_phrase: "Can I help you with something else?"`
	`  speech_timeout: "auto"`
	`  speech_model: "numbers_and_commands"`
	`  enhanced: "false"`
	`  record: "true"`
	`  account_sid: "***"`
	`  auth_token: "***"`
	`  no_repeat: "true"`

Table 4.15 Rasa Channel Configuration

4.3.6 Agent / Core

The Rasa agent is a core component which loads the model and connects all components and channels together. It runs in an asynchronous sanic (python) server. No modifications to the core agent were required for this project.

4.4 Custom Quiz Web Application

A custom quiz web application was created for the chatbot study to enable questions to be asked of participants at the same time as they interacted with the chatbot. This meant immediate feedback relevant to the domain product benefit in question could be captured from participants while they were still reflecting on it. Ethics were considered here and so the participant was not presented with any questions of a personal nature. They were also assured that any personal information which was inadvertently typed would be promptly deleted and not used in the research. They were presented with a consent form before the quiz began and given a choice to withhold consent and opt out. This Single Page Application (SPA) consists of a javascript front end application which embeds the chatbot widget in an iframe. It was developed using the open source vue framework. Figure 4.16 shows a screen from the quiz web application where the user has selected a question to answer, “Which plan is better for convalescence?”

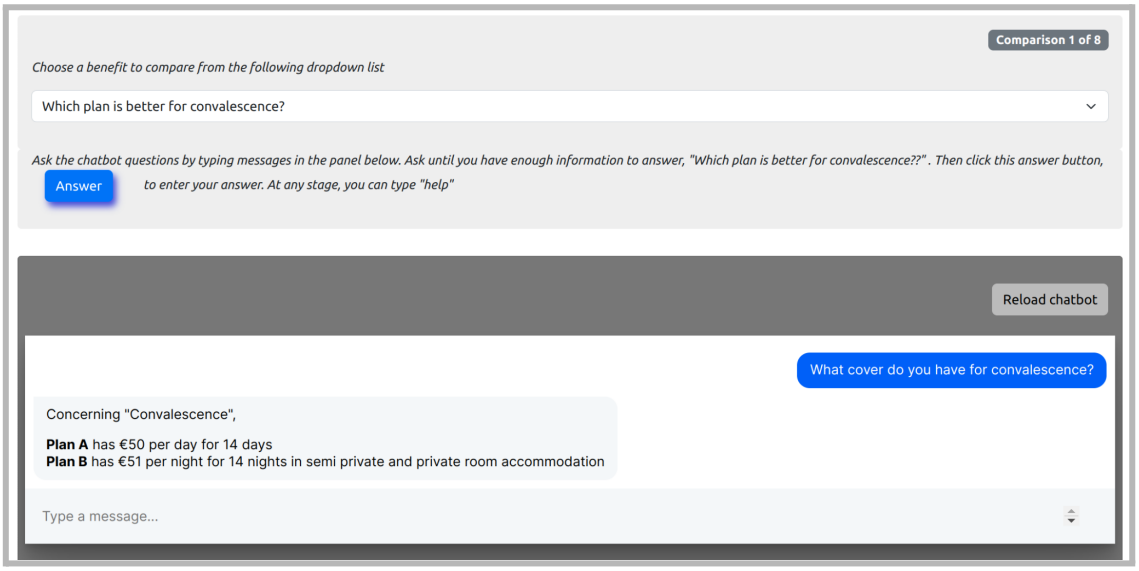


Figure 4.16 The Quiz Web Application

This enabled a series of focused questions to be posed to the participant with the chat interface available on the same screen. This makes it easy for the participant to focus on

each individual question in turn. The design of the questions is described in 3.5 Question Design and 3.6 Design of the Evaluation Framework

The application interacts with python code in a sanic (python) server. The Rasa sanic server was used and extended with a new route "*/answer*" to receive answers posted in a json payload from the front end. The answers given by the user are stored in the newly created *answers* table in the Rasa tracker store sqlite database.

The list of questions was dynamically configurable as a javascript object. This could easily be adjusted during beta testing until the final question set was defined. Figure 4.17 shows how questions can be configured by level. There are two questions for the level "easy1". The stub, "Which plan is better for" is added to form each question e.g. "Which plan is better for convalescence?"

```
{
  easy1: {
    convalescence: "same",
    "cancer accommodation support": "plan B",
  },
  easy2: {
    "overseas hospital": "same",
    optical: "plan B",
  },
  easy3: {
    vaccinations: "same",
    hearing: "same",
  },
  medium: {
    "individual excess": "plan B",
    "family excess": "plan B",
  },
  applesOranges: {
    "a semi-private room in private hospitals": "Other",
    "accident and emergency": "Other",
  },
  category1: {
    psychology: "Plan B",
    gender: "Other",
  },
  category2: {
    radiology: "Other",
    cardiac: "Other",
    cancer: "Plan A",
  }
}
```

Figure 4.17 Question Configuration

4.5 Summary

This chapter discusses the implementation of the end to end chatbot solution and the deployment which made the solution publicly available to participants. A chat widget is

customised for the front end interaction with participants. The knowledge base for the domain is implemented with python code and json files. An appropriate NLU pipeline is configured in the Rasa chatbot framework and trained using supervised machine learning. Python scripts are coded to generate NLU training files. Dialogue state tracking is developed using Rasa slots and custom python validator functions. Three competing dialogue policies are trained; a rule based policy , a story based policy and a transformer based policy. The policy which predicts with the greatest confidence is the one which chooses the next action at any stage of the conversation. Stories are used to train the system for different conversation flows. Custom python actions are coded to use the classified intents and extracted entities to lookup the appropriate product information in the knowledge base. A complex deployment is required to make the end to end system available publicly for the study participants. A custom quiz application is developed to get immediate feedback from participants while they attempt to answer complex comparisons between products in the domain.

Chapter 5 : Initial Survey and Study 1

This chapter and the following chapter detail the results for the evaluation of the chatbot in relation to the objectives of the thesis. These chapters also discuss the given results. A survey and an experimental study were performed as part of the study for this research. An initial survey was performed to establish that there is a problem with the vocabulary in the domain. Then a chatbot for the domain was designed and implemented and a study was performed to evaluate how well the chatbot worked to help participants understand the domain. The initial survey asked two simple questions and was less than 2 minutes in duration. The study experiment was about 20 minutes long and included interaction with a chatbot.

5.1 Initial Survey of the Domain

An initial survey was designed and executed as outlined in the following sections.

5.1.1 Objective

The objective of the initial survey was to establish that the domain in question, health insurance, is complex in the sense that it has technical domain specific language that is not understood by potential customers and to indicate where in the domain confusion would arise for customers and how severe this problem is.

5.1.2 Experimental Setup

The experimental setup included the participants and the survey questions. Participants were chosen on the basis of having some interest in health insurance. They were recruited from the academic research website, Prolific. They were selected on the basis that they had either taken out a health insurance policy within the preceding six months or they were about to take out health insurance within the succeeding six months. Only participants from Ireland were selected as the health insurance plans under research are only available in Ireland. Ethical considerations were applied as described in 3.1 High Level Requirementsfor the Case Study. Forty eight participants undertook the survey. No additional demographic was captured about the participants as this was not required for the evaluation and so would not have been appropriate ethically. Two survey questions were asked as specified in section 3.5 Question Design and no additional material was provided to the participants.

5.1.3 Results of the Survey

Participants were asked which health insurance terms they found to be confusing. Table 5.1 shows the participants' answers. Two categories of answers were discovered, "Terms" and "Product". An example categorised as "Terms" could be insurance jargon such as a "deductible". An example where the participant found the insurance "Product" to be confusing is when they stated that "*it's difficult to understand which options would be covered in which hospitals"*.

Participant	First confusing term	Second confusing term	Category
1	premium	deductible	Terms
2	I found it hard to understand exactly what was covered or what portion of each procedure would be covered	How could I give an idea sample of insurance jargon without having the policy on front of me?	Product
3	Issuer	Term	Terms Product
4	All was clear	None	None
5	Nothing really. I suppose its all the internal limits	excess and time periods	Terms Product
6	I found it difficult to understand which options would be covered in which hospitals and whether or not I should attend public or private hospitals for different issues		Product
7	the prices	unsure	Product
8	Community rating		Terms
9	community rating levies, risk equalisation,	premium credits, surgical applicances benefit	Terms Product
10	Certain procedures were not listed	Common layman terms, were missing and lots of medical terms used	Terms Product
11	No terms but find the costs incurred to me unclear	Which clinics and more specifically consultants covered	Product
12	I found nothing confusing		None
13	No it clearly states what I'm entitled to in a big list	None. The GRA medical aid has a 2 page list of what you're entitled to under their cover and you either pay the 30e a week for the cover or you don't. No jargon or bs	None
14	Excess	Shortfall	Terms
15	everything! i find it difficult to understand exactly what is covered, pre existing minor issues, are they covered. very confusing	the different options, fertility treatment but not told what exactly it covers, same with heart and surgeries etc	Product
16	Waiting periods	Excess	Terms
17	I don’t recall finding anything confusing, but if I did I’d Google the term before moving on.	Same answer as above.	None
18	Nearly everything. What's a copayment? How do you find out what qualifies?	What are hi-tech procedures?	Terms Product
19	Deductables	Premium	Terms
20	Excess	Shortfall	Terms

Participant	First confusing term	Second confusing term	Category
21	No confusion	No confusion	None
22	all of the terms, including private/ semi-private	excess	Terms
23	A lot of the terms and conditions about pre existing conditions	How to understand exactly which places to go to make sure it's covered	Terms Product
24	Far to much small print	Anything reffering one to other paragraphs, sections etc, also exclusions were not specific	Terms Product
25	There wasn't anything that I found confusing.	Again I cannot recall anything that I found confusing.	None
26	In patient	out patient expenses	Terms
27	Nothing		None
28	The terms and payouts were confusing	Also cashing in payments	Terms Product
29	hospital cover	cover	Product
30	It was clear	Nothing	None
31	None	None	None
32	Excess	Uninsurable	Terms
33	I have absolutely no idea what represents value for money	Tiers	Terms Product
34	Which hospitals were covered for what procedure	None	Product
35	Hard to get an updated policy with similar cover for cheaper. Not very clear	Waiting period can be overwhelming	Terms Product
36	Nothing I worked in health insurance previously	Na	None
27	Waiting periods and when they apply.		Terms Product
38	Hard to understand the excess in each situation	“Pre existing conditions” as so many illnesses are pre existing but have gotten worse	Terms Product
39	What is actually covered. There are so many pages in the booklets that they send you. If you got a page at then end that had your cover summarised into bullet points or a table so you could have a quick reference	My excess	Terms Product
40	Pathology	Medical and surgical appliances	Terms
41	Nothing was confusing	Nothing was confusing	None
42	excesss	pre-existing	Terms
43	Level of cover	Waiting period	Terms

Participant	First confusing term	Second confusing term	Category
			Product
44	The level of cover	Language used by various companies	Terms Product
45	Difficult to understand levels of cover and too many policy types available.	Excess payment	Terms Product
46	Nothing	None	None
47	The illnesses it covers	To many to remember	Terms Product
48	Nothing	None	None

Table 5.1 Confusing Terms in the Initial Survey

Figure 5.1 is a bar chart illustrating the percentage of participants who were confused by aspects of the health insurance domain.

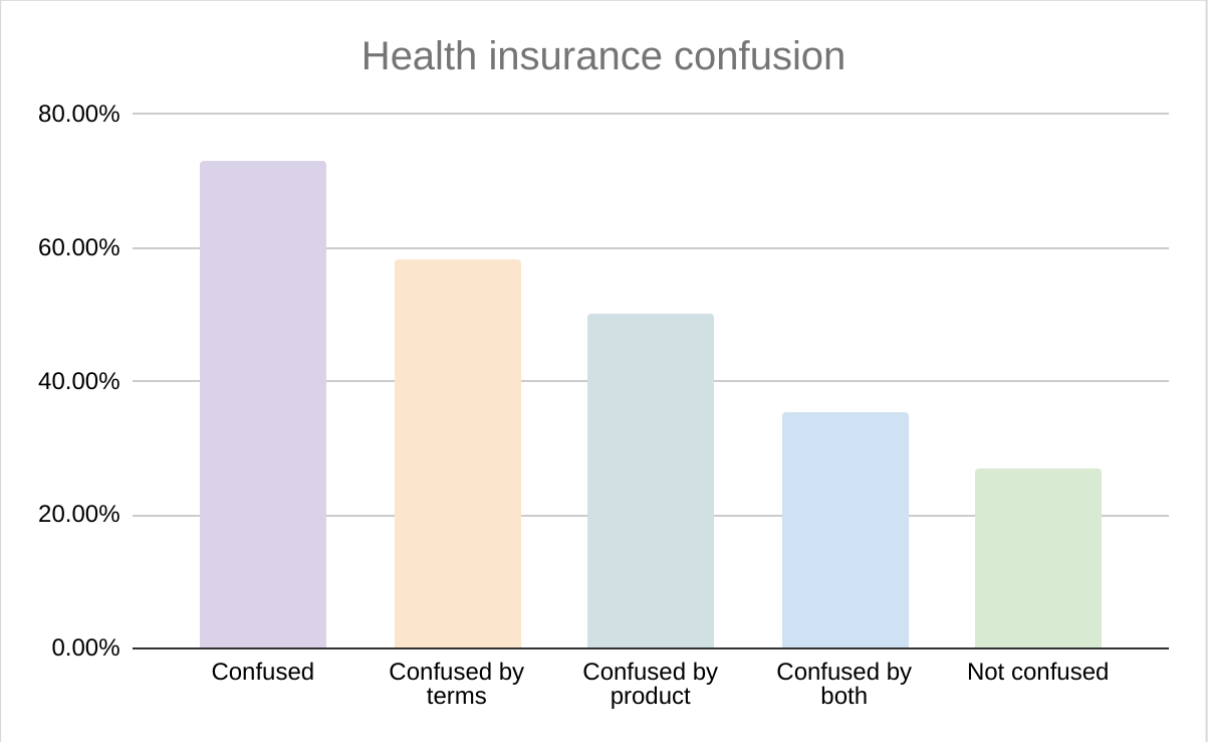


Figure 5.1 Health Insurance Confusion Chart

- 73% of respondents were confused by the domain and 27% stated that they were not confused at all.
- 58% of respondents were confused by terms such as “excess” and “surgical appliances”.
- 50% of respondents were confused by the product itself e.g. which consultants or clinics were covered by a particular product.

- 35% of respondents were confused by both the product and the terms.
- The bar showing “Confused by terms” includes participants who were confused by both the product and the terms
- The bar showing “Confused by product” includes participants who were confused by both the product and the terms

There was a broad range of confusing terms. Figure 5.2 illustrates particular terms which were confusing to more than one participant, e.g “Excess” is a term which confused 11 people.

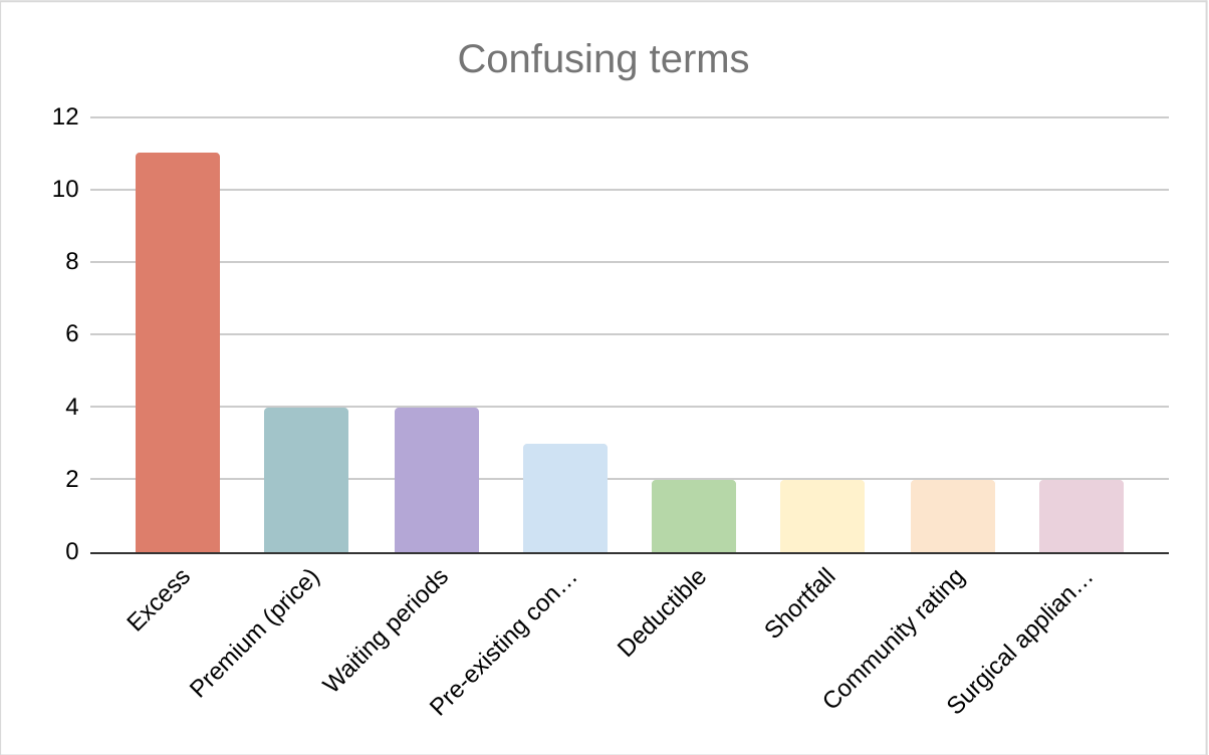


Figure 5.2 Confusing Terms

5.1.4 Discussion of the Survey

The initial survey established that there was a significant level of confusion about the domain (health insurance). 73% of participants were confused and this confusion applied to both the vocabulary and the products in the domain. The vocabulary consisted of a broad range of terms which people could not understand and the term “excess” confused the most people (eleven of the forty eight respondents).

40% of all respondents could not determine what exactly was covered on their health insurance plan. They didn’t know "*Which hospitals were covered for what procedure*" or “*which clinics were covered*” or “*which consultants were covered*”.

This indicated that we should expect a challenge when implementing a chatbot for the domain. The challenge would be twofold, concerning both the vocabulary used to describe the subject area and the participants' understanding of the products in the domain.

Following the survey, a chatbot was implemented and the two chatbot studies were conducted.

5.2 Study 1

The research involved a design based case study. A chatbot was implemented for the domain under question using the Rasa open source framework and other open source libraries. The implementation was then evaluated by participants who interrogated the chatbot to compare two representative products from the domain (two health insurance plans).

5.2.1 Objective

The objective of the study was to evaluate the chatbot design and implementation with participants who were potential purchasers of the domain products. Specifically, the following questions were addressed by the research;

1. Was the chatbot implementation able to handle complex data?
2. Did the participants understand the chatbot responses?
3. Were the participants satisfied with the usability of the chatbot?

5.2.2 Experimental Setup

The experimental setup included a chatbot, a quiz web application and participants.

Participants were chosen on the basis of having an interest in the domain. They had taken out a health insurance policy within the preceding six months or they were about to take out health insurance in the succeeding six months. They were recruited from the academic research website, Prolific, in the same manner as the initial survey participants. Only participants from Ireland were selected as the health insurance plans under research are only available in Ireland. There were eleven participants in Study 1. Ethical considerations were applied as described in 3.1 High Level Requirements for the Case Study

A chatbot was developed for the domain using the Rasa open source framework. The chatbot was then evaluated using survey questions and a quiz.

A custom web application was developed to support testing of the chatbot with participants. This quiz application was not part of the chatbot implementation itself. The web application presented comparison questions to participants while displaying the front end of the chatbot in an iframe. It provided for immediate feedback from the participants after each question. Before the participants were asked to compare the products, they were asked two open questions concerning benefits in the domain.

Two open benefit questions are asked as described in section 3.5 Question Design . They are repeated here for context.

1. “What is the most important benefit you look for in a health insurance policy when purchasing one?”
2. “There are more than 60 benefits associated with each health insurance plan. List as many as you know off the top of your head without searching the internet.”

The purpose of these questions was to identify which benefits were of significance to participants and also how well the Health Insurance Authority (HIA) taxonomy of benefits

represented the participants' language for benefits. The first benefit question was later used in a follow up comparison question to encourage the participant to explore the chatbot and identify gaps in the chatbot implementation. The follow up question was as stated as follows where "X" represents the most important benefit to the participant;
" Previously you answered that "X" was the most important benefit for you. Find out which plan is better for "X"

Participants were then given eight questions comparing two typical products. The products were two of the most popular health insurance plans from the Irish market. Every question required the participant to give an explanation for their answer. This demonstrated their understanding of both the question and the response provided by the chatbot. Figure 5.3 is an example of the screen where a participant selects their answer to a comparison question - in this case "Which plan is better for convalescence?". The participant then explains why they chose that answer.

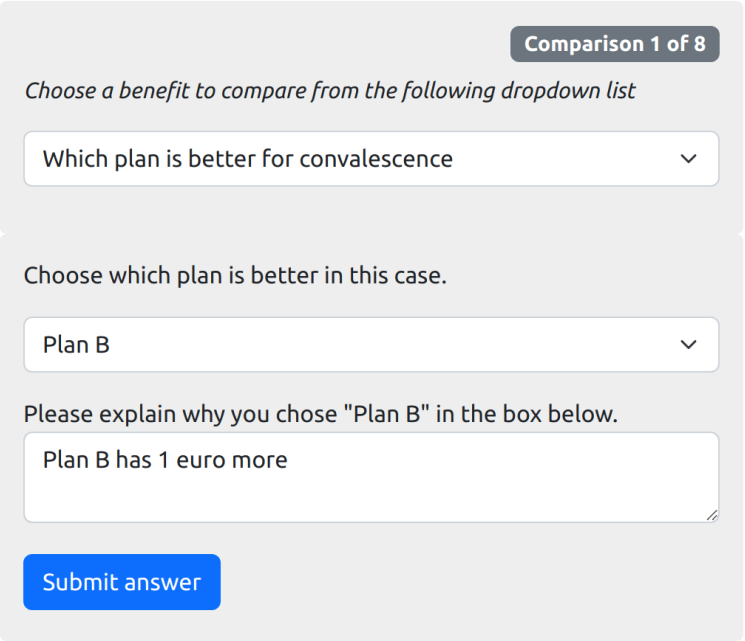


Figure 5.3 The Single Benefit Comparison Screen

Questions were then asked which compared categories of benefits rather than a single benefit. Category comparison questions had an extra step which asked the participant about their experience in navigating and understanding the category of benefits, "Explain how easy or difficult it was to compare all of the benefits from that category". Figure 5.4 shows the screen where a participant answered a category question and was required to comment on the difficulty of comparing all benefits in the chosen category.

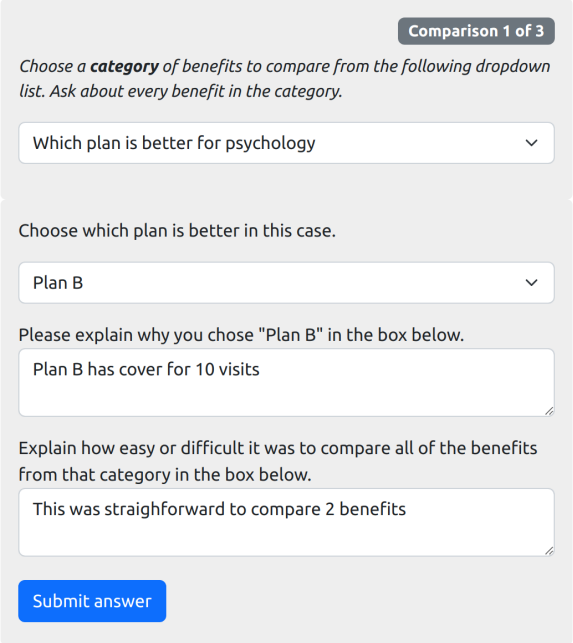


Figure 5.4 The Screen for a Category of Benefits

Feedback on the chatbot was obtained from participants by two questions immediately following the quiz as follows;

1. “What is your first impression of the chatbot?”
2. “How could the chatbot be improved?”

A System Usability Scale (SUS) questionnaire was taken by participants as the last stage of the survey. This was used as a standard measure of the usability of the chatbot.

5.2.3 Results

There were two open questions on benefits which led the participants to explore the chatbot in interesting ways.Table 5.2 shows the benefits identified by the participants as being important.

“instant access to doctors”
“hospital cover”
"good cover and a good price "
"Comprehensive cover, not basic. Urgent care and consultants visits important in case of severe illness"
"Access to digital doctor"
"The quickest access to the best medical care"
"dental cover"
"Pre Existing condition coverage"

"It's the range of hospitals and what illnesses the plan covers" "cardiac cover" "A low excess, with a broad range of hospitals" "Most cover for best price"

Table 5.2 The Most Important Benefits for Study 1

When participants were asked which health insurance benefit was of the most importance to them, some listed more than one benefit. There was some overlap between respondents with three saying "hospital cover" and two saying "doctors" and three mentioning quick "access".

- Hospital cover was expressed in different ways as "the range of hospitals", "a broad range of hospitals" and "hospital"
- "Doctors" was expressed as "instant access to doctors" and "access to digital doctor". Note that currently there are no HIA benefits or categories which are explicitly digital.
- "Access" was expressed as "access", "instant access" and "the quickest access"

Other participants listed the following as being of utmost importance to them;
"pre-existing condition coverage", "dental", "price", "what illnesses the plan covers", "the quickest access to the best medical care", "cardiac", "comprehensive , not basic", "urgent care and consultants visits important in case of severe illness", "low excess"

Others had more generic descriptions which would require further clarification before they could be mapped to benefits or benefit categories; "good cover", "medical care".

The participants were asked to list all of the health insurance benefits known to them.
The full results are in Appendix: Answers to Open Benefit Questions. Many of the benefits described by the participants did not map well to HIA categories or individual HIA benefits
Table 5.3 shows the benefits known to participants which did not match a HIA benefit

instant access online consultations Practitioner joint care mental health Osteo	Blood testing for genetic issues Parenting Digital doctor eye cover MRI	Single room Shared room travel insurance cover New born screening

Table 5.3 The Known Benefits for Study 1

The questions which compared health insurance benefits between two products (plans) were then asked. The complete results for all participants are documented in Appendix: Answers to Benefit Comparison Questions and Appendix: Answers to Category Comparison Questions. Some of the questions had deterministic answers. Participants answered 84% of

these questions correctly in the first study. Table 5.4 shows the results for the deterministic questions.

Question	Correct	Incorrect	% Correct	Note
Which plan is better for cancer accommodation support?	4	1	80%	This participant had a chatbot crash
Employee assistance	2	0	100%	
Which plan is better for overseas hospital?	3	0	100%	
Which plan is better for optical	3	1	75%	Chatbot misinterpreted one request. Another participant misinterpreted the question to apply to himself
Which plan is better for hearing?	3	0	100%	
Which plan is better for vaccinations?	5	1	83%	Chatbot misinterpreted one
Which plan is better for individual excess?	1	1	50%	Participant misunderstood the concept of excess
	21	**4**	**84%**	

Table 5.4 Answers to the Deterministic Benefit Questions for Study 1

After each question, the participant was asked to give an explanation for choosing their answer. The reasons for choosing answers are documented in the Appendix: Answers to Benefit Comparison Questions

There were 3 questions requiring the participant to compare categories of benefits;

- The first question related to a category containing 2 individual benefits e.g. the category “psychology” has the benefits of “employee assistance” and “psycho-oncology” counselling
- The second category contained 4 individual benefits
- The third question addressed the most important benefit to the participant (which the participant had specified in their response to a previous question). This could have been a single benefit or a category of benefits depending on the first answer.

Nobody completed all tasks in all of the category questions. Only a single participant completed all tasks in the category containing 4 benefits in this study. Only a single participant successfully navigated an entire category of benefits Participant comments reflected their difficulty in performing the required navigation tasks and in understanding the information presented to them. Table 5.5 lists the comments provided by participants.

“they both have pros and cons it would be difficult to figure out which is best as there is a lot of medical terms and information”
"difficult as it was convoluted"
"Not easy in this format"

"It does not directly answer regarding cancer treatment"
"Easy but did not answer my real concern"
"offers much more cover that I am looking for and would expect"

Table 5.5 Participant Comments

One participant expressed their frustration at this point, *"I start to dislike the bot very much"* Another participant offered a useful suggestion for improvement, "*It was hard. a bullet point more concise list would make it easier*".It can be seen that there is a challenge comparing more than a couple of benefit values. This challenge can only be amplified enormously when all sixty six benefits in a plan are taken into consideration. Table 5.6 is a sample participant utterance asking about a category of benefits and the chatbot's response.

Participant request	**Bot response**
Which plan is better for **public hospitals**?	A category containing these 2 individual benefits exists for **public hospital**; 1) a private room in a public hospital, 2) a semi-private room in a public hospital Type a number between 1 and 2 for a benefit from the category **public hospital**

Table 5.6 A Sample Utterance & Response for a Category Question

The full set of answers to the questions relating to categories of benefits is in the **Appendix: Answers to Category Comparison Questions**

-

The participants were given a standard System Usability Scale (SUS) test after interacting with the chatbot. Figure 5.5 presents the SUS scores for the first execution of the experiment, Study 1. The mean SUS score was 61 and the standard deviation was 20.

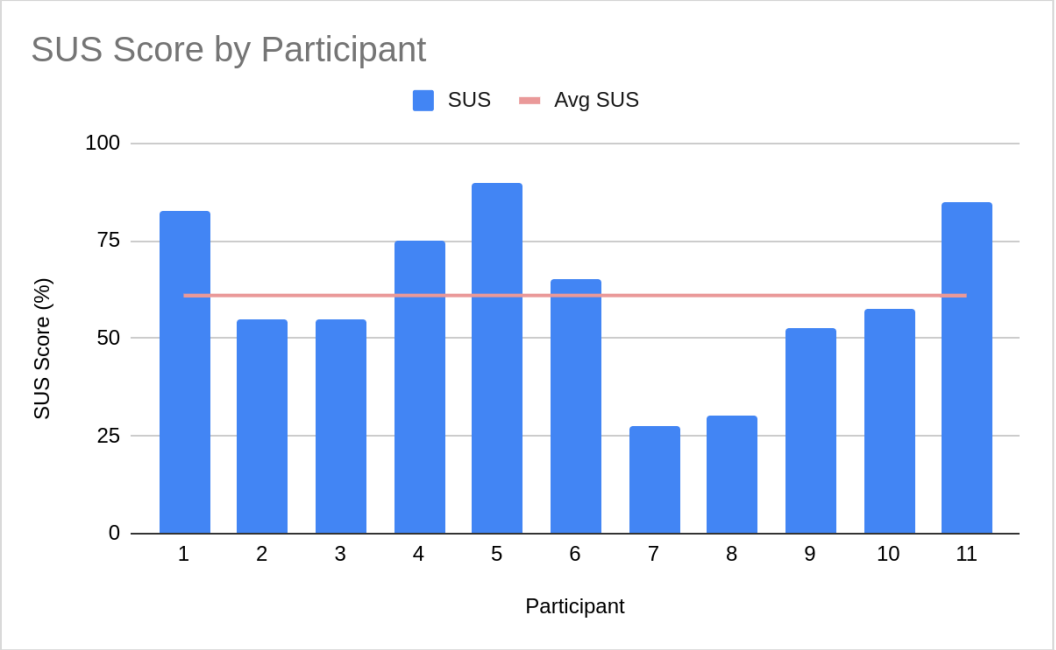


Figure 5.5 SUS Scores by Participant for Study 1

Every participant in the first study completed all goals i.e. they answered every question in the quiz. The first five questions required a single task to compare a single health insurance benefit and the last three questions required multiple tasks to navigate a category of benefits. Only one participant completed all the tasks in the final goal. The category contained four individual benefits,

The time taken to complete the goals of the chatbot experiment was measured. Most participants completed all of the goals within the suggested time of twenty minutes. There was one outliner who took thirty six minutes to complete the quiz. Figure 5.6 presents the goal times for the first execution of the experiment, Study 1. The mean of goal times was 12 minutes and the standard deviation was 9.

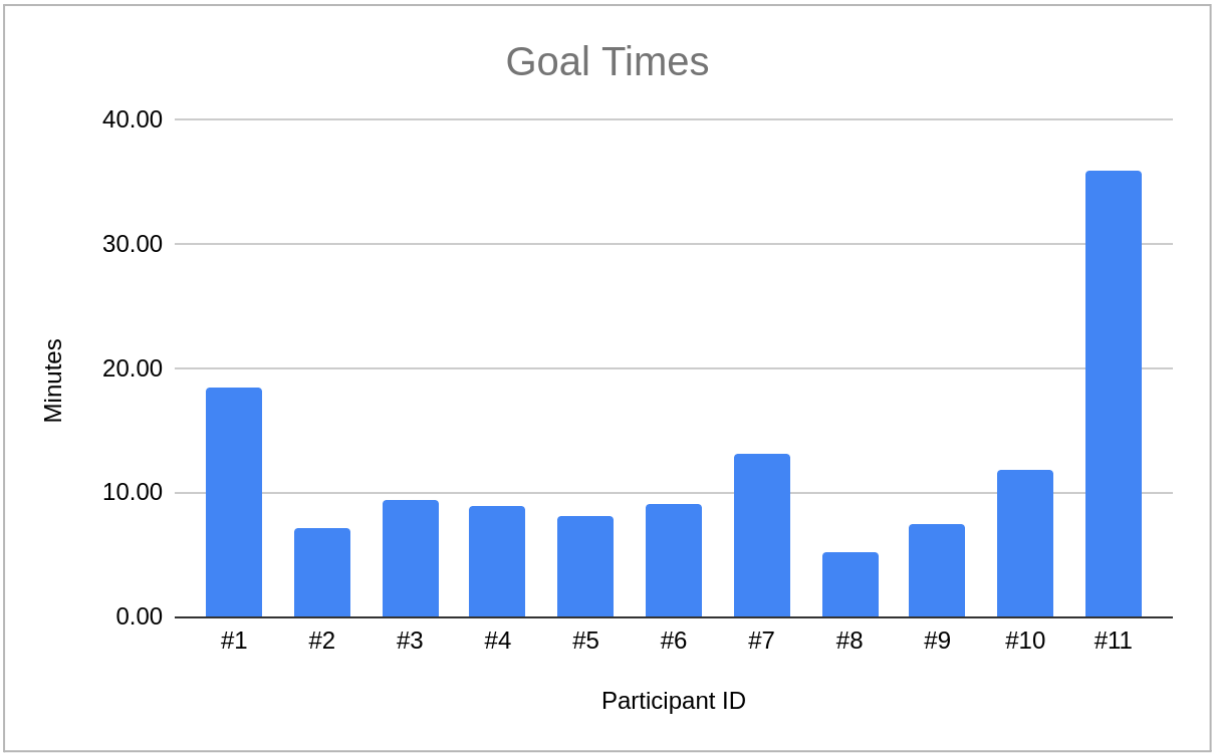


Figure 5.6 Goal Times for Study 1

Some participant requests were misunderstood by the chatbot. Figure 5.7 charts the percentage of requests misunderstood by the chatbot for each participant in Study 1. The mean was 28% and the standard deviation was 17%.

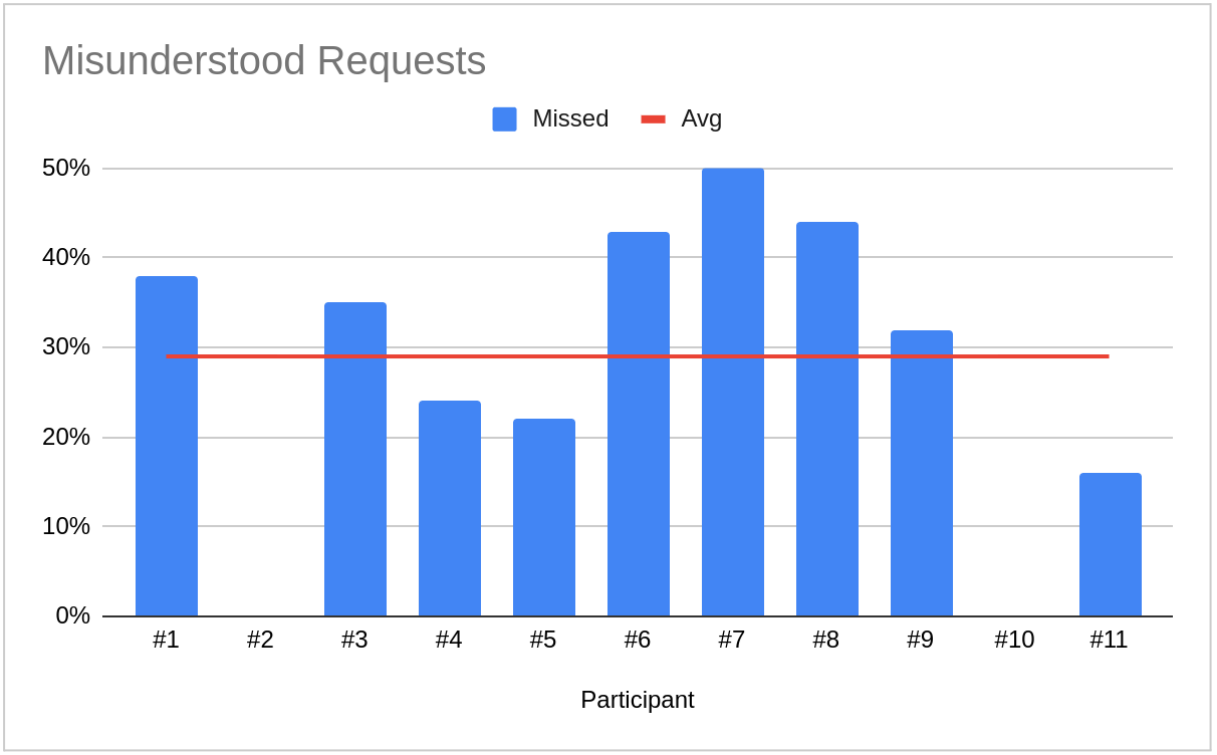


Figure 5.7 Misunderstood Requests for Study 1

Immediately after the participants finished the quiz they were asked for their feedback on the chatbot. Participants were asked to give their first impression of the chatbot and Table 5.7 outlines their answers.

“It is a convenient way to source information”
“some of it is fine, other parts the information is not easy to follow”
"very generic answers, does not understand a lot of questions ",
"Good"
"It was easy to use and was easy to compare plans"
"Relatively easy to use"
"poor, it does not understand my questions, or only partially. I find it frustrating"
“Not good at all"
"Basic chat bot with no fancy icons"
"its ok"
"Good, but can't handle non standard terms"
"Very easy to use"
"It seems to struggle with any follow up questions and just lists the benefits as they would appear on a statement of benefits",
"I didn't think it was that helpful. The best it could do was search the table to benefits for the two plans and show me a chunk of text that I still had to sort through and try to understand. I feel like it slowed down the process and I'd be quicker comparing the two plans myself."
"they understood most terminology",
"1st impression is this is difficult to phrase the things I need to find out. I'm 45 so health care is important, but financial cost is also very important. I think if I was older, I'd definitely need to be chatting through more specifics on it. But it definitely helped get me thinking on what cover I need and allowed me to get an understanding of both plans.",

Table 5.7 First Impressions of Participants for Study 1

The participants were then asked to suggest improvements to the chatbot. Table 5.8 lists their suggested improvements.

“simplified medical terms” “make the information more simple” "give more information" "More understanding. Ie when I input comprehensive it could have given me some options. I dont understand that, would you like to search for family or individual plans?",

"N/A" "Difficult to get specific answers" "it definitely needs a much better vocabulary. When I ask about maternity it did not understand. I asked for psycho therapy and it did not understand. A bot needs to have a very broad vocabulary and even with typos" "But a human in there. It was typical chatbot stuff, poor scrambled information, couldn't answer questions etc" "Not sure its a pretty complex topic and i am not sure i would chose base on the information it gives to me", "dont know" "Introduce alternate words for the same processes"

Table 5.8 Suggested Improvements for Study 1

The chatbot made errors and these are documented in the ***Appendix: Answer data & chatbot messages***. The participant utterances were all recorded and these are documented in the ***Appendix: Answer data & chatbot messages***. They were used to refine the chatbot after the first iteration as described in 6.2 How Study 1 was Used to Modify the Chatbot

5.2.4 Discussion of Study 1

The confusion discovered in Section 5.1 Initial Survey of the Domain continued in Study 1. The benefit comparison questions compared a single benefit between the two health insurance plans. The most confusing term in the initial survey, “excess” also caused confusion here for the participants. Two participants in the first execution of the chatbot experiment chose to answer the question concerning “individual excess”. One thought that a higher value of excess was better.

Seven participants answered the question about “family excess” and provided three different answers. This demonstrates confusion among participants about this benefit. The provider statements are partly to blame for this since one insurer states the benefit on a per person basis and the other on a family basis. There is a lack of consistency. Plan A simply stated “€250” and Plan B had “€100 excess per person”. This means that for a family of two plan B would have less excess but for any larger family plan A would be better. One person showed understanding of this concept , "It is per person so the value is greater if there are more than 2 people in the family" . The problem with this benefit is one of evaluation. It was difficult for participants to compare the values since they were effectively over different parameters. The text of the benefit value is free text provided by each insurer in their own words. The chatbot clearly indicated the values but the way those values were defined by the insurers made them difficult to compare.

The answers by participants showed a disconnect between their language and the language used by the Health Insurance Authority (HIA). There was also a disconnect between the mental model of benefits as expressed by participants and the taxonomy published by the HIA. It became clear that the HIA language is not in sync with the public. Many of the participants asked about “hospital cover” whereas the HIA lists the category as “inpatient”.

The HIA taxonomy has no categories for “cancer”, “cardiac” or “mental health”. In order to find all of the benefits related to cardiac a participant would need to scan all sixty six benefits in the HIA taxonomy to determine which are relevant. The HIA benefit names can be very long winded such as “The Blackrock Clinic, the Mater Private and the Beacon Hospital procedures other than Cardiac and Special”. In addition, that particular benefit mentions cardiac but actually excludes it. This adds further cognitive load on participants who may be searching for cardiac benefits to firstly evaluate such a long benefit name and then, secondly, to mentally remove it from their list.

Table 5.9 shows an example of the benefits in a category (inpatient) in the HIA taxonomy. An explanation of the confusion caused by this category follows the table.

semi private room in public hospital
private room in public hospital
semi-private room in private hospital
private hospital private room
The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures
The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Special Procedures
The Blackrock Clinic, the Mater Private and the Beacon Hospital procedures other than Cardiac and Special
Day Case Private Hospitals
Day Case The Blackrock Clinic, the Mater Private and the Beacon Hospital

Table 5.9 HIA Benefit Names for the “Inpatient” Category

The category caused misunderstanding on multiple levels. To begin with, the title of the category is “inpatient” which is not a term which participants used - they asked instead about “hospitals”. Then “Day Case” was not a term understood by participants. The benefit labelled “The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures” relates to cardiac. However, the rigid nature of the HIA taxonomy means that this benefit can only appear in a single category, “inpatient”. The benefit does not appear in a “cardiac” category which was identified by participants as an area of interest.

The answers which participants gave to the open benefit questions were a rich source of their language and hence it was decided to use these to improve the chatbot. This is discussed in the following chapter, 6.2 How Study 1 Was Used to Modify the Chatbot

The benefit comparison questions compared a single benefit between the two health insurance plans. The most confusing term in the initial survey, “excess” also caused confusion here for the participants. Two participants in the first execution of the chatbot experiment chose to answer the question concerning “individual excess”. One thought that a higher value of excess was better. For the second execution, an explainer note was added which was displayed to participants when they were shown the value of individual excess for both plans. The note stated “*please note that less excess means more money back for you when you claim”* . All three participants in the second execution of the experiment gave the correct answer for “individual excess”.

Seven participants answered the question about “family excess” and provided three different answers. This demonstrates confusion among participants about this benefit. The provider statements are partly to blame for this since one insurer states the benefit on a per person basis and the other on a family basis. There is a lack of consistency. Plan A simply stated “€250” and Plan B had “€100 excess per person”. This means that for a family of two plan B would have less excess but for any larger family plan A would be better. One person showed understanding of this concept , "*It is per person so the value is greater if there are more than 2 people in the family*" . The problem with this benefit is one of evaluation. It was difficult for participants to compare the values since they were effectively over different parameters. The text of the benefit value is free text provided by each insurer in their own words. The chatbot clearly indicated the values but the way those values were defined by the insurers made them difficult to compare.

The category comparison questions were not answered fully by participants. Only a single participant successfully navigated an entire category of benefits in the first execution of the experiment. Participant comments reflected their difficulty in performing the required tasks and in understanding the information presented to them. One participant expressed their frustration at this point, *"I start to dislike the bot very much".* Another participant offered a useful suggestion for improvement, “*It was hard. a bullet point more concise list would make it easier*”.It can be seen that there is a challenge comparing more than a couple of benefit values. This challenge can only be amplified enormously when all sixty six benefits in a plan are taken into consideration.

Participants were asked to give a reason for the answers which they selected. The reasons indicated their level of understanding of the domain.The reasons also revealed the difficult nature of the language used by the providers. Participants stated , "*Both are overly complicated and don't answer the question*" and “*I don’t like the idea of (up to) in plan A as it is vague*"

Providers expressed the same benefit value using different parameters so that the values were not directly comparable between plans. For example, for the benefit “Family excess” one insurer stated “€250” while the other stated “€100 excess per person”. This caused participants to express confusion in the reasons they gave for their answers, for example; “"*it is not clear if plan A is per family or per person*"

The System Usability Scale (SUS) was scored by all participants.The average SUS score was 61 in the Study 1. This is below the 50th percentile score of 68.

All participants completed all goals by answering every question in the quiz. The final goal for each participant was to compare a category consisting of four individual benefits. Only one participant in the first execution completed all of the tasks in this goal.

Category comparison becomes more difficult the more benefit items are contained in a given category. When comparing psychology benefits a participant stated "*It was quite easy as there were only two".* However, when comparing the category of cancer a participant highlighted the difficulty, saying it was *“Difficult. Some of the cardiac related care such as screening and scans was buried in the radiology category”.*

Most participants finished the quiz in the expected time of 20 minutes.

Thirty two chatbot errors could be attributed to a single dialogue turn of a participant. Many of the misunderstood requests were dialogue breakdowns of the “specific” type i.e where the participant asked for something in particular which the chatbot had not been trained on. These were straightforward to address by adding a new *alias* or *benefit_tag* to the chatbot knowledge base. Thirteen of the chatbot errors were multi-turn i.e. they involved a sequence of participant utterances.

During the error analysis many new participant utterance expressions were identified. Table 5.10 highlights the variety of different ways in which participants expressed their intents. For the same entities they used very different words to express their intent. Some expressions were common, for example nine out of eleven people simply stated the benefit name with no surrounding words. The utterances are presented in the table below using the syntax for Rasa NLU. Entities are enclosed in angle brackets and the entity type is enclosed in curly brackets.

Intent	**Expression**	**Count of participants**
Ask about cover	**which plan is better for** [optical policies](benefit)?	2
	can you tell me about [vaccinations](benefit_tag) ?	1
	which plan is better for [psychology](benefit_tag) ?	7
	Which plan has [hearing](benefit_tag) cover	1
	[instant access to a doctor](benefit_tag) ?	9
	which plan will give me [access to a gp](benefit_tag) ?	1
	Which plan is good for [cover](benefit_tag) and [a	1
	good price](benefit_tag) ?	1
	Tell me about [price](benefit_tag)	3
	Tell me about [plan A](plan)	1
	What is the [accommodation](benefit_tag) in [plan	2
	A](plan) ?	1
	What is covered?	1
	How much is covered?	1
	how much is covered per night?	1
	what plans are available?	1
	Which covers [maternity](benefit_tag)?	1
	how much is [excess](benefit_tag) for [plan a](plan)	
	and [b](plan)?	1
	are [accident and emergency](benefit_tag) **covered**?	
	is [psycho therapy](benefit_tag) **covered in** plan a[plan]	
	and [b](plan)	1
	is [psychology](benefit_tag) **covered**	
	how is [cancer](benefit_tag) **covered**	1
	cover for [cancer](benefit_tag)	1
	Does [Plan A](plan) **have a** [private	1
	room](benefit_tag)?	1
	what [excess](benefit_tag) **is better**?	1
	are [both plans](plan) **the same**?	1

Intent	Expression	Count of participants
	better for [fmily excess](benefit_tag) **plan is better for** [accident and emergency](benefit_tag)? **is there a** [difference](difference_in_cover) **between** Plan A[plan] or [Plan B](plan) **Which plan is better for** Its the [range of hospitals](benefit_tag) **and** what [illnesses](benefit_tag) the plan covers? **Which plan has a better** range of hospitals? **What plan gives** [free glasses](benefit_tag)? **What** [eye care](benefit_tag) **can I get** **What** [hearing](benefit_tag) benefit_tags are there? **What** [excess for one person](benefit_tag) **is there** [Excess] (benefit_tag) **cover amounts**	1 1 1 1 1 1 1 1 1
	Notes: 9 out of 11 people simply uttered the name of the benefit with no surrounding words e.g. “psychology”	
Elaborate on cover	**Is this** [per person](aspect_of_cover) or [individual](aspect_of_cover) ? **Is there any other difference between plans for** [overseas hospital](benefit_tag)? **Is** [plan A](plan) [per person](aspect_of_cover) **Why such a small** [€1 difference](difference_in_cover)? **Which is better** ? **So** [Plan B](plan) **has no** [excess](benefit_tag)? **so** [vaccination](benefit_tag) **are no covered?** **does** [plan b](plan) **have a** [limit](aspect_of_cover)? **does** [plan b](plan) h**ave a** [limit](aspect_of_cover)it on there [excess](benefit_tag)? **Any benefits for a** [women](insured_person) **Is there any other** [cancer supports](benefit_tag)?	1 1 1 1 1 1 1 1 1
	Notes: “Which is better” was asked as soon as the participant read the response to an FAQ question about a benefit “So Plan B has no excess?” was asked as a follow on question about cover for a private room “So Plan B has no excess?” was asked as a follow on question about cover for a private room “so vaccination are no covered” was asked as confirmation once the vaccination benefit was stated “does plan b have a limit?” was asked as clarification to a response about family excess “Any benefits for a women?” Was asked for further info on gender category	
Ask about FAQ	[terms](knowledge_item) **List the** [terms](knowledge_item) **List all** [benefit categories](knowledge_item)	1 1 1
Ask faq type	**what is** [family excess](benefit_tag)	1

Intent	Expression	Count of participants
question		
Select a benefit from a category	[2][number_pbi)	9

Table 5.10 Participant Expressions for Intents

Benefit names from the HIA can be very wordy and this leads easily to mismatches when using Levenshtein distance to fuzzy match. For example a participant who asked for “Inpatient accommodation” got a response from the bot relevant to “cancer accommodation” since this was a better match than the “inpatient” category of benefits. It is more likely that the participant was interested in hospital cover.

Entities in the domain proved to be quite complex, particularly the line item of an insurance product - the plan benefit item (*pbi*). For simpler domains like restaurant ordering, entities tend to be more straightforward and also more semantically similar. For example, “pepperoni” and “mushroom” are pizza toppings which are both types of food.

The variation in plan benefit items in the HIA taxonomy made it difficult for entity extraction. Word count varied widely from titles as short as “Home births” to those as long as "A Multi-Occupancy or Semi-Private room in a Public Hospital and Day Case". This variation can be seen in the following histogram which charts the count of plan benefit items (PBIs) with a given number of words in them. Five plan benefit items have 12 or more words in them.

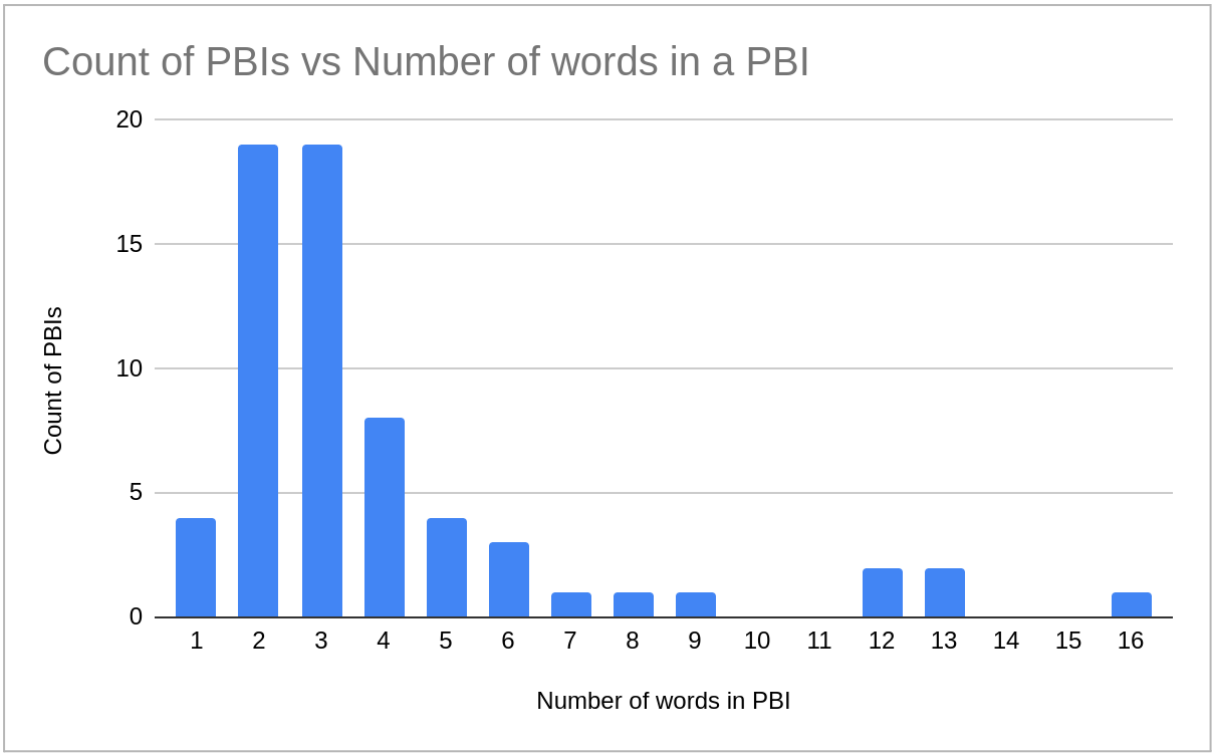


Figure 5.8 Count of PBIs vs Number of words in a PBI

Another challenge was the lexical similarity between *pbis* which made entity extraction difficult. In many cases there was too much lexical similarity between *pbis* as in the following list;

- a private room in a private hospital
- a private room in a public hospital
- a semi-private room in a private hospital
- a semi-private room in a public hospital

Semantic similarity was also a concern. There was too little semantic similarity between pbis, e.g. between "24 Hour Telephone Assistance overseas" and "Family excess". This is very unlike the pizza ordering domain where the entities “pepperoni” and “mushroom” are semantically similar as food types.

This meant that the supervised training examples had to include all of the plan benefit items and not just a sample of them. Rasa provided a confidence value for each entity it extracted based on comparison with the examples with which it had been trained. Since some of the entities could have very many words the approach was taken to reduce the load on the user and to support matching the user’s input with only a few words typed. This meant that the confidence threshold for filing the appropriate slot with this entity was lowered to 0.65. Therefore mistakes could be made by the bot. The chatbot stated the entity name in the response to the user to ensure the user knew exactly which *pbi* was being stated. For example if the user simply typed “a private room” the bot might respond with “Concerning a **private room in a private hospital**, plan A has this cover…..”. In addition the response included links (buttons) to the other possible *pbis* to enable the user to choose the correct *pbi* in the event where the bot was mistaken. This dialogue repair strategy used elements of the options strategy [27].

5.3 Summary

This chapter details the results of the survey and the first execution of the experimental study, Study 1. The initial survey establishes that there is a problem with confusion for users of the domain. It describes the extent of the problem in terms of the language and products in the domain. Then a chatbot for the domain is designed and implemented and the first study, Study 1, evaluates how well the chatbot worked to help participants understand the domain. It establishes that the mental model of participants does not match the domain model as documented by the Health Insurance Authority(HIA) and identifies the discrepancies. Study 1 finds that participants understand the chatbot for simple benefit comparisons. However, many user requests are misunderstood by the chatbot and the usability of the chatbot is below standard.

Chapter 6 : Study 2

6.1 Objective

The three objectives presented for Study 1 also apply here, namely;

1. Was the chatbot implementation able to handle complex data?
2. Did the participants understand the chatbot responses?
3. Were the participants satisfied with the usability of the chatbot?

There is an additional objective to use the results of Study 1 to improve the chatbot. This is detailed in the next section.

6.2 How Study 1 was Used to Modify the Chatbot

The ontology (knowledge base) developed for the initial chatbot implementation was broader than the taxonomy provided by the Health Insurance Authority (HIA). Benefits were not restricted to a single hierarchy in the ontology. They could instead be mapped to multiple categories (*benefit_tags*). In addition each *benefit_tag* could have multiple *aliases* to cater for the variety in the participants' language.

After Study 1, *aliases* were added to the chatbot in order to map the participant terms which had been discovered by the study. In addition, certain words such as "care" and "policy" were used by participants in ways that interfered with entity extraction by the Rasa NLU. A new entity type, *junk*, was added to Rasa to remove these from consideration.

Seventy four new participant terms were discovered. Table 6.1 lists the mapping of the terms which was performed in order to improve the chatbot.

Term discovered in Study 1	**Alias**	**Benefit_tag**	**Action or Comment**
instant access to doctors	instant access	Instant_access	A new benefit_tag and alias were created.
hospital cover	hospital stay	inpatient	Existed already
good cover	N/A	N/A	Too generic a term for an alias, was handled as an example in the top level intent
good price	price	adult_price	Existed already
Comprehensive cover	N/A	N/A	This was added as an intent training example
Urgent care	urgent care	a and e	A new alias was created
consultants visits	N/A	consultant	Existed already
Access to digital doctor	N/A	digital	A new benefit_tag was created

Term discovered in Study 1	Alias	Benefit_tag	Action or Comment
The quickest access to the best medical care	N/A	instant access	A new benefit_tag was created
dental cover	N/A	dental	Existed already
Pre Existing condition coverage	N/A	pre-existing condition	This required a new custom explanation to direct the participant to ask their insurer
range of hospitals	hospital stay	inpatient	Existed already
illnesses	N/A	N/A	This was handled by an intent training example
cardiac cover	N/A	cardiac	Existed already
A low excess,	excess	N/A	Existed already
broad range of hospitals	N/A	hospital stay	Existed already
instant access	instant access	instant_access	A new benefit_tag was created
online consultations	online	digital	A new benefit_tag was created
access to private rooms	N/A	private room	Existed already
dental	N/A	dental	Existed already
hospital	hospital stay	inpatient	Existed already
GP	N/A	gp	Existed already
prescription	prescription	prescription_costs	A new alias was created
practitioner	general practitioner	gp	Existed already . Not clear if this should refer to alll practitioners which would be very broad
cardiac	N/A	cardiac	Existed already
fertility	N/A	fertility	Existed already
cancer care	N/A	cancer	Existed already. "Care" should be added as a junk entity
joint care	N/A	joints	A new benefit_tag was created to encompass osteopaths and chiropractors
mental health	mental health	psychology	Existed already
Dietician	N/A	dietician	Existed already
Physio	N/A	physio	Existed already
Osteo	osteo	joints	A new benefit_tag was created to encompass osteopaths and chiropractors
Cancer screening	N/A	cancer	Existed already.
maternity benefits	N/A	maternity	Existed already. "Benefits" was added as

Term discovered in Study 1	Alias	Benefit_tag	Action or Comment
			a junk entity
new born screening	new born	post natal	A new alias was created. Could also add reference to the "heel prick" screening which is free and done for all newborns
Blood testing for genetic issues	blood testing	insurer_benefit	Create a new placeholder benefit which directs participants to ask their insurer for more information
Counselling	counselling	psychology	Existed already
Dental	N/A	dental	Existed already
Optical	N/A	optical	Existed already
Podiatrist	podiatry	other_day_to_day_practitioners	Existed already
Acupuncture	N/A	acupuncture	Create new benefit_tag
Chiropractor	N/A	chiropractor	Existed already
Parenting	N/A	parenting	Create new benefit_tag which points to parent_with_child, psychology
Digital doctor	N/A	digital	A new benefit_tag was created
Dentist visits	dentist	dental	Create an alias for "dentist". Add "visit" as a junk entity
GP visits	N/A	gp	Existed already. Add "visit" as a junk entity
Consultant Fees	N/A	consultant	Existed already
Physio visits	N/A	physio	Existed already
dentistry	dentist	dental	Create an alias for "dentist".
cancer care	N/A	cancer	Existed already.
maternity	N/A	maternity	Existed already. "Benefits" was added as a junk entity
physio therapy	N/A	physio	Existed already
psycho therapy	psycho therapy	psychology	New alias
ophtalmology	ophthalmology	optical	New alias
cardiac cover	N/A	cardiac	Existed already
eye cover	eye	optical	Existed already
emergency cover	N/A	emergency	Existed already. Add "cover" as a junk entity
single room	single room	private room	New alias
shared room	shared room	semi-private room	New alias
family cover	N/A	family	This requires a new custom explanation to direct the participant to ask their insurer

Term discovered in Study 1	Alias	Benefit_tag	Action or Comment
cancer cover	N/A	cancer	Existed already. Add "cover" as a junk entity
orthopedic cover	orthopedic	insurer_benefit	This requires a new custom explanation to refer participant to hospital and consultant cover at their insurer
hybrid theatre cover	hybrid	insurer_benefit	This requires a new custom explanation to refer participant to high tech hospitals and their insurer for more details
robotic cover	robot	insurer_benefit	This requires a new custom explanation to refer participant to high tech hospitals and their insurer for more details
specialty cardiac cover	N/A	cardiac	Existed already. Add "specialty" and "cover" as junk entities
dental cover	N/A	dental	Existed already
travel insurance cover	travel insurance	insurer_benefit	This requires a new custom explanation to refer participant to ask their insurer for more details
GP cover	N/A	gp	Existed already. Add "visit" as a junk entity
Emergency dental cover	N/A	emergency dental care	Existed already
Consultant fees	N/A	consultant	Existed already
Cancer treatment	N/A	cancer	Existed already. "Care" was added as a junk entity
MRI	MRI	radiology	New alias
Maternity hospital cover	Maternity hospital	hospital_3	New alias
Private room in a hospital	Private room in a hospital	private room	New alias

Table 6.1 Mapping of Participant Terms Discovered in Study 1

The table illustrates how a rich list of *benefit_tags* and *aliases* was developed to improve the ontology. A few participant terms such as “good cover” were too ambiguous and could not be mapped.

The answers to the category comparison questions were used to further modify the chatbot. The category presentation and navigation were improved before the second execution of the chatbot experiment as follows;

- Better category navigation was implemented using buttons. This required a new frontend chat widget to support this functionality.
- A clearer statement of categories was made by the bot, for example, “Cancer is a category containing 4 individual benefits”

Figure 6.1 illustrates the improvements made to category navigation using a screenshot of the navigation buttons for the category of benefits related to "cancer".

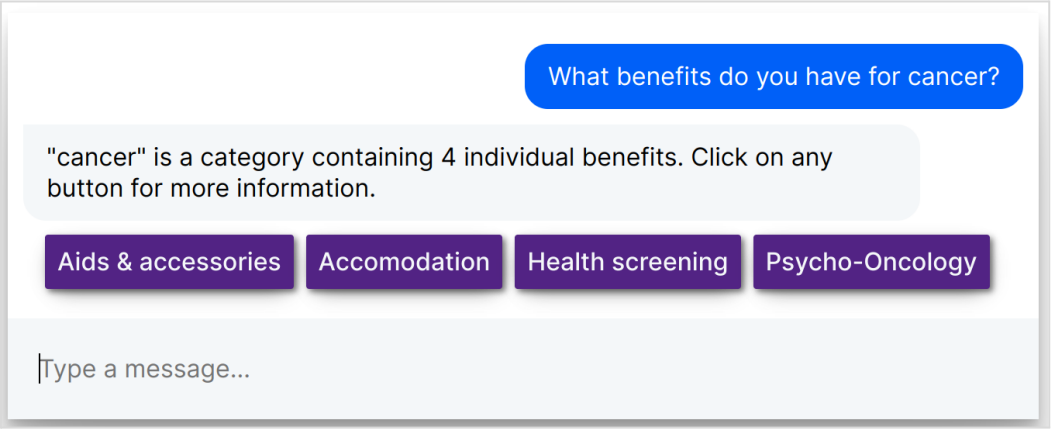


Figure 6.1 Category Navigation Using Buttons

Participants provided a reason for each answer they gave. These reasons were used to modify the chatbot as they were good indicators of areas where the bot struggled and needed improving. One participant stated "*Again the bot did not answer directly*" which highlighted a case where the incorrect plan benefit item (*pbi*) was extracted. The chatbot was improved with a better NLU training example to correct this.

The chatbot message logs were analysed and many improvements were identified for the second iteration of development. Table 6.2 lists the analysis and modifications which were made before the second execution of the study experiment.

#	**Analysis & Modifications**
1	"policies" is a common word which has a negative effect on matching when extracted along with the meaningful word, "optical". "Policies" could be extracted as a new entity type, named *junk* which would leave "optical" as the *benefit_tag* entity. This would make a better fuzzy match.
2	This was a typo from the participant
3	"A gp" was not matched as a *pbi* entity but would have been matched as a *benefit_tag* entity. Merging both entity types into a single *benefit* entity could solbe this
4	Intent was classified as *ask_faq_question* instead of *ask_pbi_or_benefit_tag*. NLU needs better training examples with comparative,superlative and contrasting phrases e.g. which plan is "cheaper", "cheapest", "better","best", "different to"
5	Entity was extracted as "and a good price". An alias of "good price" would have matched this correctly to "price". NLU examples containing multiple benefit_tag entities could also help since this would have stripped out the word "and"
6	Wrong intent was predicted. Need NLU training examples for the intent *ask_pbi_benefit_tag* such as "what is the [accomodation](benefit_tag) in [plan A](plan)?

#	Analysis & Modifications
7	“Artificial eye” was extracted as an entity. Add this to the ontology as an alias for optical
8	“Inpatient accomodation” was extracted as an entity. Add this to the ontology as an alias for *hospitals*
9	“semi private room” was extracted as two separate entities. “Private room” was the one used by Rasa to fill the *benefit_tag_slot* which is a text slot. Investigate if a list slot would be better. In this scenario the action would have to determine that the two benefit_tags “semi” and “private room” belonged together. Alternatively, add an alias of “semi private” to the ontology
10	“Comprehensive” was extracted as a benefit_tag entity and no match was found. This is a very generic question about benefits. Need to add more NLU examples to the existing intent *state_interest_in_benefits* which handles generic queries and statements of interest in overall plans
11	Add “urgent care” as an alias for “accident and emergency”
12	“Health insurance” needs to be added as an example for the generic intent *state_interest_in_benefits*
13	“plan a” resulted in “a” extracted as an entity. Renaming sample plans to Alpha and Beta should address this
14	Add “digital doctor” as an alias
15	Wrong intent was predicted. Remove pbi examples from NLU training for intent *ask_faq_question*
16	Wrong intent was predicted. Remove pbi examples from NLU training for intent *ask_faq_question*
17	participant wanted “The quickest access to the best medical care” . What does this mean as a benefit? It certainly includes “accident and emergency” and probably should include private hospitals since some like the Mater Cork have private ER rooms with much shorter wait times
18	Entity was extracted as a *pbi* but would have been found in the ontology as a *benefit_tag* .Merging entities pbi and benefit_tag should resolve this
19	“individual excess” not matched with enough confidence as a benefit_tag. Merge entities *pbi* and *benefit_tag* to improve matching in ontology
20	“what plans are available?” is predicted as intent ask_faq_question. The generic intent *state_interest_in_benefits* can be developed further to handle this
21	

#	Analysis & Modifications
	"what is covered?" is predicted as intent ask_faq_question. This looks more like the intent *ask_pbi_or_benefit_tag* .That intent can be developed to handle cases like this where there is no *benefit_tag* entity
22	"plan a and b" was extracted as the benefit tag and incorrectly matched with "a and e". Rename the two sample plans from "a and b" to "alpha and beta" to avoid confusion with simple terms
23	"psycho therapy" needs to be created as an alias
24	"how is cancer covered?" is predicted as ask_faq_question. Need better NLU training examples for intent *ask_pbi_or_benefit_tag* to cover surrounding words
25	Need better NLU training examples for intent *ask_pbi_or_benefit_tag* to cover surrounding words and more balanced examples for intent *ask_faq_question*
26	"Pre existing condition" should be added to ontology
27	"Are both plans the same?" same as intent *ask_pbi_or_benefit_tag* with no entities of type *benefit_tag* in the participant's utterance
28	"Range of hospitals" was extracted as a pbi. "Range of" could be added as a *junk* entity
29	"Which plan is better for Its the range of hospitals and what illnesses the plan covers?" Need to allow lists of benefit_tag entities in a single utterance and custom action code to handle this
30	"Free glasses" can be added as an alias of "optical"
31	"What eye care can I get" wrongly predicted as intent *state_interest_in_benefits*. Need better examples for intent *ask_pbi_or_benefit_tag*
32	wrongly predicted as intent *state_interest_in_benefits*. Need better examples for intent *ask_pbi_or_benefit_tag*

Table 6.2 Chatbot Error Analysis & Modifications

Table 6.3 lists errors which occurred over a sequence of turns (multi-turn) during Study 1 along with the modifications made to the chatbot before Study 2 was conducted.

Participant	Sequence	Analysis & improvements
#1	first	Policies need better training on stories with errors
	second	Create a new intent *ask_why_so* to answer additional questions on the current benefit
	third	Merge entities *pbi* and *benefit_tag* so matching can occur across all

#3	first	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*
	second	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*
#5	first	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*
	second	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*
#6	first	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*
#7	first	Create a new intent *ask_why_so* to answer additional questions on the current faq item . In this case pbi_text_slot will be empty but faq_term_slot will not
#8	first	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*
#9	first	Create a new intent *ask_why_so* to answer additional questions on the current *benefit_tag*
#11	first	Create a new intent *ask_why_so* to answer additional questions on the current *pbi*

Table 6.3 Analysis of Mult-turn Chatbot Errors

The utterances participants used to express their intent were mined for new ways of asking for the same things. The newly discovered intent expressions were then used in training the NLU.
Participant utterances were mined for new ways of asking for the same things. The newly discovered intent expressions could then be used in training the NLU.

The Table 26 Participant Expressions of Intent highlights the variety of different ways in which participants expressed their intents. For the same entities they used very different words to express their intent. Some expressions were common, for example nine out of eleven people simply stated the benefit name with no surrounding words. Benefit names from the HIA can be very wordy and this leads easily to mismatches when using Levenshtein distance to fuzzy match. For example a participant who asked for “Inpatient accommodation” got a response from the bot relevant to “cancer accommodation” since this was a better match than the “inpatient” category of benefits. It is more likely that the participant was interested in hospital cover.

In summary, the analysis of the chatbot errors and the participant utterances drove improvements to modify the chatbot so that it could better understand the language used by participants. In particular, the chatbot could better detect the different intent expressions and the varied vocabulary participants used to express benefits. Common words used by

participants such as “policy” and “cover” were extracted as a new entity type, *junk*, to prevent them being added to other entities such as *benefit_tags* or *pbi*s.

6.3 Experimental Setup

The experimental setup for Study 2 was identical to that of Study 1 except for the number of participants. Six participants were recruited for Study 2. The participants were drawn from the same pool as that which was used for Study 1 but the participants of Study 1 were excluded. The number of participants was less in Study 2 than in Study 1 as the available group of suitable users dwindled and it took longer to sign up new participants within the period of the study. The minimum number of users recommended by Nielsen for a usability study is five and this criterion was met for Study 2 [62].

6.4 Results

The first two questions posed to participants related to benefits and were of an open nature. The results are in Appendix: Answers to Open Benefit Questions. Some of the benefits did not match the HIA benefits or benefits discovered in Study 1. These are documented in Table 6.4 and Table 6.5.

"best value with a combination of day-to-day benefits and cover in hospitals."
"hospital stays and operations covered"
"Most cover for best price"
"everyday health expenses"
"Sporting injury cover"
"Consultants"

Table 6.4 Most Important Benefits for Study 2

day to day cover
digital doctor
digital nurse
free back and neck physio
Medical Scans
Eye tests
glasses
Online GP
Nurse on call
Health screening

Table 6.5 Known Benefits in Study 2

The benefit questions which had a deterministic answer were answered correctly 93% of the time as shown in Table 6.6 below;

Question	Correct	Incorrect	% Correct	Note
Which plan is better for cancer accommodation support?	1	1	50%	
Employee assistance	1	0	100%	
Which plan is better for overseas hospital?	1	0	100%	One participant misinterpreted the question to look at the entire overseas category
Which plan is better for optical	4	0	100%	
Which plan is better for hearing?	2	0	100%	One participant answered for vaccinations
Which plan is better for vaccinations?	3	0	100%	
Which plan is better for individual excess?	3	0	100%	Participant misunderstood the concept of excess
	15	**1**	**93%**	

Table 6.6 Deterministic Benefit Comparison Questions in Study 2

The next five questions required the participant to compare a single benefit between two separate insurance products. The participant was allowed to choose which benefit they wanted to compare for each question and so the answers are grouped by the benefit selected in the Appendix: Answers to Benefit Comparison Questions

The final two questions in the quiz required the participant to compare a whole category of benefits. The initial category had 2 benefits and the second had 4 benefits. The answers are grouped by the category selected in the Appendix: Category Comparison Questions

After each question, the participant was asked to give an explanation for choosing their answer. The reasons for choosing answers are documented in the Appendix: Answers to Benefit Comparison Questions and Appendix: Category Comparison Questions

Following the quiz, the participants were asked for feedback and they rated the chatbot using a standard System Usability Scale (SUS) test after interacting with the chatbot. Figure 6.2 charts the SUS results for the participants of Study 2. The mean was 67 and the standard deviation was 8.

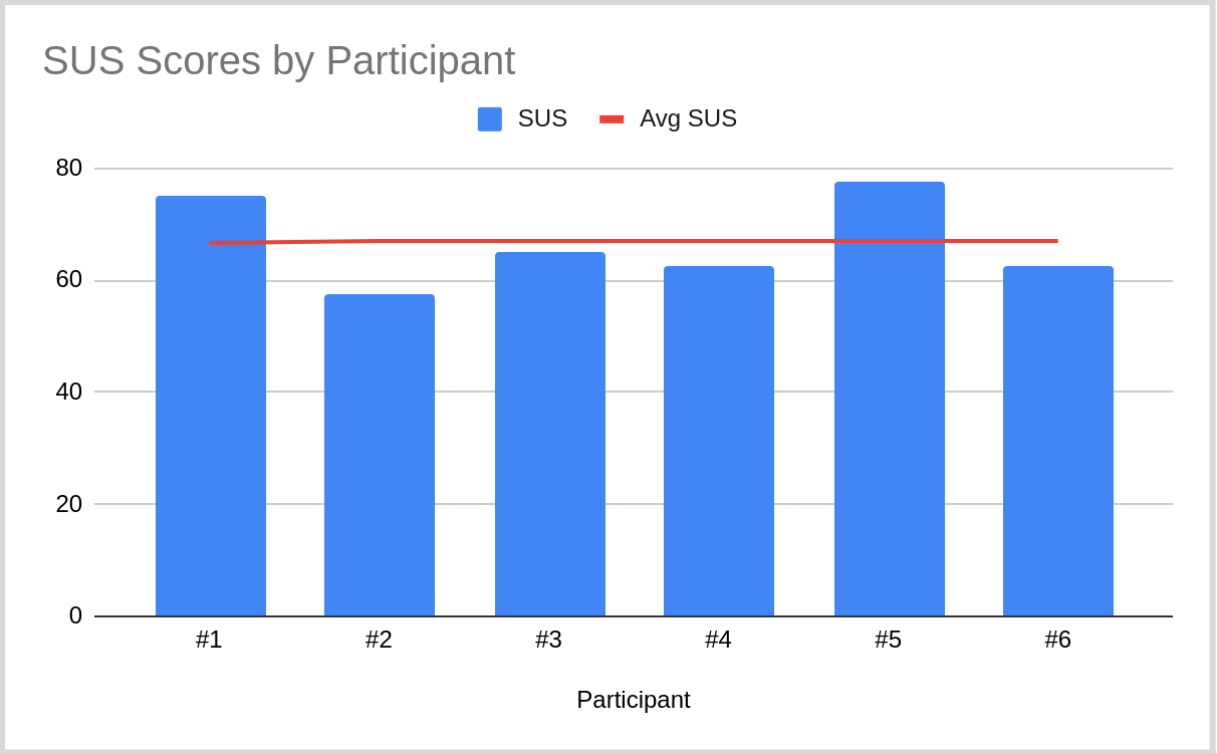


Figure 6.2 SUS Scores for Study 2

The goal time was measured for each participant as the time taken to answer all questions in the quiz. Figure 6.3 charts the times required by each participant in Study 2 to complete all goals. The mean was 24 minutes and the standard deviation was 9 minutes.

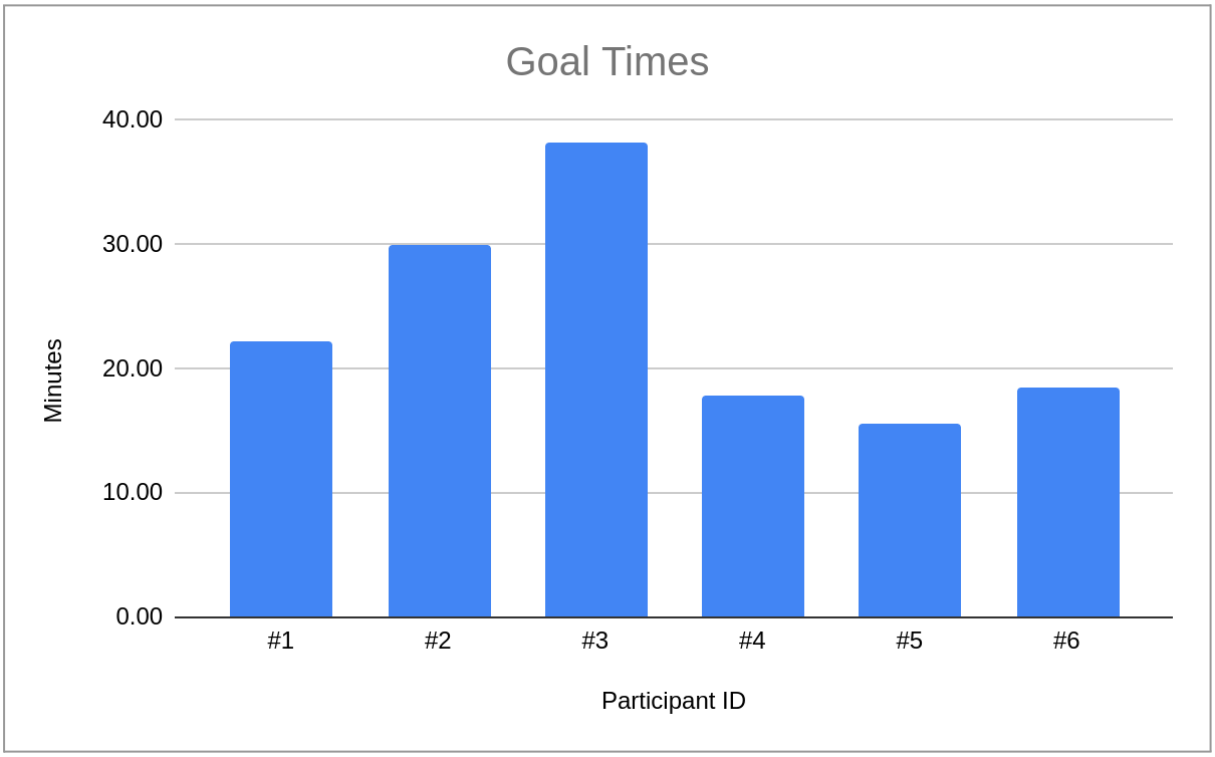


Figure 6.3 Goal Times for Study 2

Figure 6.4 charts the percentage of requests misunderstood by the chatbot for each participant in Study 2. The mean was 12% and the standard deviation was 7%.

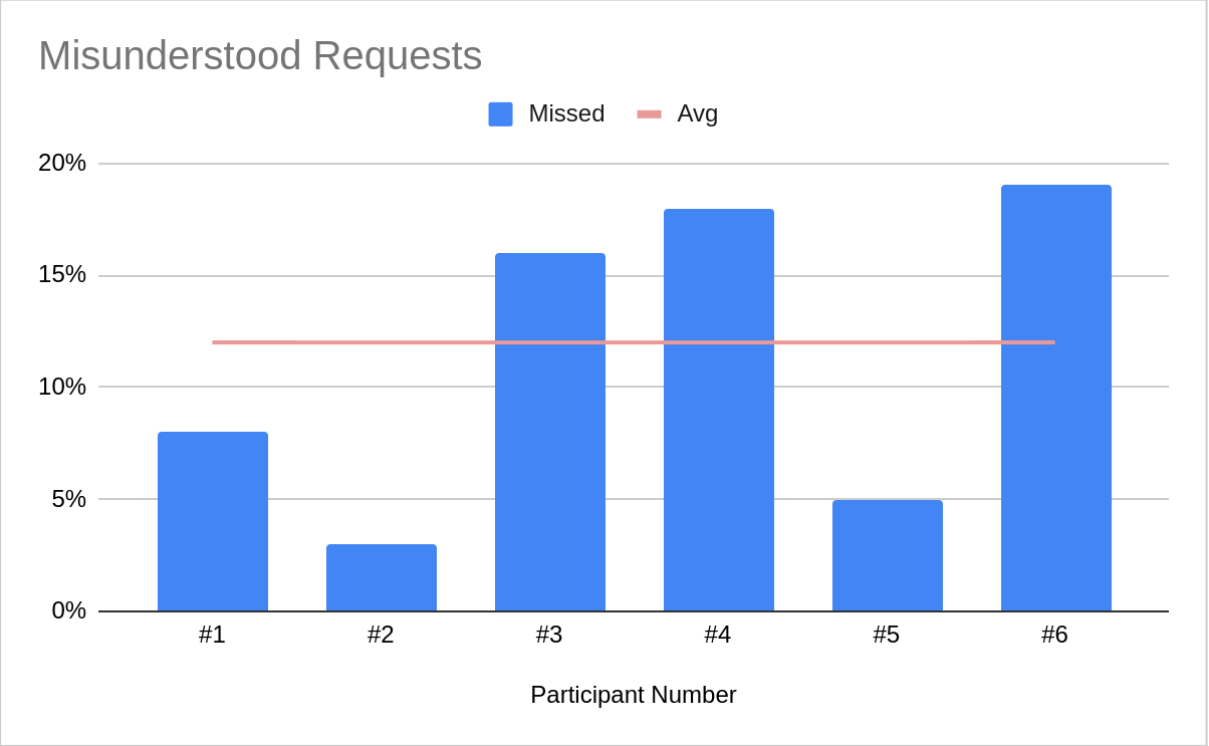


Figure 6.4 Misunderstood Requests for Study 2

The participants were asked for their feedback on the chatbot. Their answers to the question,"What is your first impression of the chatbot?" are presented in Table 6.7 below. One participant declined to answer the question.

"It seems to struggle with any follow up questions and just lists the benefits as they would appear on a statement of benefits",

"1st impression is this is difficult to phrase the things I need to find out. I'm 45 so health care is important, but financial cost is also very important. I think if I was older, I'd definitely need to be chatting through more specifics on it. But it definitely helped get me thinking on what cover I need and allowed me to get an understanding of both plans."

"they understood most terminology"

"I didn't think it was that helpful. The best it could do was search the table of benefits for the two plans and show me a chunk of text that I still had to sort through and try to

understand. I feel like it slowed down the process and I'd be quicker comparing the two plans myself." "Very easy to use"

Table 6.7 First Impressions of the Chatbot in Study 2

The participants were then asked , "How could the chatbot be improved?". Their answers are presented in Table 6.8. One participant declined to answer.

"Have more information for follow up questions" "Suggestive terms would definitely help. For example, I mentioned hospital cover initially, but that is way too broad, and then the suggested words appeared in one of the later answers and I was able to choose, semi-private, semi-private in public, etc...." "more intuitive" "Offer some knowledge or input rather than just finding and presenting the information. If a term appears in multiple categories show the relevant sentence where the term is used. For example only the section about cardiac scans in the radiology section when I am focusing on cardiac care." "No it worked well"

Table 6.8 Suggested Improvements in Study 2

All chatbot errors are documented in the ***Appendix: Answer data & chatbot messages***
All of the participant utterances are documented in the ***Appendix: Answer data & chatbot messages***

6.5 Discussion

Participants answered 93% of the deterministic benefit comparison questions correctly in the second study which was an improvement on the result of 86% in the first study. Confusion was alleviated for some participants. In preparation for Study 2, an explainer note was added which was displayed to participants when they were shown the value of individual excess for both plans. The note stated “please note that less excess means more money back for you when you claim” . All three participants who answered the questions for “individual excess” in Study 2 of the experiment gave the correct answer.

The SUS score improved to 67 in Study 2 from 61 in Study 1. Moreover there were no extremely low scores in the second execution as can be seen in the comparison below. The

standard deviation of the first execution was 20 compared to 8 for the second run. This indicated that the modifications made to the chatbot following the first execution of the experiment helped to improve the SUS score and to even out the variation in scores. However, A t-test was performed and the two-tailed P value was 0.4967 which is higher than the generally accepted threshold of 0.05 [62]. This means that the change in the mean SUS score is not proven to be statistically significant. Figure 6.5 shows a comparison between the SUS scores of Study 1 and those of Study 2. The individual charts are repeated from 5.2.3 Results and 6.4 Results. As noted in 6.3 Experimental Setup there were fewer participants in the second study.

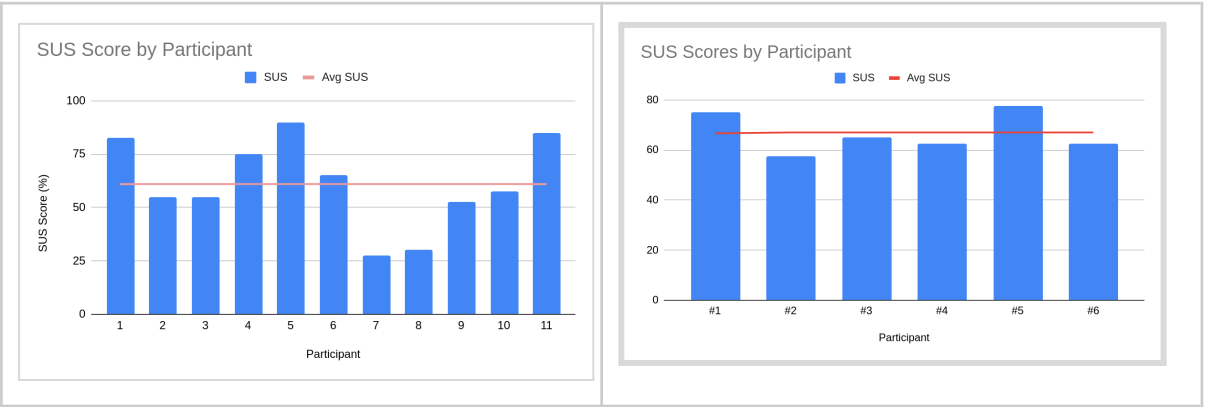


Figure 6.5 Comparison of SUS Scores

Category navigation improved in the second execution after buttons were introduced. An average of 6 category items were viewed per participant in the second execution. This compared favourably to 4 items in the first execution. In the first execution three people could not grasp the concept of category navigation and did not navigate a single category item. In the second execution, every participant indicated understanding by navigating at least one category. A t-test was performed and the two-tailed P value was 0.2494 which is higher than the generally accepted threshold of 0.05 [62]. This means that the improvement in category navigation is not statistically significant. Figure 6.6 compares the number of category items explored by participants in Study 1 with those of Study 2.

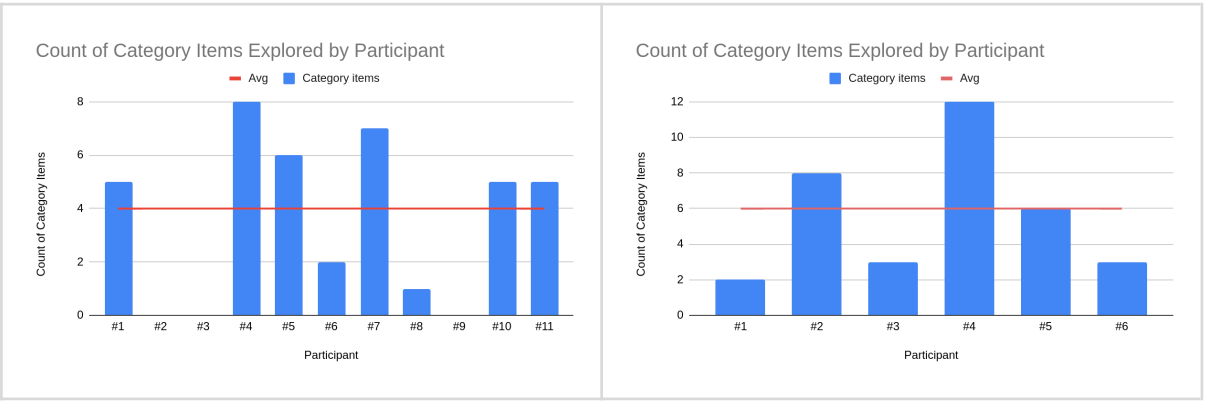


Figure 6.6 Comparison of Category Items Explored

The average percentage of user requests which were misunderstood by the chatbot in Study 2 improved by a factor of more than two from Study 1. It decreased to 12% of all requests compared to 28% in the first execution. The standard deviation for misunderstood requests was 17 in the first execution compared to 7 in the second. This meant that there was less variation across participants' interactions. This indicates that the modifications made to the chatbot to cater for more intents and a broader range of participant language helped the chatbot to better understand what was being asked of it. A t-test was performed and the two-tailed P value was 0.0455 which is lower than the generally accepted threshold of 0.05 [62]. This means that the improvement in misunderstood requests is considered to be statistically significant. Figure 6.7 compares the counts of misunderstood requests from Study 1 with those of Study 2. The individual charts are repeated from 5.2.3 Results and 6.4 Results.

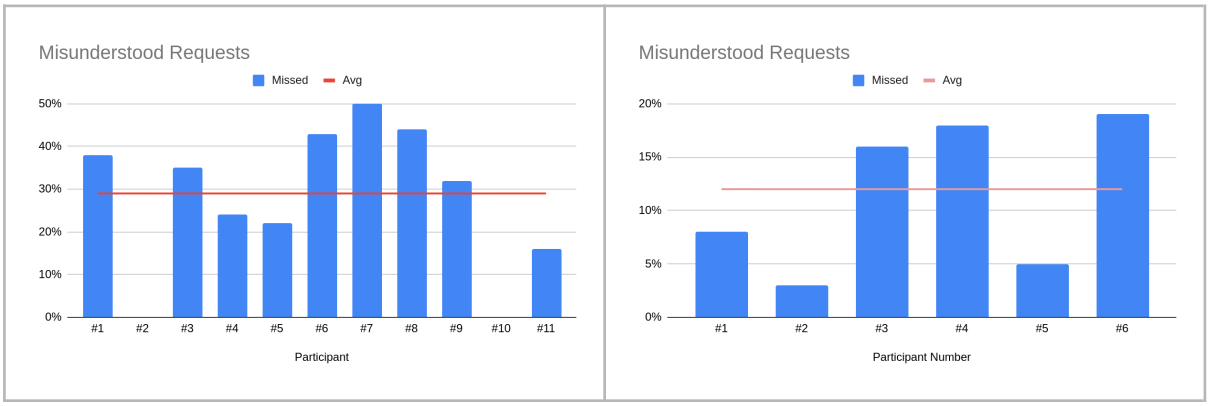


Figure 6.7 Comparison of Misunderstood Requests

Goal times increased for Study 2 with the participants taking longer on average to complete the quiz. This can in part be attributed to the extra amount of category navigation which they performed.

Simplification of responses and better presentation of information such as bullet lists of sub-items may in future make the benefits more understandable. If the insurers used exactly the same parameters when stating the benefit values then those parameters could be aligned in a table for direct comparison. The chatbot would be better if the dialogue management could be extended to handle sub dialogues and follow up questions. As one participant suggested, *"It seems to struggle with any follow up questions and just lists the benefits as they would appear on a statement of benefits".*

For future work, the training data gained from participant interactions could be more quickly extended if access was obtained to transcribed call logs from real customer conversations.

6.6 Summary

This chapter documents the objectives, experimental setup and results of the second execution of the experiment, Study 2. The setup is the same as Study 1 except that there are 6 participants in Study 2. The participant language discovered in Study 1 is used to

further develop the prototype and to modify the chatbot NLU. The struggles of participants attempting to navigate categories of benefits in Study 1 are used to inform the design of category navigation in the chatbot in preparation for Study 2. The same metrics are used for the evaluation of Study 2 so a comparison of results can be performed between both studies. The participant understanding of chatbot answers has improved as measured by the accuracy of the deterministic comparisons. The reduction in misunderstood requests by the chatbot indicates improved NLU. The SUS score improvement in Study 2 indicates a better experience of usability for participants.

Chapter 7 : Conclusion

This chapter concludes the report and discusses how the objectives set out at the beginning of the research project were met.

The first objective was to establish that the domain in question , health insurance, is complex in the sense that it has technical domain specific language that is not understood by potential customers and to indicate where in the domain confusion would arise for customers and how severe this problem is. This objective was achieved and a survey undertaken at the outset of the research indicated that more than half of the respondents were confused by the vocabulary used in the domain which was of a medical and technical nature. In addition it was discovered that 50% of those surveyed were confused by the domain products themselves. They did not understand what coverage was provided by particular products - which hospitals or which consultants or which procedures are covered. The complexity of the domain (health insurance) was further highlighted by the results of the chatbot experiment where many participants did not understand domain terms such as “excess”.

The next objective was to design and implement a chatbot to address the domain language and product data complexities using open source technology. This was achieved by constructing a chatbot to represent the domain using an open source framework and software libraries. The chatbot presented complex data from the domain in a manner which users understood. This was achieved using the Rasa open source framework and many other open source libraries and tools. The domain data was represented by a knowledge base which could be queried from custom code running in the Rasa action server. The ontology for the domain was based on a taxonomy provided by the industry regulator, the Health Insurance Authority (HIA). The chatbot knowledge base extended the starting model from the HIA adding more entities and relationships. It was further enhanced when the first execution of tests revealed richer natural language being used by the test participants.

The chatbot implementation supported two modalities, text and voice. The Twilio communications platform was leveraged to support the voice channel. Beta test users responded well to the voice implementation, especially the naturalness of the synthetic voice. Voice raised additional concerns not seen in the chat implementation including dropped words and speech timeouts. Call recording was essential to understanding problems with the voice dialogue and a custom call recording component was implemented to support this.

A further objective was to evaluate the chatbot design and implementation with participants who are potential purchasers of the products in the domain to determine if the chatbot could handle the complexity of the domain in terms of product data and domain language and that the participants could understand the chatbot responses. This objective was achieved through two executions of the same experiment in which participants were asked a series of questions to compare two products in the domain. They interrogated the chatbot to find answers to the questions posed. The participants correctly answered more than eighty percent of the questions which had deterministic answers. They found the chatbot responses very understandable for many single benefit comparison tasks.

Open questions were posed to participants to expose the natural language which they used to express intents and describe benefits in the domain. This language was used to improve the NLU detection of intents and extraction of entities and the chatbot was modified before the second execution of the experiment. More than 70 new benefit terms were discovered from participant interactions and these were used to enhance the chatbot ontology.

The final objective was to determine if the participants were satisfied with the usability of the chatbot. This was measured by using the standard System Usability Scale (SUS). The first execution of the chatbot experiment yielded an average SUS score of 61 which is below the 50th percentile score of 68. The results of the first execution were analysed and used to modify the chatbot. These modifications resulted in an improved average SUS score of 67 for the second execution of the experiment. The variation in SUS scores also evened out. In addition the requests misunderstood by the chatbot decreased by a factor of two. This can be attributed to the additional intent expressions and entities mined from the participant utterances and the open questions.

This work demonstrates that, even working without a corpus of industry data, a reasonable chatbot can be developed and improved by an iterative process. Language used by participants was mined and chatbot logs were analysed to extend the coverage of the knowledge base and the NLU pipeline. A wide variety of techniques and tools was used to improve the chatbot between test iterations. A new web front end was added, complete with new GUI affordances such as buttons. A technique was used to classify certain language used by participants as “junk” so the words would not interfere with the desired domain entity extraction.

The vocabulary complexities of the domain were addressed as far as possible and yet there are inherent challenges in the domain which a chatbot cannot resolve. The providers express the benefit values using different parameters to one another. This means that many attributes of a product are not directly comparable.

7.1 Future Work

This research used 11 participants for the first execution of the experiment and 6 different participants for the second. It would be interesting to extend the scale of the project to see how many more participant language terms and intents could be discovered. Better presentation of the domain information in the chatbot responses, for example aligned lists of sub parameters in benefit values, could improve the usability of the chatbot. The experiments could be conducted across other modalities. The dialogue management of the chatbot could be extended to handle more sub dialogues and follow up questions.

Since it is recommended to use a corpus of domain data as a grounding for chatbot development, it would be interesting to create such a corpus of data for this domain. This could be done using modern Automatic Speech Recognition (ASR) technology to transcribe the audio recordings of an industry call centre. The transcriptions could be mined for real world intent expressions and entities and sequences of conversation turns. These intent

expressions, entities and turns could be used to train the NLU of a chatbot and to describe new dialogue paths for it.

The expansion of generative AI models such as chatGPT into the area of task oriented dialogue opens new avenues for dialogue systems. Future work would be to determine how a generative model could be applied to create a more human-like interaction with users of a task oriented system for a business domain. The model would need to be constrained in some manner so that all questions on the attributes of products in the domain would return only factual answers based on the product specifications.

Bibliography

1 Burgess, K. John, and Julia Connel. Developments in the Call Centre Industry. Taylor & Francis Limited, 2006.

2 Beraud, Mathieu, et al. "Job quality and career opportunities for call center workers: Contrasting patterns in France." Low-Wage Work in France (2008): 254-87.

3 Hillmer, Steve, Barbara Hillmer, and Gale McRoberts. "The real costs of turnover: Lessons from a call center." Human Resource Planning 27.3 (2004): 34-42.

4 Nuruzzaman, Mohammad, and Omar Khadeer Hussain. "A survey on chatbot implementation in customer service industry through deep neural networks." 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE). IEEE, 2018.

5 Kore.ai Named a Leader in 2022 Gartner® Magic Quadrant™ for Enterprise Conversational AI Platforms, Aug 3, 2022. https://kore.ai/kore-ai-named-a-leader-in-2022-gartner-magic-quadrant-for-enterprise-conversational-ai-platforms/

6 Juuti, Kalle, and Jari Lavonen. "Design-based research in science education: One step towards methodology." Nordic studies in science education 2.2 (2006): 54-68.

7 Gartner Predicts Conversational AI Will Reduce Contact Center Agent Labor Costs by $80 Billion in 2026,STAMFORD, Conn., August 31, 2022 https://www.gartner.com/en/newsroom/press-releases/2022-08-31-gartner-predicts-conversational-ai-will-reduce-contac

8 Nicolescu, Luminița, and Monica Teodora Tudorache. "Human-Computer Interaction in Customer Service: The Experience with AI Chatbots—A Systematic Literature Review." Electronics 11.10 (2022): 1579.

9 Chen, Hongshen, et al. "A survey on dialogue systems: Recent advances and new frontiers." ACM Sigkdd Explorations Newsletter 19.2 (2017): 25-35.

10 Harms, Jan-Gerrit, et al. "Approaches for dialog management in conversational agents." IEEE Internet Computing 23.2 (2018): 13-22.

11 Bobrow, Daniel G., et al. "GUS, a frame-driven dialog system." Artificial intelligence 8.2 (1977): 155-173.

12 Young, Steve, et al. "The hidden information state model: A practical framework for POMDP-based spoken dialogue management." Computer Speech & Language 24.2 (2010): 150-174.

13 Rasa slot validation https://rasa.com/docs/rasa/slot-validation-actions/ Jan 12, 2022

14 DietClassifier https://rasa.com/docs/rasa/components/#dietclassifier Mar 7, 2022

15 Jurafsky, Dan. Speech & language processing. Pearson Education India, 2000.

16 Forms in rasa
https://rasa.com/docs/rasa/forms/ May 5, 2022

17 Greyling, Cobus. ”Dialog Management Considerations for Chatbots” 2020
https://cobusgreyling.medium.com/dialog-management-considerations-for-chatbots-6ed4dca65a80 Feb 20, 2022

18 Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

19 Vlasov, Vladimir, Johannes EM Mosig, and Alan Nichol. "Dialogue transformers." arXiv preprint arXiv:1910.00486 (2019).

20 Lin, Hsien-chin, et al. "Domain-independent user simulation with transformers for task-oriented dialogue systems." arXiv preprint arXiv:2106.08838 (2021).

21 Allouch, Merav, Amos Azaria, and Rina Azoulay. "Conversational agents: Goals, technologies, vision and challenges." Sensors 21.24 (2021): 8448.

22 Zhou, Ce, et al. "A comprehensive survey on pretrained foundation models: A history from bert to chatgpt." arXiv preprint arXiv:2302.09419 (2023).

23 Følstad, Asbjørn, and Marita Skjuve. "Chatbots for customer service: user experience and motivation." Proceedings of the 1st international conference on conversational user interfaces. 2019.

24 Akhtar, Mubashra, Julia Neidhardt, and Hannes Werthner. "The potential of chatbots: analysis of chatbot conversations." 2019 IEEE 21st conference on business informatics (CBI). Vol. 1. IEEE, 2019.

25 Sheehan, Ben, Hyun Seung Jin, and Udo Gottlieb. "Customer service chatbots: Anthropomorphism and adoption." Journal of Business Research 115 (2020): 14-24.

26 Reinkemeier, Fabian, and Ulrich Gnewuch. "Designing effective conversational repair strategies for chatbots." (2022).

27 Ashktorab, Zahra, et al. "Resilient chatbots: Repair strategy preferences for conversational breakdowns." Proceedings of the 2019 CHI conference on human factors in computing systems. 2019.

28 Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).

29 Bommasani,Rishi and Percy Liang .”Reflections on Foundation Models” Oct 18, 2021
https://hai.stanford.edu/news/reflections-foundation-models

30 Bommasani, Rishi, et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021).

31 https://www.ft.com/content/4fae2380-d7a7-410c-9eed-91fd1411f977 Jan 7, 2023.

32 How ChatGPT Managed to Grow Faster Than TikTok or Instagram
https://time.com/6253615/chatgpt-fastest-growing/ Feb 21, 2023.

33 Rasa pre-trained embeddings https://rasa.com/docs/rasa/tuning-your-model/#pre-trained-embeddings Apr 20, 2022

34 Bang, Yejin, et al. "A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity." arXiv preprint arXiv:2302.04023 (2023).

35 Jo, Eunkyung, et al. "Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention." (2023).

36 Karpukhin, Vladimir, et al. "Dense passage retrieval for open-domain question answering." arXiv preprint arXiv:2004.04906 (2020).

37 Shuster, Kurt, et al. "Retrieval augmentation reduces hallucination in conversation." arXiv preprint arXiv:2104.07567 (2021).

38 Barrouillet, Pierre, et al. "Time and cognitive load in working memory." Journal of Experimental Psychology: Learning, Memory, and Cognition 33.3 (2007): 570.

39 Cowan, Nelson. "The magical mystery four: How is working memory capacity limited, and why?." Current directions in psychological science 19.1 (2010): 51-57.

40 Rosenshine, Barak, and Carla Meister. "The use of scaffolds for teaching higher-level cognitive strategies." Educational leadership 49.7 (1992): 26-33.

41 Lee, Jiyong. "Task complexity, cognitive load, and L1 speech." Applied linguistics 40.3 (2019): 506-539.

42 Beckmann, Jens F. "Taming a beast of burden–On some issues with the conceptualisation and operationalisation of cognitive load." Learning and instruction 20.3 (2010): 250-264.

43 Comes, Sherry, David Schatsky and Rameeta Chauhan. “Conversational AI Five vectors of progress” Deloitte (2021) https://www2.deloitte.com/us/en/insights/focus/signals-for-strategists/the-future-of-conversational-ai.html Oct 17, 2022

44 Gevirts Morris, “The Trouble with Word Error Rate (WER)” 2019 https://blog.deepgram.com/the-trouble-with-wer/ Oct 23, 2022

45 Brunk, Joshua, et al. "A Survey of Task-Based Machine Learning Content Extraction Services for VIDINT." arXiv preprint arXiv:2207.04158 (2022).

46 Starke, Alain D., and Minha Lee. "Unifying recommender systems and conversational user interfaces." Proceedings of the 4th Conference on Conversational User Interfaces. 2022.

47 Sonnenburg, Soren, et al. "The need for open source software in machine learning." (2007): 2443-2466.

48 Keras

https://keras.io/about/ Jan 11, 2023.

48 Neural machine translation with a Transformer and Keras https://www.tensorflow.org/text/tutorials/transformer Jan 17, 2023.

49 Abadi, Martín, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." arXiv preprint arXiv:1603.04467 (2016).

50 Bocklisch, Tom, et al. "Rasa: Open source language understanding and dialogue management." arXiv preprint arXiv:1712.05181 (2017).

51 https://rasa.com/docs/rasa/arch-overview Feb 9, 2022.

52 Bending the ml pipeline in rasa 3.0 https://rasa.com/blog/bending-the-ml-pipeline-in-rasa-3-0/ Feb 18, 2022.

53 Abd-Alrazaq, Alaa, et al. "Technical metrics used to evaluate health care chatbots: scoping review." Journal of medical Internet research 22.6 (2020): e18301.

54 Walker, M., Kamm, C., Litman, D.”Towards developing general models of usability with PARADISE”. Journal of Natural Language Engineering 6(3-4), 363–377 (2000)

55 Maroengsit, Wari, et al. "A survey on evaluation methods for chatbots." Proceedings of the 2019 7th International conference on information and education technology. 2019.

56 Jenneboer, Liss, Carolina Herrando, and Efthymios Constantinides. "The impact of chatbots on customer loyalty: A systematic literature review." Journal of theoretical and applied electronic commerce research 17.1 (2022): 212-229.

57 Bangor, Aaron, Philip T. Kortum, and James T. Miller. "An empirical evaluation of the system usability scale." Intl. Journal of Human–Computer Interaction 24.6 (2008): 574-594.

58 Cameron, Gillian, et al. "Assessing the usability of a chatbot for mental health care." Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5. Springer International Publishing, 2019

59 Jakob Nielsen. 1993. Iterative user-interface design. Computer 26, 11 (1993)

60 Botfrront chat widget https://github.com/botfront/rasa-webchat Oct 12, 2022.

61 Dewi, Ida Ayu Putu Arini, I. Nyoman Kardana, and I. Nyoman Muliana. "Functions of speech acts in “critical eleven”." RETORIKA: Jurnal Ilmu Bahasa 6.1 (2020): 1-6.

62 Smucker, Mark D., James Allan, and Ben Carterette. "A comparison of statistical significance tests for information retrieval evaluation." Proceedings of the sixteenth ACM conference on Conference on information and knowledge management. 2007.

Appendix: Glossary of Terms

The following terms were used in the context of the research project.

Term	Explanation
PBI	Plan Benefit Item. A single line item on a health insurance plan (product) e.g. “Private hospital semi-private room”
PBI Value	The value for a particular PBI on a specific health insurance plan e.g. “€350 annually”
Benefit	A general term describing something of value provided by a health insurer to the insured customer
Alias	A friendly name given to a benefit or a category of benefits to enable participants to use their own language when interacting with the chatbot.

Appendix : Comparison of Two Popular Health Insurance Plans from the HIA

The Health Insurance Authority

HealthPlus Access	Essential Connect Family
VHI Healthcare	Laya Healthcare
Current Version Commenced:31/12/2018	Current Version Commenced:15/12/2018

Prices Per Annum

	Per year		Per year
Adult	€2238.12	**Adult**	€1583.37
Child 1	€476.05	**Child 1**	€299.78
- Child 1 (0-4yrs)	€476.05	N/A	
- Child 1 (5-17yrs)	€476.05	N/A	
Child 2	€476.05	**Child 2**	€0.00
- Child 2 (0-4yrs)	€476.05	N/A	
- Child 2 (5-17yrs)	€695.46	N/A	
Child 3	€476.05	**Child 3**	€0.00
- Child 3 (0-4yrs)	€476.05	N/A	
- Child 3 (5-17yrs)	€476.05	N/A	
Child 4	€0.00	**Child 4**	€0.00
Young Adult	€695.46	**Young Adult**	€791.69
- Young Adult (18-20yrs)	€695.46	- Young Adult (18-20yrs)	€791.69
- Young Adult (21yrs)	€1043.50	- Young Adult (21yrs)	€784.68
- Young Adult (22yrs)	€1287.32	- Young Adult (22yrs)	€959.19
- Young Adult (23yrs)	€1531.12	- Young Adult (23yrs)	€1137.53
- Young Adult (24yrs)	€1774.93	- Young Adult (24yrs)	€1315.86
- Young Adult (25yrs)	€2018.73	- Young Adult (25yrs)	€1494.20
Newborn	€0.00	**Newborn**	€0.00

In Patient

A Multi-Occupancy or Semi-Private room in a Public Hospital and Day Case.	
Covered	Covered
Public Hospital Private Room and Day Case	
Covered	Covered up to semi-private rate
Private Hospital Semi-Private Room	
Covered but 90% cover for certain cardiac and special procedures in Hermitage and Galway Clinic	Covered with once-off €300 excess per family per year; No excess applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals Orthopaedic Hospital List http://bit.ly/LayaList
Private Hospital Private Room	
Semi-private rate + 100% of technical charges. Some private	Covered with once-off €300 excess per family per year and

1 / 6

In Patient

hospitals full cover. Ask Vhi for details	shortfall of €250 per night. Full cover in the Mater Private Cork; No excess/shortfall applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals Orthopaedic Hospital List http://bit.ly/LayaList
Not covered	Not covered
The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures	
90%	Covered in Beacon Hospital with once-off €300 excess per family per year. Reduced cardiac list covered in Mater Private and Blackrock Clinic, no excess; Ask Laya for details.
The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Special Procedures	
90%	Covered in Beacon with once-off €300 excess per family per year; No excess applies to 5 orthopaedic procedures (hip and knee replacement). No cover in Blackrock Clinic or Mater Private except for 5 orthopaedic procedures (hip and knee replacement) which are fully covered. Orthopaedic Hospital List http://bit.ly/LayaList
The Blackrock Clinic, the Mater Private and the Beacon Hospital procedures other than Cardiac and Special	
55% semi-private (Mater Private only); 45% private (Mater Private and Blackrock Clinic); Beacon Hospital full cover	Covered in Beacon Hospital with once-off €300 excess per family per year. No cover in the Blackrock Clinic or Mater Private
Day Case Private Hospitals	
Covered	Covered with once-off €300 excess per family per year.
Day Case The Blackrock Clinic, the Mater Private and the Beacon Hospital	
Covered	Covered in Beacon Hospital with once-off €300 excess per family per year. No cover in the Blackrock Clinic or Mater Private.
Convalescence	
€51 x 14 nights in semi private and private room accommodation	€50 x 14 days
Cancer Accommodation Support	
€100 for 1 night's accommodation for each treatment	Not covered

Maternity

Hospital Costs Up To 3 Nights	
Full Cover	Covered
Home Births	
€3,150	€3,500
Delivery Consultant Fees	
Covered up to agreed charges incurred on day of delivery for participating consultants	Covered up to agreed charges incurred on day of delivery for participating consultants
Outpatient Maternity Consultant Care	
See pre/post natal care	€400
Postnatal Home Help	
Not covered	Not covered
Postnatal Home Nursing	

Maternity	
Not covered	Not covered
Child Home Nursing	
€100 x 14 days	€80 x 28 days
Pre & Postnatal Care	
€255	Not covered
Child Healthcare Benefit	
Vhi paediatric clinic (Swiftcare) initial consultant consultation 50% cover, follow up paediatric treatment and services (including lactation consultant, dietician, ultrasound, blood tests and x-ray) 50% cover	Not covered
Parent Accompanying Child	
€40 X 14 days not incl. first 3 days	Not covered
Partner Benefit	
Not covered	Not covered
Breastfeeding Consultancy	
Not covered	Not covered
Other Maternity Benefits	
Not covered	Not covered

Out-patient Benefits	
Individual Excess	
€100	€150
Family Excess	
€100 per person	€250
Consultant Fees	
€60 per visit; Second Opinion Service covered; Consultant Consultation at Vhi Swiftcare 50% (orthopaedic, oral maxillofacial & sports medicine)	€51 per visit
GP Visits	
€30 per visit; Practise Nurse €20 X 1 visit; Vhi online doctor full cover 12 visits	24 hr GP Line; 24 Hr Nurseline; GP live online consultations X 3
Physiotherapist	
€30 X 7 visits	Physioline
Emergency Dental Care	
Not covered	€510 for restorative treatment within 5 days
Non Emergency Dental	
€20 X 7 visits	Not covered
Home Nursing	
Full cover for Vhi Homecare benefit for certain procedures and upon referral from certain hospitals	€40 x 40 days
A&E Charge	
€75 X 1 visit (Applies to A&E in Public Hospitals Only); Vhi Swiftcare full cover, unlimited visits with a €50 excess per visit.	€20 per visit; QuickCare consultations & treatment Up to €150 per visit in Laya Healthcare approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in Laya Healthcare

Out-patient Benefits	
	approved minor injuries & illness centres
Alternative Medicines-Acupuncture, Chiropractor, Osteopath	
€20 x 12 combined visits	Not covered
Dietician, Occupational Therapist, Chiropodist, Speech Therapist	
€20 x 12 combined visits	Not covered
Other Day To Day Practitioners	
Orthoptists, Reflexologist, Podiatrists and Clinical Psychologist included under Alternative Practitioners benefit	Healthcoach online assessment and personalised programmes - unlimited use; Healthcoach face-to-face consultation - full cover for 1 consultation every 2 years.
Cancer Support Benefit	
Wig/hairpiece, mastectomy bra, swimsuit, surgical prosthesis (no excess, but subject to the maximum for medical and surgical appliances of €3,200)	Cover for 1 hairpiece per year following cancer treatment; Cover for first breast prosthesis following inpatient claim, second prosthesis assessed in accordance with surgical appliance list; Sleeping caps, full refund for 1 sleeping caps per membership year; Eyebrow tattooing (following cancer treatment), up to €60 once per membership year
Psycho-Oncology Counselling	
€50 X 10 visits	Not covered
Manual Lymph Drainage	
€50 X 10 visits	€500; no excess
Hearing Test	
Not covered	Not covered
Optical	
Eye test: €30 in each 24 month period Glasses/contact lenses: 75% of charges up to €100 in each 24 month period	Not covered
Prescription Costs	
Not covered	Not covered
Employee Assistance Programme	
Not covered	Not covered
Child Speech And Language Therapy	
See benefit for Dietician, Occupational Therapist, Chiropodist, Speech Therapist above Meditation App If included in Your plan, Vhi will pay the benefit listed in Your table of benefits towards the annual subscription costs of specified meditation apps. You can claim back the benefit against one app, once per renewal year. See Vhi.ie/emotional-wellbeing for further details.	Not covered
Vaccinations	
Not covered	Not covered
Outpatient Policy Limit	
€3,200	€6,400

Out-patient Benefits

Hormone Replacement Therapy for Gender Dysphoria	
Not covered	Not covered

Out-patient Radiology

Approved Centres	
CT full cover; PET-CT full cover; MRI covered, no excess for category 1, with €125 excess for category 2, see rules.	MRI, CT and PET-CT covered in full in direct payment centres except cardiac CT's which are covered by non-direct payment terms. Covered up to €500 and subject to outpatient excess in non direct payment centres
Non-Approved Centres	
Not covered	Not covered
Radiology Consultants' Fees	
€60 per procedure	Covered subject to outpatient excess
Radiology Test Fees	
50% of charges to a max of €850 per year	Coverd up to €500, subject to out-patient excess
Health Screening	
Lifestages Screening Programme €335 per screen and heart check and cancer check every 2 years in a Vhi Medical Centre; Full cover for Mammograms and Dexa Scans every 2 years; no excess for above; Joint Care Programme: Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health, full cover every 2 years; Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from Joint Care Screening - Full cover x 3 visits, no excess; Medfit Cardiac Care programme €250 every 2 years; Urgent Cardiac Care benefit €250 per year; Medfit Cardiac Rehabilitation programme €250 per year. Full cover for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment covered up to the level of cover provided in the Inpatient section.	Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening up to €30; Men's Cancer Screening up to €20. Full cover for 'HeartBeat' cardiac screening, no excess; Home testing kits, up to €20. Health Screening €150 every 2 years paid directly to centre

Overseas

Benefit Abroad For Surgical Procedures Available In Ireland	
€100,000	Covered in the EU subject to approval
Benefit Abroad For Surgical Procedures Not Available In Ireland	
€100,000	Covered in the EU subject to approval
Hospital Bill	
€100,000	€100,000
Repatriation Expenses	
Covered	€2,000,000
Companion Repatriation Expenses	
€1,000	€1,000
24 Hour Telephone Assistance	
Covered	Covered
Expenses For Companion	

Overseas	
€1,000	Not covered
Gender Reassignment Surgery	
Cover for medically necessary surgeries subject to pre-approval up to the level of cover on the plan.	Cover up to €100,000 subject to prior approval

Psychiatric Cover	
Inpatient Psychiatric Non Alcohol Drug	
100 days	100 days
Inpatient Psychiatric Alcohol Related	
91 days per 5 years	91 days per 5 years
Outpatient Mental Health Care	
Mental Health Therapy €75 x 12 visits; Clinical Psychologist €20 x 12 visits under Alternative Practitioners benefit	€50 x 12 Mental Health Consultations in Dean Clinic; €50 x 12 Eating Disorder Consultations in Lois Bridges both subject to outpatient excess; Counselling Support Benefit: 24/7 Telephone Counselling, full cover; Face to face/Video/Phone Counselling, full cover for 5 visits per issue; Cover for specific support services, full cover for 1 session per issue for legal, consumer & health advice, career guidance, life coach & mediation.

Fertility Benefits	
Infertility Benefit	
Not covered	Not covered
Fertility Preservation	
Not covered	Not covered
Fertility Support Services	
Not covered	Not covered

Appendix: Answers to Open Benefit Questions

What is the most important benefit you look for in a health insurance policy when purchasing one?

“instant access to doctors”
“hospital cover”
"good cover and a good price "
"Comprehensive cover, not basic. Urgent care and consultants visits important in case of severe illness"
"Access to digital doctor"
"The quickest access to the best medical care"
"dental cover"
"Pre Existing condition coverage"
"Its the range of hospitals and what illnesses the plan covers"
"cardiac cover"
"A low excess, with a broad range of hospitals"
"Most cover for best price"
"Sporting injury cover"
"everyday health expenses"
"Consultants"
"best value with a combination of day-to-day benefits and cover in hospitals."
"hospital stays and operations covered"

There are more than 60 benefits associated with each health insurance plan. List as many as you know off the top of your head without searching the internet.

“instant access, online consultations, access to private rooms”
“dental, hospital, GP, perscription, practitioner”
"cardiac, fertility, cancer care, joint care, mental health"
"Dietician, Physio, Osteo, Cancer screening, maternity benefits, new born screening, Blood testing for genetic issues, Counselling, Dental, Optical, Podiatrist, Acupuncture, Chiropractor , Parenting, "
"Digital doctor, Dentist visits, GP visits, Consultant Fees, Physio visits"
"dentistry, cancer care, maternity, physio therapy, psycho therapy, ophtalmology"

"cardiac cover. eye cover. emergency cover. single room. shared room . family cover. cancer cover. orthopedic cover. hybrid theatre cover . robotic cover. specialty cardiac cover. dental cover. travel insurance cover. "
"GP cover, Emergency dental cover, Consultant fees, Cancer treatment, MRI, Maternity hospital cover, Private room in a hospital, "
"Maternity cover,Consultants cover,Semi private room in private or public hospital ,Private room in private or public hospital,Online GP,Nurse on call,Counselling,Health screening,"
"Accident and Emergency,Inpatient cover,Maternity cover,Consultations,Radiology,GP cover,Physiotherapy,Outpatient benefits,Mental Health benefits,Cardiac care,Fertility treatment",
"GP visits, dental routine checkups, glasses, private hospital, private room in public hospital, consultant visits",
"GP, Hospital, Consultants, Dentist, Psychiatry, Physiotherapy, Radiology, Medical Scans, Eye tests, Pregnancy care,"
"free travel insurance, day to day cover, digital doctor, digital nurse, free back and neck physio"
"Cardiac care, Orthopedic care, dental care, Private hospital care, Cancer care, Physio cover, Fertility cover, Travel cover, A&E cover, Maternity Cover, GP cover, acupuncture "

Which plan is better for <your most important benefit>?
This question used a previous answer as input - the participants' answer to the earlier question, “What is the most important benefit you look for in a health insurance policy when purchasing one?”

Question: Which plan is better for “instant access to doctors” ? Answer: “Plan A” Reason: “It has 24 hour access to a gp by phone”
Question: “Which plan is better for “hospital cover”?” Answer: “Plan A” Reason: “there is 12 individual benifits”
Question: Which plan is better for "good cover and a good price "? Answer: "Other" reason: "the price does not state if its per person or per couple/family etc"
Question: Which plan is better for "Comprehensive cover, not basic. Urgent care and consultants visits important in case of severe illness"? answer: "Plan B" reason: "More benfits"
Question: Which plan is better for "Access to digital doctor"? answer: "Other" reason: "Did not understand"
Question: Which plan is better for "The quickest access to the best medical care"? answer: "Other" reason: "The question was not answered"
Question: Which plan is better for "dental cover"? answer: "Other"

reason: "none of the plans is good enough, actually no real cover at all"
Question: Which plan is better for "Pre Existing condition coverage" ? answer: "Other" reason: "The bot did not understand"
Question: Which plan is better for "Its the range of hospitals and what illnesses the plan covers"? answer: "Other", reason: "It doesn't understand the question "
Question: Which plan is better for "cardiac cover"? answer: "Plan A" reason: "Plan A has full cover in the Beacon Hospital with once-off €300 excess per family per year"
Question: Which plan is better for "A low excess, with a broad range of hospitals"? answer: "Plan B" reason: "Plan B has the lower excess"

Appendix: Answers to Benefit Comparison Questions

These questions asked the participant to compare a single benefit between two plans.The participant was allowed to select which comparison questions they answered.
They asked for a reason to explain their answer.

Which plan is better for convalescence?

Answer: Plan B Reason: it gives you better accomadation
answer: "Plan B" reason: "for an additional € per night have the choice of semi and private room"
answer: "Plan B" reason: "The amount is higher and it includes semi private and private accomodation"
answer: "Plan A" reason: "There is only €1 difference and no criteria for the exact accommodation"
answer: "Plan B" reason: "B offers semi private and private"
answer: "Plan A" reason: "There is more information, Plan B is not really telling me anything"

Which plan is better for cancer accommodation support?

Answer: Plan B Reason: Plan B has €100 for 1 night's accommodation for each treatment plan A has No cover
answer: "Plan B" reason: "Plan A has no cover"
answer: "Other" reason: "I did not get enough answers to my questions"
answer: "Plan B" reason: "100 euro for 1 nights accommodation for each treatment"
answer: "Plan B" reason: "Plan A has no cover"

Which plan is better for employee assistance?

answer: "Both are equal" reason: "As neither offers employee assistance"
answer: "Both are equal" reason: "They don't cover"

Which plan is better for certain cardiac procedures at the blackrock clinic?

answer: "Both are equal" reason: "The answer does not specific either A or B but generalises"

Which plan is better for overseas hospital?

Answer: Both are equal Reason: both plans have 100,000 cover
answer: "Both are equal" reason: "Amount is the same"
answer: "Both are equal" reason: "there was nothing to tell me any different"

Which plan is better for optical?

answer: “Other” Reason: “plan 2 has a better room”
answer: "Plan A" reason: "I do not wear contacts or glasses so it is not needed monthly for me "
answer: "Plan B" reason: "Plan A has no cover"
answer: "Plan B" reason: "plan a has no cover"
answer: "Plan B" reason: "Plan B has cover for eye tests of €30 in each 24 month period. Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period"
answer: "Plan B" reason: "Plan A has no optical cover"

Which plan is better for hearing?

answer: "Both are equal" reason: "no difference, no cover on either plan "
answer: "Both are equal" reason: "neither has any cover for hearing tests"
answer: "Both are equal" reason: "Both plans have no cover"

Which plan is better for vaccinations?

Answer: Both are equal

Reason: neither cover vaccines
Answer: Other Reason: neither plan has any cover
answer: "Both are equal" reason: "No cover on either"
answer: "Both are equal" reason: "They both do not provide cover"
answer: "Plan B" reason: "child care is covered"
answer: "Both are equal" reason: "no cover on both"

Which plan is better for individual excess?

answer: "Plan A" reason: "plan A has 150 and plan B has 100"
answer: "Plan B" reason: "Plan B excess is €50 than plan A"

Which plan is better for family excess?

answer: "Other” Reason: “it is not clear if plan A is per family or per person”
answer: "Plan B" reason: "less excess"
answer: "Plan B" reason: "100 per person"
answer: "Plan B" reason: "It is per person so the value is greater if there are more than 2 people in the family"
answer: "Plan A" reason: "family excess is cheaper"
answer: "Plan B" reason: "100 euros seems more sensible"
answer: "Plan A" reason: "Plan A has €250 "

Which plan is better for a semi private room in private hospitals?

answer: "Plan B" reason: "plan b has 90% cover where as plan A has a once off payment "
answer: "Other"

reason: "The answer did not specify which plan offers this."

Which plan is better for home nursing?

answer: "Other" reason: "Plan B refers to full cover but only for certain procedures which is not entirely satisfactory."

Which plan is better for a semi private room in private hospitals?

answer: "Plan B" reason: "Full cover, no shortfall. "
answer: "Other" reason: "Chatbot could not answerWhich plan is better for a semi private room in private hospitals?"
answer: "Plan B" reason: "Plan B has full cover but 90% cover for certain cardiac and special procedures in the Hermitage and the Galway Clinic"
answer: "Plan A" reason: "Plan B has very limited cover for rooms and is unclear"

Which plan is better for a private room in a private hospital?

answer: "Other" reason: "Both are overly complicated and don't answer the question"

Which plan is better for accident and emergency?

Answer: “Plan B” Reason: “ has more cover for multiple visits and I dont like the idea of (up to) in plan A as it is vague”
answer: "Plan B" reason: "much better cover "
answer: "Plan B" reason: "Full cover at insurer's clinic with unlimited visits"
answer: "Plan B" reason: "I believe that this plan would be best and cheaper in the long run"
answer: "Plan A" reason: "Costs 20 Euro a visit treatment up to 150\n"
answer: "Plan A" reason: "Cheaper excess for the A & E as long as you are not treated by a constultate"

Appendix: Answers to Category Comparison Questions

These questions asked the participant to compare an entire category of benefits between two plans. In addition to explaining their answer the participant was asked to "Explain how easy or difficult it was to compare all of the benefits from that category?"

Which plan is better for psychology?

answer: “Plan B” reason: “A has no cover for it.” category: “ it was straightforward”
answer: "Plan A" reason: "has more cover" category: "difficult"
answer: "Plan B" reason: "Plan A has no cover" category: "Easy - information is direct and to the point"
answer: "Both are equal" reason: "they do not cover what I am looking for" category: "I find it just frustrating because the bot does not seem to understand most of my questions"
answer: "Plan B" reason: "Plan A has no cover" category: "It was somewhat hard, as the wording was not clear"

Which plan is better for gender?

answer: "Plan B" reason: "Plan b as it only covers necessary surgeries" category: "quite easy"
answer: "Plan A" reason: "It has cover up to 100k" category: "Easy"
answer: "Both are equal" reason: "It doesn't answer the question" category: "Not really really vague"
answer: "Plan A" reason: "Plan A has cover up to €100,000 subject to prior approval " category: "ok"

Which plan is better for radiology?

answer: "Other" reason: "unsure about this one i would need further help"

category: "difficult"

Which plan is better for cardiac?

Answer: “Both are equal” Reason: “It was hard a bullet point more concise list would make it easier” Category: “they both have pros and cons it would be difficult to figure out which is best as there is a lot of medical terms and information”
answer: "Other" reason: "Chat Bot did not work, only my messages could be seen " category: "N/A "
answer: "Other" reason: "Again the bot did not answer directly" category: "difficult as it was convoluted"
answer: "Other" reason: "Not sure its too technical to make a decision" category: "Not easy in this format"
answer: "Plan A" reason: "lan A has full cover in the Beacon Hospital with once-off €300 excess per family per year. Reduced cardiac list coverage in the Mater Private and Blackrock Clinic with no excess" category: "ok"

Which plan is better for cancer?

answer: "Plan B" reason: "There are more benefits" category: "Moderately easy"
answer: "Other" reason: "It does not directly answer regarding cancer treatment" category: "Easy but did not answer my real concern"
answer: "Plan B" reason: "offers much more cover that I am looking for and woukd expect" category: "I start to dislike the bot very much"
answer: "Plan B" reason: "There seem some cross over, but Plan B has more" category: "Difficult for option 1"

Appendix: Feedback on the Chatbot

What is your first impression of the chatbot?

“It is a convenient way to source information”
“some of it is fine, other parts the information is not easy to follow”
"very generic answers, does not understand a lot of questions ",
"Good"
"It was easy to use and was easy to compare plans"
"Relatively easy to use"
"poor, it does not understand my questions, or only partially. I find it frustrating"
“Not good at all"
"Basic chat bot with no fancy icons"
"its ok"
"Good, but can't handle non standard terms"
"Very easy to use"
"It seems to struggle with any follow up questions and just lists the benefits as they would appear on a statement of benefits",
"I didn't think it was that helpful. The best it could do was search the table to benefits for the two plans and show me a chunk of text that I still had to sort through and try to understand. I feel like it slowed down the process and I'd be quicker comparing the two plans myself."
"they understood most terminology",
"1st impression is this is difficult to phrase the things I need to find out. I'm 45 so health care is important, but financial cost is also very important. I think if I was older, I'd definitely need to be chatting through more specifics on it. But it definitely helped get me thinking on what cover I need and allowed me to get an understanding of both plans.",

How could the chatbot be improved?

“simplified medical terms”
“make the information more simple”
"give more information"
"More understanding. Ie when I input comprehensive it could have given me some options. I dont understand that, would you like to search for family or individual plans?",
"N/A"
"Difficult to get specific answers"

"it definitely needs a much better vocabulary. When I ask about maternity it did not understand. I asked for psycho therapy and it did not understand. A bot needs to have a very broad vocabulary and even with typos"
"But a human in there. It was typical chatbot stuff, poor scrambled information, couldn't answer questions etc"
"Not sure its a pretty complex topic and i am not sure i would chose base on the information it gives to me",
"dont know"
"Introduce alternate words for the same processes"
"No it worked well",
"It seems to struggle with any follow up questions and just lists the benefits as they would appear on a statement of benefits"
"Offer some knowledge or input rather than just finding and presenting the information.\nIf a term appears in multiple categories show the relevant sentence where the term is used. For example only the section about cardiac scans in the radiology section when I am focusing on cardiac care."
"more intuitive"
"Suggestive terms would definitely help. For example, I mentioned hospital cover initially, but that is way too broad, and then the suggested words appeared in one of the later answers and I was able to choose, semi-private, semi-private in public, etc...."

Appendix: Chatbot Single Turn Errors

The chatbot incorrectly responded to the participant messages in the table below.
The trained entities were *pbi* (plan benefit item) for a single benefit ,*benefit_tag* for a category of benefits *faq_term* for something the bot can answer an FAQ question on. The trained intents were *ask_faq_question , ask_pbi_or_benefit_tag* .The DietClassifier is a rasa nlu component which predicts intents and entities.

#	Participant message	Response	Error cause
1	which plan is better for optical policies?	A category containing these 2 individual benefits exists for public hospital; 1) a private room in a public hospital, 2) a semi-private room in a public hospital Type a number between 1 and 2 for a benefit from the category public hospital	NLU ENTITY ERROR The DietClassifier extracted “optical policies” as a *benefit_tag*. Fuzzy wuzzy matched this term with “public hospital” with confidence of 65, just above the threshold
2	!	I'm sorry. I didn't quite understand what you said.	USER ERROR User hit exclamation instead of the digit “1” *nlu_fallback* intent was predicted for “!”
3	which plan will give me access to a gp?	I'm sorry, I don't have a benefit by that name	NLU ENTITY ERROR “A gp”was extracted by DietClassifier as a *pbi* and not matched with ontology. It would have been matched as a benefit_tag
4	which plan is cheapest?	This covers fees incurred when having your hearing tested.	NLU INTENT ERROR DietClassifier predicts intent of *ask_faq_question* with more confidence than *ask_pbi_or_benefit_tag* Chatbot finds a *pbi* and gives its FAQ definition
5	Which plan is better for good cover and a good price?	I didn't understand the benefit you are asking for.	NLU ENTITY ERROR DietClassifier extracts “and a good price” as a *benefit_tag* Best match is “A and e” but confidence is below the threshold
6	what is the accomodation in plan A?	This covers fees for receiving vaccines.	NLU INTENT ERROR DietClassifier predicts intent of *ask_faq_question* with confidence 0.99 and extracts “accomodation” as the *faq_term* entity There is no *faq_term* in the ontology so a *pbi* is looked for and “vaccination” is the best match

#	Participant message	Response	Error cause
7	Artificial eye	I didn't understand the benefit you are asking for.	DietClassifier predicts intent of *ask_pbi_or_benefit_tag* and extracts a *benefit_tag* entity of “artificial eye”. There is no match in the ontology
8	Inpatient accomodation	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment	DietClassifier predicts intent of *ask_pbi_or_benefit_tag* and extracts a *benefit_tag* entity of “Inpatient accomodation”. Closest match in ontology is “cancer accomodation”
9	semi private room	Concerning a private room in a private hospital Plan A has coverage with once-off €300 excess per family per year and shortfall of €250 per night. Full cover in the Mater Private Cork; No excess/shortfall applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has cover up to the semi-private rate and cover for 100% of technical charges. Full cover in some private hospitals. Ask your insurer for details	NLU error DietClassifier correctly predicts the intent as *ask_pbi_or_benefit_tag* but extracts 2 distinct benefit_tag entities , “*semi*” and “*private room*”
10	comprehensive	I didn't understand the benefit you are asking for	User’s most important benefit was “comprehensive cover” This concept is not in the ontology. The DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “comprehensive” as a *benefit_tag* entity
11	Urgent care	I didn't understand the benefit you are asking for	DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “urgent care” as a *benefit_tag* entity. This is not defined in the ontology
12	Health insurance	I didn't understand the benefit you are asking for	DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “health insurance” as a *benefit_tag* entity. This is not defined in the ontology
13	tell me about plan a		NLU ENTITY ERROR

#	Participant message	Response	Error cause
			Blank response from chatbot DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “plan” as a plan entity and “a” as a benefit_tag entity This is not defined in the ontology Note that “plan A” with capitalised A would be detected by DietClassifier as a complete plan entity
14	Which plan is better for Access to digital doctor?	I didn't understand the benefit you are asking for	ONTOLOGY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “digital doctor” as a *benefit_tag* entity. This is not defined in the ontology
15	Which plan is better for certain cardiac procedures at the blackrock clinic?	Cardiac procedures are lists of heart related procedures, for which the insurer has negotiated a set price with the hospital. This benefit relates to those procedures at	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* intent and extracts the *pbi* entity The user is asking for a comparison , not an FAQ type explanation
16	Which plan is better for a semi private room in private hospitals?	This means a bed in a private hospital which is in a room shared with no more than 5 other patients	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* intent and extracts the *pbi* entity The user is asking for a comparison , not an FAQ type explanation
17	Which plan is better for The quickest access to the best medical care?	OK. I’ll stop	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “The quickest access to the best medical care ” as a *pbi* entity. This is not defined in the ontology
18	routine dental	Sorry! I didn't catch which benefit you asked for.	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “routine dental” as a benefit_tag entity. This is not defined as a *benefit_tag* in the ontology
18	individual excess	Sorry! I didn't catch which benefit you asked for.	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent and extracts “individual excess” as a *benefit_tag* entity. This is not defined as a *benefit_tag* in the ontology. It was defined as a

#	Participant message	Response	Error cause
			pbi
20	what plans are available?	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* intent and extracts “available” as the *faq_term* entity Ontology has no match
21	what is covered	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* intent and extracts “covered” as the *faq_term* entity Ontology has no match
22	how much is excess for plan a and b	Sorry! I didn't catch which benefit you asked for.	NLU ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent but no match in ontology for “plan a and b”
23	is psycho therapy covered in plan a and b	Sorry! I didn't catch which benefit you asked for.	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent but no match in ontology for “psycho therapy”
24	how is cancer covered?	Convalescence is the gradual recovery of health and strength after illness.	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* and has no faq_term for “cancer”.The action matches “convalescence” with confidence 0.53 and responds with the faqExplanation from the *pbi*, convalescence
25	Which plan is better for a semi private room in private hospitals?	This means a bed in a private hospital which is in a room shared with no more than 5 other patients	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* and responds with the faqExplanation from the *pbi* “Private Hospital Semi-Private Room”
26	Which plan is better for Pre Existing condition coverage?	I didn't understand the benefit you are asking for.	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent but no match in ontology for “*pre existing cover”*
27	Are both plans the same?		NLU INTENT ERROR Chatbot response was blank *nlu_fallback* intent was predicted then TEDPolicy predicted the number_pbi_form as the next action

#	Participant message	Response	Error cause
28	Which plan has a better range of hospitals?	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent with a *pbi* and “range of hospitals” was matched in ontology to “overseas hospitals”
29	Which plan is better for Its the range of hospitals and what illnesses the plan covers?	Apologies! I didn't understand that last request.	ONTOLOGY ENTITY ERROR NLU ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag* intent with 2 *pbi* entities, “the range of hospitals and” and “illnesses the”. Rasa uses the highest confidence *pbi* to attempt to fill the slot. The custom validation function rejects the value, finding no suitable match in the ontology
30	What plan gives free glasses?	I didn't understand the benefit you are asking for.	ONTOLOGY ENTITY ERROR DietClassifier predicts *ask_pbi_or_benefit_tag*. The entity “free glasses” is not matched in the ontology
31	What eye care can I get	Which health benefit interests you? You can say,for example, "tell me about cardiac"	NLU INTENT ERROR DietClassifier predicts the intent of *state_interest_in_benefits* with a confidence of 0.7
32	What excess for one person is there	This covers fees for nursing care received in the home.	NLU INTENT ERROR DietClassifier predicts *ask_faq_question* intent and gives a faqExplanation from the current *pbi*

Appendix: Chatbot Multi-turn Errors

Some errors could only be understood by reviewing a sequence of conversation turns.

Participant #1

had three separate sequences of errors but still gave a really high SUS score
SUS 82.5
First sequence

User / bot	Message	Explanation
user	which plan is better for optical policies?	
bot	A category containing these 2 individual benefits exists for public hospital; 1) a private room in a public hospital, 2) a semi-private room in a public hospital Type a number between 1 and 2 for a benefit from the category public hospital	User / bot Message Explanation
user	!	DietClassifier predicts nlu_fallback intent
bot	I'm sorry. I didn't quite understand what you said.	
user	1	
bot	I'm sorry. I didn't quite understand what you said.	DietClassifier correctly predicts select_number_pbi intent and correctly extracts “1” as the selection but then the form is cancelled by TED. TEDPolicy predicts action_execution_rejected with confidence 0.979 ActionExecutionRejected RulePolicy then predicts action_default_fallback with confidence 0.4 Custom action action_default_fallback was configured in From config.yml - name: RulePolicy core_fallback_threshold: 0.4 core_fallback_action_name: "action_default_fallback" enable_fallback_prediction: True
user	plan 1	

User / bot	Message	Explanation
bot	I'm sorry. I didn't quite understand what you said.	

Second sequence

User / bot	Message	Explanation
user	**can you tell me about family excess**	
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person	
user	does plan A mean per person	User seems to be asking about the pbi stated in the **previous response** and whether it relates to a single person
bot	I didn't understand the benefit you are asking for.	

Third sequence

User / bot	Message	Explanation
user	which plan will help me see a doctor fast	
bot	Sorry! I didn't catch which benefit you asked for.	“A doctor fast” is extracted as a benefit_tag entity Rasa calls custom validation method which matches “doctor” with confidence of 0.63 The threshold is configured as 65 so this value is rejected TEDPolicy predicts action_state_pbis_for_choosing And this custom action sends the response “utter_no_benefit” since benefit_tag_slot is empty
user	instant access to doctor	
bot	Apologies! I didn't understand that last request.	“Instant access to doctor” is extracted by DietClassifier as a pbi entity and the custom validation method, validate_pbi_slot / validate_text_pbi_slot (inside the form) , is invoked. This method found a pbi “"Other Day To Day Practitioners" with confidence of 0.48

		The validation method sends the response message
user	which plan will give me access to a gp	
bot	Apologies! I didn't understand that last request.	“ a gp” is extracted by DIETClassifier as a **pbi** Valildation method tries to match but confidence is too low for a pbi Note that a benefit_tag would have been matched best_benefit_tag_alias('a gp') {'best_match': 'gp', 'confidence': 67, 'benefit_tag': 'gp', 'is_alias': False}
user	gp	
bot	Concerning GP Visits Plan A has 24 hr GP Line; 24 Hr Nurseline; Three free GP live online consultations and Plan B has cover for €30 per visit. Practise Nurse is covered for 1 visit at €20. Full cover is provided for 12 video calls to your insurer's online doctor.	“Gp” is extracted by DietClassifier as a benefit_tag and matched with confidence 1.0

Participant #3

First sequence

User / bot	Message	Explanation
user	tell me about price	
bot	Concerning Annual price per adult Plan A has €1583.37 and Plan B has €2238.12	Correct response
user	is this per person or individual	The user message refers to the **previous response** pbi as “this”, The intent of the user is to get further information about the pbi value
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	

Second sequence

User /bot	Message	Explanation
user	what is the optical options	
bot	This covers fees for optical treatment	DietClassifier predicts ask_faq_question intent and faq_term entity of “optical options”
user	what is covered	The user seems to be referring to the **previous response** for “optical options”
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	Dietclassifer predicts intent ask_faq_question with no entities

Participant #5

First sequence

User/bot	Message	Explanation
user	Which plan is better for overseas hospital?	
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover	
user	Is there any other difference between plans for overseas hospital?	The user seems to be referring to the **previous response** and looking for “other” differences
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover	The chatbot identifies the same intent and entity as before and responds with exactly the same answer

Second sequence

User/bot	Message	Explanation
user	Which plan is better for family excess	
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person	
user	Is plan A per person	The user message seems to refer to the value of excess for plan A in the **previous response** The user’s question can probably be restated as “Is the value of family excess stated for plan A as €250 calculated per person ?”
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	

Participant #6

First sequence

User/bot	Message	Explanation
user	Which plan is better for convalescence?	
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation	
user	Why such a small €1 difference?	The user message seems to refer to the value of convalescence for both plans in the **previous response** and comments on the difference The intent is predicted as select_number_pbi with a low confidence of 0.47 and no entities extracted TED predicts the action as number_pbi_form This form does not have a prompt for the number and so nothing is said in the chatbot response

Participant #7

First sequence

User/bot	Message	Explanation
user	what is cancer accomodation support	
bot	Cancer accomodation support is payment for an overnight stay when receiving cancer treatment on a given day.	
user	how much is covered	
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	The user message seems to refer to the pbi he inquired about in his **previous request ,** namely “cancer accomodation support”
user	how much is covered per night	
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	DietClassifier predicts the intent ask_faq_question with confidence of 0.79

Participant #8

First sequence

User/bot	Message	Explanation
user	Which plan is better for a semi private room in private hospitals?	
bot	This means a bed in a private hospital which is in a room shared with no more than 5 other patients	
user	WHich is better?	User seems to be asking a question following on from their **previous request**
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	

Second sequence

User/bot	Message	Explanation
user	Which plan is better for a semi private room in private hospitals?	
bot	This means a bed in a private hospital which is in a room shared with no more than 5 other patients	DietClassifier predicts the intent ask_faq_question. Bot responds with the faqExplanation for the pbi
user	WHich is better?	User seems to be asking a question following on from their **previous request**
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain	

Participant #9

First sequence

User/bot	Message	Explanation
user	Which plan is better for gender	
bot	There are 2 benefits directly related to gender; Type the number or type the benefit name to hear more 1 for Gender Reassignment Surgery and 2 for HRT for Gender Dysphoria	
user	Any benefits for a women	The user seems to refer to the **previous response**

bot	Apologies! I didn't understand that last request.	Intent is predicted as ask_pbi_or_benefit_tag. DietClassifier extracts “any” and “a woman” as benefit_tag entities

Participant #11

First sequence

User/bot	Message	Explanation
user	Can I have information on cancer support	
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment	
user	Is there any other cancer supports	The user is looking for additional information concerning the **previous response**
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment	

Appendix : Goal Times

The following were the times taken by each participant to answer each question on the quiz for the first execution.

Questions	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11
1	2.23	1.40	2.07	0.97	1.80	2.57	2.05	1.17	1.47	4.77	6.73
2	2.12	0.73	1.33	0.28	0.77	0.75	1.50	0.35	0.52	1.12	7.98
3	0.67	0.68	0.42	0.33	0.37	1.25	1.97	0.82	0.65	0.65	1.13
4	1.83	0.53	0.83	0.58	1.20	0.97	1.38	0.60	0.57	1.32	1.70
5	3.42	1.12	0.62	1.43	1.02	1.55	1.28	0.82	0.97	0.75	1.77
6	1.63	0.98	1.12	0.80	1.07	1.27	1.95	0.98	1.25	1.25	1.10
7	4.57	0.80	1.10	2.68	1.47	0.80	1.87	0.45	1.27	1.30	4.42
8	1.95	0.95	1.88	1.83	0.45		1.15		0.87	0.70	11.17
sum	18.42	7.20	9.37	8.92	8.13	9.15	13.15	5.18	7.55	11.85	36.00

For the second execution the times were as follows

Question	#1	#2	#3	#4	#5	#6
1	1.25	3.63	5.15	2.73	1.40	1.62
2	0.70	4.70	3.00	0.62	2.05	3.77
3	1.68	3.45	1.58	1.33	1.48	0.90
4	1.28	1.92	1.75	0.82	2.02	1.25
5	1.30	3.58	4.52	1.93	2.18	1.60
6	4.02	3.80	5.22	4.23	2.15	1.67
7	6.58	4.63	7.80	4.70	2.40	2.57
8	5.42	4.22	9.22	1.53	1.98	5.17
Sum	22.23	29.93	38.24	17.89	15.66	18.55

Appendix: Conversation Messages for Study 1

Participant #1

user	what plan has a better convalescence policy?
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation
user	which plan is better for optical policies?
bot	A category containing these 2 individual benefits exists for public hospital; 1) a private room in a public hospital, 2) a semi-private room in a public hospital Type a number between 1 and 2 for a benefit from the category public hospital
user	!
bot	I'm sorry. I didn't quite understand what you said.
user	1
bot	I'm sorry. I didn't quite understand what you said.
user	plan 1
bot	I'm sorry. I didn't quite understand what you said.
user	**can you tell me about vaccinations**
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover
user	**can you tell me about family excess**
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	does plan A mean per person
bot	I didn't understand the benefit you are asking for.
user	which plan is better for A and E
bot	I didn't understand the benefit you are asking for.
user	**Which plan is better for accident and emergency?**
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.

user	**which plan is better for psychology**
bot	A category containing these 2 individual benefits exists for psychology; 1) Psycho-Oncology Counselling, 2) an employee assistance programme Type a number between 1 and 2 for a benefit from the category psychology
user	**1**
bot	Concerning Psycho-Oncology Counselling Plan A has no cover and Plan B has €50 cover per visit for 10 visits
user	**which plan is better for cardiac**
bot	A category containing these 4 individual benefits exists for cardiac; 1) Health Screening, 2) The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures, 3) approved radiology centres, 4) non-approved radiology centres Type a number between 1 and 4 for a benefit from the category cardiac
user	**1**
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	**2**
bot	Concerning The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures Plan A has full cover in the Beacon Hospital with once-off €300 excess per family per year. Reduced cardiac list coverage in the Mater Private and Blackrock Clinic with no excess; Ask your insurer for details and Plan B has 90% cover
user	**3**
bot	Concerning approved radiology centres Plan A has has full cover for MRI, CT and PET-CT scans in direct payment centres. The exception is Cardiac CT's which are covered by non-direct payment terms. Cover up to €500 is provided in non direct payment centres and this is subject to outpatient excess and Plan B has full cover for CT, PET-CT and MRI. There is no excess for category 1. There

	is €125 excess for category 2. Refer to your insurer's rules for more information
user	**4**
bot	Concerning non-approved radiology centres Plan A has no cover and Plan B has no cover
user	which plan will help me see a doctor fast
bot	Sorry! I didn't catch which benefit you asked for.
user	instant access to doctor
bot	Apologies! I didn't understand that last request.
user	which plan will give me access to a gp
bot	Apologies! I didn't understand that last request.
user	gp
bot	Concerning GP Visits Plan A has 24 hr GP Line; 24 Hr Nurseline; Three free GP live online consultations and Plan B has cover for €30 per visit. Practise Nurse is covered for 1 visit at €20. Full cover is provided for 12 video calls to your insurer's online doctor.

Participant #2

user	Which plan is better for cancer accomodation support
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	Which plan is better for overseas hospital
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover
user	Which plan is better for vaccinations
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover
user	Which plan is better for individual excess
bot	Concerning Individual Excess Plan A has €150 and Plan B has €100
user	Which plan is better for a semi private room in private hospitals?

bot	Concerning a semi-private room in a private hospital Plan A has coverage with once-off €300 excess per family per year. No excess applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has full cover but 90% cover for certain cardiac and special procedures in the Hermitage and the Galway Clinic
user	Which plan is better for psychology
bot	A category containing these 2 individual benefits exists for psychology; 1) Psycho-Oncology Counselling, 2) an employee assistance programme Type a number between 1 and 2 for a benefit from the category psychology
user	Which plan is better for radiology
bot	A category containing these 5 individual benefits exists for radiology (out-patient radiology); 1) Health Screening, 2) Radiology Consultants' Fees, 3) Radiology Test Fees, 4) approved radiology centres, 5) non-approved radiology centres Type a number between 1 and 5 for a benefit from the category radiology (out-patient radiology)
user	hospital cover
bot	A category containing these 12 individual benefits exists for hospital stay (inpatient); 1) Cancer Accommodation Support, 2) Convalescence, 3) Maternity Hospital Costs Up To 3 Nights, 4) The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures, 5) a private room in a private hospital, 6) a private room in a public hospital, 7) a semi-private room in a private hospital, 8) a semi-private room in a public hospital, 9) day case at high-tech hospitals, 10) day case at private hospitals, 11) other procedures at high-tech hospitals, 12) special procedures at high-tech hospitals Type a number between 1 and 12 for a benefit from the category hospital stay (inpatient)

Participant #3

user	which plan is cheapest
bot	This covers fees incurred when having your hearing tested.
user	Which plan is better for good cover and a good price
bot	I didn't understand the benefit you are asking for.
bot	Concerning a hearing test Plan A has no cover and Plan B has no cover
user	help
bot	- You can ask me; - to "list all benefit categories" - about a particular benefit ,for example, "tell me about physiotherapists" - about a category of benefits ,for example, "tell me about cancer". - about a confusing term, for example, "explain excess" - to list all terms that I can explain

user	tell me about price
bot	Concerning Annual price per adult Plan A has €1583.37 and Plan B has €2238.12
user	is this per person or individual
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	cancer
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	what is the difference in convalescence
bot	Convalescence is the gradual recovery of health and strength after illness
user	which service offers the best convalesence care
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation
user	what is the accomodation in plan A
bot	This covers fees for receiving vaccines.
user	what is the optical options
bot	This covers fees for optical treatment
user	what is covered
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	artificial eye
bot	I didn't understand the benefit you are asking for.
bot	Concerning Optical Plan A has no cover and Plan B has cover for eye tests of €30 in each 24 month period. Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	hearing
bot	Concerning a hearing test Plan A has no cover and Plan B has no cover
user	family
bot	Concerning Family Excess Plan A has €250 and

	Plan B has €100 excess per person
user	what is covered
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	terms
bot	There are 28 terms which I can explain as follows; agreed charges, alternative medicine, cardiac procedures, convalescence, day case, ECG, excess, family excess, gender dysphoria, health coach, high tech, individual excess, inpatient, MRI, out patient excess, outpatient, outpatient policy limit, partner assistance, PET-CT, pre-existing condition, procedures other than cardiac and special, repatriation, schedule of benefits, semi private, shortfall, special procedures, surgery not available, technologically advanced hospitals,
user	
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.

Participant #4

user	cancer
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	Inpatient accomodation
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	Optical
bot	Concerning Optical Plan A has no cover and Plan B has cover for eye tests of €30 in each 24 month period. Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	Vaccinations
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover

user	Family excess
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	semi private room
bot	Concerning a private room in a private hospital Plan A has coverage with once-off €300 excess per family per year and shortfall of €250 per night. Full cover in the Mater Private Cork; No excess/shortfall applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has cover up to the semi-private rate and cover for 100% of technical charges. Full cover in some private hospitals. Ask your insurer for details
user	counselling
bot	A category containing these 2 individual benefits exists for counselling (psychology); 1) Psycho-Oncology Counselling, 2) an employee assistance programme Type a number between 1 and 2 for a benefit from the category counselling (psychology)
user	2
bot	Concerning an employee assistance programme Plan A has no cover and Plan B has no cover
user	1
bot	Concerning Psycho-Oncology Counselling Plan A has no cover and Plan B has €50 cover per visit for 10 visits
user	radiology
bot	A category containing these 5 individual benefits exists for radiology (out-patient radiology); 1) Health Screening, 2) Radiology Consultants' Fees, 3) Radiology Test Fees, 4) approved radiology centres, 5) non-approved radiology centres Type a number between 1 and 5 for a benefit from the category radiology (out-patient radiology)
user	1
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no

	excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	2
bot	Concerning Radiology Consultants' Fees Plan A has full cover subject to outpatient excess and Plan B has €60 cover per procedure
user	3
bot	Concerning Radiology Test Fees Plan A has cover up to €500, subject to out-patient excess and Plan B has 50% of charges up to a maximum of €850 per year
user	4
bot	Concerning approved radiology centres Plan A has has full cover for MRI, CT and PET-CT scans in direct payment centres. The exception is Cardiac CT's which are covered by non-direct payment terms. Cover up to €500 is provided in non direct payment centres and this is subject to outpatient excess and Plan B has full cover for CT, PET-CT and MRI. There is no excess for category 1. There is €125 excess for category 2. Refer to your insurer's rules for more information
user	5
bot	Concerning non-approved radiology centres Plan A has no cover and Plan B has no cover
user	comprehensive
bot	I didn't understand the benefit you are asking for.
user	urgent care
bot	Sorry! I didn't catch which benefit you asked for.
user	health insurance
bot	Sorry! I didn't catch which benefit you asked for.
user	consultant
bot	A category containing these 3 individual benefits exists for consultant; 1) Consultant Fees, 2) Delivery Consultant Fees, 3) Radiology Consultants' Fees Type a number between 1 and 3 for a benefit from the category consultant
user	family cover
bot	I didn't understand the benefit you are asking for.
user	1
bot	Sorry, there is no list to choose from

user	help
bot	- You can ask me; - to "list all benefit categories" - about a particular benefit ,for example, "tell me about physiotherapists" - about a category of benefits ,for example, "tell me about cancer". - about a confusing term, for example, "explain excess" - to list all terms that I can explain
user	list all benefit categories
bot	There are 61 benefit categories as follows; Prices per annum, a and e, alternative medicine, breastfeeding, cancer, cancer accommodation, cardiac, child healthcare, chiropody, chiropractor, consultant, convalescence, day case, dental, dietitian, emergency, employee assistance, excess, family excess, fertility, gender, gender reassignment surgery, gp, health screening, healthcoach, hearing, high tech hospital, home nursing, hospital overseas, hrt for gender dyshporia, individual excess, inpatient, liver disease, maternity, occupational therapy, optical, orthoptistry, osteopath, out-patient radiology, outpatient, overseas, parent with child, physio, podiatry, policy limit, post natal, pre natal, price, private hospital, private room, psychiatry, psychology, public hospital, radiology, reflexology, repatriation, semi-private room, special procedures, speech therapy, surgery, vaccination,
user	oncology
bot	A category containing these 4 individual benefits exists for oncology (cancer); 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category oncology (cancer)
user	1
bot	Concerning Aids and accessories for cancer patients Plan A has cover for 1 hairpiece per year following cancer treatment. Cover also for first breast prosthesis following am inpatient claim. The second prosthesis would be assessed in accordance with the surgical appliance list. A full refund is provided for 1 sleeping cap per membership year. Eyebrow tattooing (following cancer treatment) is covered for up to €60 once per membership year and Plan B has cover for Wig/hairpiece, mastectomy bra, swimsuit, surgical prosthesis. There is no excess, but the amount is subject to the maximum for medical and surgical appliances of €3,200. Cover for Eye tests of €30 in each 24 month period . Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period

Participant #5

user	Which plan is better for convalescence
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation

user	help
bot	- You can ask me; - to "list all benefit categories" - about a particular benefit ,for example, "tell me about physiotherapists" - about a category of benefits ,for example, "tell me about cancer". - about a confusing term, for example, "explain excess" - to list all terms that I can explain
user	tell me about plan A
user	Which plan is better for overseas hospital?
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover
user	Is there any other difference between plans for overseas hospital?
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover
user	Which plan is better for vaccinations?
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover
user	Which plan is better for family excess
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	Is plan A per person
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	Which plan is better for accident and emergency
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.
user	Which plan is better for gender?
bot	A category containing these 2 individual benefits exists for gender; 1) Gender Reassignment Surgery, 2) HRT for Gender Dysphoria Type a number between 1 and 2 for a benefit from the category gender
user	1
bot	Concerning Gender Reassignment Surgery Plan A has cover up to €100,000 subject to prior approval and Plan B has cover for medically necessary surgeries subject to pre-approval up to the level of cover on the plan

user	2
bot	Concerning HRT for Gender Dysphoria Plan A has no cover and Plan B has no cover
user	Which plan is better for cancer
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	1
bot	Concerning Aids and accessories for cancer patients Plan A has cover for 1 hairpiece per year following cancer treatment. Cover also for first breast prosthesis following am inpatient claim. The second prosthesis would be assessed in accordance with the surgical appliance list. A full refund is provided for 1 sleeping cap per membership year. Eyebrow tattooing (following cancer treatment) is covered for up to €60 once per membership year and Plan B has cover for Wig/hairpiece, mastectomy bra, swimsuit, surgical prosthesis. There is no excess, but the amount is subject to the maximum for medical and surgical appliances of €3,200. Cover for Eye tests of €30 in each 24 month period . Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	2
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	3
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an

	approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	4
bot	Concerning Psycho-Oncology Counselling Plan A has no cover and Plan B has €50 cover per visit for 10 visits
user	Which plan is better for Access to digital doctor
bot	I didn't understand the benefit you are asking for.
bot	Apologies! I didn't understand that last request.

Participant #6

bot	There are 4 benefits directly related to cancer; Type the number or type the benefit name to hear more 1 for Aids and accessories for cancer patients and 2 for Cancer Accommodation Support and 3 for Health Screening and 4 for Psycho-Oncology Counselling
user	1
bot	Concerning Aids and accessories for cancer patients Plan A has cover for 1 hairpiece per year following cancer treatment. Cover also for first breast prosthesis following am inpatient claim. The second prosthesis would be assessed in accordance with the surgical appliance list. A full refund is provided for 1 sleeping cap per membership year. Eyebrow tattooing (following cancer treatment) is covered for up to €60 once per membership year and Plan B has cover for Wig/hairpiece, mastectomy bra, swimsuit, surgical prosthesis. There is no excess, but the amount is subject to the maximum for medical and surgical appliances of €3,200. Cover for Eye tests of €30 in each 24 month period . Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	convalescence
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation
user	Which plan is better for convalescence?
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation
user	Why such a small €1 difference?
user	Which plan is better for employee assistance?
bot	Concerning an employee assistance programme Plan A has no cover and Plan B has no cover

user	Which plan is better for certain cardiac procedures at the blackrock clinic?
bot	Cardiac procedures are lists of heart related procedures, for which the insurer has negotiated a set price with the hospital. This benefit relates to those procedures at the specified high-tech hospitals
user	Which plan is better for a semi private room in private hospitals?
bot	This means a bed in a private hospital which is in a room shared with no more than 5 other patients
user	Which plan is better for home nursing?
bot	There are 2 benefits directly related to home nursing; Type the number or type the benefit name to hear more 1 for Child Home Nursing and 2 for Home Nursing
bot	Apologies! I didn't understand that last request.
user	2
bot	Concerning Home Nursing Plan A has €40 per day for 40 days and Plan B has full cover for your insurer's Homecare benefit for certain procedures and upon referral from certain hospitals
user	Which plan is better for cancer?
bot	There are 4 benefits directly related to cancer; Type the number or type the benefit name to hear more 1 for Aids and accessories for cancer patients and 2 for Cancer Accommodation Support and 3 for Health Screening and 4 for Psycho-Oncology Counselling
user	3
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	Which plan is better for The quickest access to the best medical care>
bot	OK, I'll stop

user	routine dental
bot	Sorry! I didn't catch which benefit you asked for.
user	dental
bot	There are 2 benefits directly related to dental; Type the number or type the benefit name to hear more 1 for Emergency Dental Care and 2 for Non Emergency Routine Dental Care
user	routine dental care
bot	Sorry! I didn't catch which benefit you asked for.
user	individual excess
bot	Sorry! I didn't catch which benefit you asked for.
user	excess
bot	There are 2 benefits directly related to excess; Type the number or type the benefit name to hear more 1 for Family Excess and 2 for Individual Excess
user	family excess
bot	Sorry! I didn't catch which benefit you asked for.

Participant #7

user	hi
user	what is cancer accomodation support
bot	Cancer accomodation support is payment for an overnight stay when receiving cancer treatment on a given day.
user	how much is covered
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	how much is covered per night
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	what plans are available
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	what is covered
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	list the terms
bot	There are 28 terms which I can explain as follows; agreed charges, alternative medicine, cardiac procedures, convalescence, day case,

	ECG, excess, family excess, gender dysphoria, health coach, high tech, individual excess, inpatient, MRI, out patient excess, outpatient, outpatient policy limit, partner assistance, PET-CT, pre-existing condition, procedures other than cardiac and special, repatriation, schedule of benefits, semi private, shortfall, special procedures, surgery not available, technologically advanced hospitals,
user	optical
bot	Concerning Optical Plan A has no cover and Plan B has cover for eye tests of €30 in each 24 month period. Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	which plan covers vaccinations
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover
user	which covers maternity
bot	A category containing these 13 individual benefits exists for maternity; 1) Assistance for partners during maternity, 2) Breastfeeding Consultancy, 3) Child Healthcare Benefit, 4) Child Home Nursing, 5) Delivery Consultant Fees, 6) Home Births, 7) Maternity Hospital Costs Up To 3 Nights, 8) Other Maternity Benefits, 9) Outpatient Maternity Consultant Care, 10) Parent with child, 11) Postnatal Home Help, 12) Postnatal Home Nursing, 13) Pre & Postnatal Care Type a number between 1 and 13 for a benefit from the category maternity
user	2
bot	Concerning Breastfeeding Consultancy Plan A has no cover and Plan B has no cover
user	5
bot	Concerning Delivery Consultant Fees Plan A has full cover up to agreed charges incurred on day of delivery for participating consultants and Plan B has full cover up to agreed charges incurred on day of delivery for participating consultants
user	3
bot	Concerning Child Healthcare Benefit Plan A has no cover and Plan B has 50% cover at your insurer's paediatric clinic for an initial consultant visit. Also 50% cover for follow up paediatric treatment and services (including lactation consultant, dietician, ultrasound, blood tests and x-ray).
user	what is family excess
bot	Family excess applies to where there is more than one person on the policy. Individual excess applies where there is only one person on the policy and the
user	how much is excess for plan a and b
bot	Sorry! I didn't catch which benefit you asked for.
user	family excess

bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	are accident and emergency covered
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.[1]
user	is psycho therapy covered in plan a and b
bot	Sorry! I didn't catch which benefit you asked for.
bot	Apologies! I didn't understand that last request.
user	psychology
user	is psychology covered
bot	A category containing these 2 individual benefits exists for psychology; 1) Psycho-Oncology Counselling, 2) an employee assistance programme Type a number between 1 and 2 for a benefit from the category psychology
user	2
bot	Concerning an employee assistance programme Plan A has no cover and Plan B has no cover
user	1
bot	Concerning Psycho-Oncology Counselling Plan A has no cover and Plan B has €50 cover per visit for 10 visits
user	how is cancer covered
bot	Convalescence is the gradual recovery of health and strength after illness.
user	cover for cancer
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	3
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied

[1]

	for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	tell me about dental cover
bot	A category containing these 2 individual benefits exists for dental; 1) Emergency Dental Care, 2) Routine dental Type a number between 1 and 2 for a benefit from the category dental
user	2
bot	Concerning Routine dental Plan A has no cover and Plan B has €20 per visit for 7 visits

Participant #8

user	Which plan is better for convalescence?
bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation
user	Which plan is better for employee assistance?
bot	Concerning an employee assistance programme Plan A has no cover and Plan B has no cover
user	Which plan is better for private room in a private hospital?
bot	Concerning a private room in a private hospital Plan A has coverage with once-off €300 excess per family per year and shortfall of €250 per night. Full cover in the Mater Private Cork; No excess/shortfall applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has cover up to the semi-private rate and cover for 100% of technical charges. Full cover in some private hospitals. Ask your insurer for details
user	Which plan is better for a semi private room in private hospitals?
bot	This means a bed in a private hospital which is in a room shared with no more than 5 other patients
user	WHich is better?
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I

	am able to explain
user	Which plan is better for accident and emergency?
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.
user	Which plan is better for cardiac?
bot	There are 4 benefits directly related to cardiac; Type the number or type the benefit name to hear more 1 for Health Screening and 2 for The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures and 3 for approved radiology centres and 4 for non-approved radiology centres
user	1
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	Which plan is better for Pre Existing condition coverage?
bot	I didn't understand the benefit you are asking for.
bot	Apologies! I didn't understand that last request.

Participant #9

user	Which plan is better for convalescence?

bot	Concerning Convalescence Plan A has €50 per day for 14 days and Plan B has €51 per night for 14 nights in semi private and private room accommodation
user	**Does Plan A have a private room?**
bot	Concerning a private room in a private hospital Plan A has coverage with once-off €300 excess per family per year and shortfall of €250 per night. Full cover in the Mater Private Cork; No excess/shortfall applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has cover up to the semi-private rate and cover for 100% of technical charges. Full cover in some private hospitals. Ask your insurer for details
user	**So Plan B has no excess?**
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	**Which plan is better for overseas hospital?**
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover
user	what excess is better?
bot	I'm sorry but I don't have an explanation for that. You can ask me to list the terms that I am able to explain
user	list the terms
bot	There are 28 terms which I can explain as follows; agreed charges, alternative medicine, cardiac procedures, convalescence, day case, ECG, excess, family excess, gender dysphoria, health coach, high tech, individual excess, inpatient, MRI, out patient excess, outpatient, outpatient policy limit, partner assistance, PET-CT, pre-existing condition, procedures other than cardiac and special, repatriation, schedule of benefits, semi private, shortfall, special procedures, surgery not available, technologically advanced hospitals,
user	are both plans the same
user	**Which plan is better for vaccination**
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover
user	**so vaccination are no covered**
bot	Concerning Vaccinations Plan A has no cover and Plan B has no cover
user	**better for family excess**
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	does plan b have a limit

bot	Sorry! I didn't catch which benefit you asked for.
user	**does plan b have a limit on there excess**
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	**plan is better for accident and emergency?**
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.
user	does Plan A cover A&E
bot	Concerning A&E Charge Plan A has €20 cover per visit; Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres and Plan B has €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.
user	**Which plan is better for gender**
bot	There are 2 benefits directly related to gender; Type the number or type the benefit name to hear more 1 for Gender Reassignment Surgery and 2 for HRT for Gender Dysphoria
user	Any benefits for a women
bot	Apologies! I didn't understand that last request.
user	**Which plan is better for cardiac**
bot	There are 4 benefits directly related to cardiac; Type the number or type the benefit name to hear more 1 for Health Screening and 2 for The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures and 3 for approved radiology centres and 4 for non-approved radiology centres
user	**is there a difference between Plan A or Plan B**
bot	Apologies! I didn't understand that last request.
user	for Health Screening and
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2 years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full

	cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	**The Blackrock Clinic**
bot	Apologies! I didn't understand that last request.
user	**Which plan is better for Its the range of hospitals and what illnesses the plan covers**
bot	Sorry! I didn't catch which benefit you asked for.
bot	Apologies! I didn't understand that last request.
user	**Which plan has a better range of hospitals**
bot	Concerning overseas hospital bills Plan A has €100,000 cover and Plan B has €100,000 cover
user	**Which plan is better for Its the range of hospitals and what illnesses the plan covers?**
bot	Apologies! I didn't understand that last request.

Participant #10

user	cancer
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	4
bot	Concerning Psycho-Oncology Counselling Plan A has no cover and Plan B has €50 cover per visit for 10 visits
user	3
bot	Concerning Health Screening Plan A has Full cover for Mammograms and Dexa scans in certain centres. Women's Cancer Screening is covered up to €30. Men's Cancer Screening is covered up to €20. Full cover for 'HeartBeat' cardiac screening with no excess; Home testing kits are covered up to €20. Health Screening up to €150 every 2

	years is paid directly to centre and Plan B has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre. There is full cover for Mammograms and Dexa Scans every 2 years. There is no excess applied for the above. The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health. This is full covered every 2 years. There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening. The Medfit Cardiac Care programme is covered for €250 every 2 years. There is Urgent Cardiac Care benefit of €250 per year. The Medfit Cardiac Rehabilitation programme is covered for€250 per year. Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation. Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits
user	2
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	1
bot	Concerning Aids and accessories for cancer patients Plan A has cover for 1 hairpiece per year following cancer treatment. Cover also for first breast prosthesis following am inpatient claim. The second prosthesis would be assessed in accordance with the surgical appliance list. A full refund is provided for 1 sleeping cap per membership year. Eyebrow tattooing (following cancer treatment) is covered for up to €60 once per membership year and Plan B has cover for Wig/hairpiece, mastectomy bra, swimsuit, surgical prosthesis. There is no excess, but the amount is subject to the maximum for medical and surgical appliances of €3,200. Cover for Eye tests of €30 in each 24 month period . Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	**Which plan is better for cancer accomodation support?**
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	**Which plan is better for optical**
bot	Concerning Optical Plan A has no cover and Plan B has cover for eye tests of €30 in each 24 month period. Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	**Which plan is better for hearing?**
bot	Concerning a hearing test Plan A has no cover and Plan B has no cover
user	Which plan is better for family excess?
bot	Concerning Family Excess Plan A has €250 and

	Plan B has €100 excess per person
user	Which plan is better for a semi private room in private hospitals
bot	Concerning a semi-private room in a private hospital Plan A has coverage with once-off €300 excess per family per year. No excess applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has full cover but 90% cover for certain cardiac and special procedures in the Hermitage and the Galway Clinic
user	**Which plan is better for gender?**
bot	A category containing these 2 individual benefits exists for gender; 1) Gender Reassignment Surgery, 2) HRT for Gender Dysphoria Type a number between 1 and 2 for a benefit from the category gender
user	1
bot	Concerning Gender Reassignment Surgery Plan A has cover up to €100,000 subject to prior approval and Plan B has cover for medically necessary surgeries subject to pre-approval up to the level of cover on the plan
user	**Which plan is better for cardiac**
bot	A category containing these 4 individual benefits exists for cardiac; 1) Health Screening, 2) The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures, 3) approved radiology centres, 4) non-approved radiology centres Type a number between 1 and 4 for a benefit from the category cardiac
user	2
bot	Concerning The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures Plan A has full cover in the Beacon Hospital with once-off €300 excess per family per year. Reduced cardiac list coverage in the Mater Private and Blackrock Clinic with no excess; Ask your insurer for details and Plan B has 90% cover
user	**cardiac cover"**
bot	A category containing these 4 individual benefits exists for cardiac; 1) Health Screening, 2) The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures, 3) approved radiology centres, 4) non-approved radiology centres Type a number between 1 and 4 for a benefit from the category cardiac
user	2
bot	Concerning The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures Plan A has full cover in the Beacon Hospital with once-off €300 excess per family per year. Reduced cardiac list coverage in the Mater Private and Blackrock Clinic with no excess; Ask your insurer for details and Plan B has 90% cover

Participant #11

user	**cancer**
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	**Can I have information on cancer support**
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	Is there any other cancer supports
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	What plan gives free glasses
bot	I didn't understand the benefit you are asking for.
user	What eye care can I get
bot	Which health benefit interests you? You can say,for example, "tell me about cardiac"
user	**Tell me about optical**
user	Optical cover
bot	Concerning Optical Plan A has no cover and Plan B has cover for eye tests of €30 in each 24 month period. Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	**What hearing benefits are there**
bot	This covers fees incurred when having your hearing tested.
user	Which plan
bot	I didn't understand the benefit you are asking for.
bot	Apologies! I didn't understand that last request.
user	Which plan has hearing cover
bot	
user	What excess for one person is there
bot	This covers fees for nursing care received in the home.
user	Excess
user	**Excess cover amounts**
bot	Concerning Family Excess Plan A has €250 and Plan B has €100 excess per person
user	**individual excess**

bot	Concerning Individual Excess Plan A has €150 and Plan B has €100
user	**What cover for a private room**
bot	Concerning a private room in a private hospital Plan A has coverage with once-off €300 excess per family per year and shortfall of €250 per night. Full cover in the Mater Private Cork; No excess/shortfall applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals. Check with your insurer for more details. and Plan B has cover up to the semi-private rate and cover for 100% of technical charges. Full cover in some private hospitals. Ask your insurer for details
user	**psychology cover**
bot	A category containing these 2 individual benefits exists for psychology; 1) Psycho-Oncology Counselling, 2) an employee assistance programme Type a number between 1 and 2 for a benefit from the category psychology
user	**1**
bot	Concerning Psycho-Oncology Counselling Plan A has no cover and Plan B has €50 cover per visit for 10 visits
user	psychology
bot	A category containing these 2 individual benefits exists for psychology; 1) Psycho-Oncology Counselling, 2) an employee assistance programme Type a number between 1 and 2 for a benefit from the category psychology
user	**2**
bot	Concerning an employee assistance programme Plan A has no cover and Plan B has no cover
user	**Cancer cover**
bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	**1**
bot	Concerning Aids and accessories for cancer patients Plan A has cover for 1 hairpiece per year following cancer treatment. Cover also for first breast prosthesis following am inpatient claim. The second prosthesis would be assessed in accordance with the surgical appliance list. A full refund is provided for 1 sleeping cap per membership year. Eyebrow tattooing (following cancer treatment) is covered for up to €60 once per membership year and Plan B has cover for Wig/hairpiece, mastectomy bra, swimsuit, surgical prosthesis. There is no excess, but the amount is subject to the maximum for medical and surgical appliances of €3,200. Cover for Eye tests of €30 in each 24 month period . Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	**cancer**

bot	A category containing these 4 individual benefits exists for cancer; 1) Aids and accessories for cancer patients, 2) Cancer Accommodation Support, 3) Health Screening, 4) Psycho-Oncology Counselling Type a number between 1 and 4 for a benefit from the category cancer
user	**2**
bot	Concerning Cancer Accommodation Support Plan A has no cover and Plan B has €100 for 1 night's accommodation for each treatment
user	**excess**
bot	A category containing these 2 individual benefits exists for excess; 1) Family Excess, 2) Individual Excess Type a number between 1 and 2 for a benefit from the category excess
user	**2**
bot	Concerning Individual Excess Plan A has €150 and Plan B has €100

Appendix : Conversation Messages for Study 2

Note that buttons were implemented for the second execution of tests and button presses are represented in the chatbot messages as intents, for example, `/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}`

Participant #1

user	Which plan is better for cancer accommodation support?
bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment
user	Which plan is better for optical?
bot	Concerning "Optical", **Plan A** has no cover **Plan B** has cover for eye tests of €30 in each 24 month period Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	Which plan is better for vacci ations?
bot	I didn't understand the benefit you are asking for.
bot	Concerning "Vaccinations", **Plan A** has no cover **Plan B** has no cover
user	Which plan is better for vaccinations?
bot	Concerning "Vaccinations", **Plan A** has no cover **Plan B** has no cover
user	Which plan is better for family excess?
bot	Concerning "Family Excess", please be aware that excess applies to the entire family unless stated differently. Less excess means more money back for you when you claim. **Plan A** has €250 **Plan B** has €100 excess per person
user	Which plan is better for accident and emergency?

bot	Concerning "A&E Charge", **Plan A** has €20 cover per visit Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres **Plan B** has €75 for 1 visit (Applies to A&E in Public Hospitals Only) Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit
user	Which plan is better for gender?
bot	"gender" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Gender Reassignment Surgery", **Plan A** has cover up to €100,000 subject to prior approval **Plan B** has cover for medically necessary surgeries subject to pre-approval up to the level of cover on the plan. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"gender"}
bot	"gender" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "HRT for Gender Dysphoria", **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"gender"}
bot	"gender" is a category containing 2 individual benefits. Click on any button for more information.
user	Which plan is better for cardiac?
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"cardiac"}
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures", **Plan A** has full cover in the Beacon Hospital with once-off €300 excess per family per year Reduced cardiac list coverage in the Mater Private and Blackrock Clinic with no excess Ask your insurer for details **Plan B** has 90% cover. Click on the button below to see more benefits from the

	category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"cardiac"}
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}
bot	Concerning "approved radiology centres", these centres are explicitly named on an approved list at your insurer **Plan A** has has full cover for MRI, CT and PET-CT scans in direct payment centres The exception is Cardiac CT's which are covered by non-direct payment terms Cover up to €500 is provided in non direct payment centres and this is subject to outpatient excess **Plan B** has full cover for CT, PET-CT and MRI There is no excess for category 1 There is €125 excess for category 2 Refer to your insurer's rules for more information. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"cardiac"}
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":4}
bot	Concerning "non-approved radiology centres", these centres are NOT on an approved list at your insurer **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	Which plan is better for most cover for best price?
bot	Concerning "Annual price per adult", note that this is the individual price for a single adult. **Plan A** has €1583 **Plan B** has €2238

Participant #2

user	Which plan is best for cancer support
bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment

user	Is there a limit to the number of nights each year
bot	I'm sorry. I didn't quite understand what you said.
bot	Apologies! I didn't understand that last request.
user	Is there a limit to the claims on Plan B
bot	I'm sorry! I don't have that information. Please contact your insurer.
user	Do I get any other accomodation support benefits with plan B
bot	I'm sorry! I don't have that information. Please contact your insurer.
user	Which plan is better for convalescence
bot	Concerning "Convalescence", **Plan A** has €50 per day for 14 days **Plan B** has €51 per night for 14 nights in semi private and private room accommodation
user	Can I make multiple claims in a year
bot	I'm sorry. I didn't quite understand what you said.
bot	Apologies! I didn't understand that last request.
user	Am i limited to one claim per year
bot	Apologies! I didn't understand that last request.
user	Can I claim more than once
bot	Sorry! I didn't catch which benefit you asked for.
bot	Apologies! I didn't understand that last request.
user	What does convalescence cover?
bot	Concerning "Convalescence", **Plan A** has €50 per day for 14 days **Plan B** has €51 per night for 14 nights in semi private and private room accommodation
user	Which plan is better for overseas care
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":6}
bot	Concerning "overseas hospital bills", **Plan A** has €100,000 cover **Plan B** has €100,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.

user	/select_number_pbi{"number_pbi":7}
bot	Concerning "surgical procedures overseas which are also in Ireland", **Plan A** has full cover in the EU subject to approval **Plan B** has €100,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "Companion Repatriation Expenses", **Plan A** has €1,000 cover **Plan B** has €1,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}
bot	Concerning "Expenses For Companion", **Plan A** has no cover **Plan B** has €1,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":8}
bot	Concerning "surgical procedures overseas which are not found in Ireland", **Plan A** has full cover in the EU subject to approval **Plan B** has €100,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":4}

bot	Concerning "Gender Reassignment Surgery", **Plan A** has cover up to €100,000 subject to prior approval **Plan B** has cover for medically necessary surgeries subject to pre-approval up to the level of cover on the plan. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":5}
bot	Concerning "Repatriation Expenses", **Plan A** has €2,000,000 cover **Plan B** has full cover. Click on the button below to see more benefits from the category
user	vaccinations
bot	Concerning "Vaccinations", **Plan A** has no cover **Plan B** has no cover
user	Is there any vaccination cover available
bot	I'm sorry! I don't have that information. Please contact your insurer.
user	Individual excess
bot	Concerning "Individual Excess", please note that less excess means more money back for you when you claim. **Plan A** has €150 **Plan B** has €100
user	Do specific procedures carry different excess
bot	I didn't understand the benefit you are asking for.
bot	Concerning "Family Excess", please be aware that excess applies to the entire family unless stated differently. Less excess means more money back for you when you claim. **Plan A** has €250 **Plan B** has €100 excess per person
user	Which plan is better for accident and emergency

bot	Concerning "A&E Charge", **Plan A** has €20 cover per visit Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres **Plan B** has €75 for 1 visit (Applies to A&E in Public Hospitals Only) Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit
user	Are there any other A&E benefits
bot	I don't have enough information for that request. Please ask your insurer for details.
user	Psychology
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Psycho-Oncology Counselling", **Plan A** has no cover **Plan B** has €50 cover per visit for 10 visits. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"psychology"}
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "an employee assistance programme", **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	radiology
bot	"radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "Radiology Consultants' Fees", **Plan A** has full cover subject to outpatient excess **Plan B** has €60 cover per procedure. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}

bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}
bot	Concerning "Radiology Test Fees", **Plan A** has cover up to €500, subject to out-patient excess **Plan B** has 50% of charges up to a maximum of €850 per year. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":4}
bot	Concerning "approved radiology centres", these centres are explicitly named on an approved list at your insurer **Plan A** has has full cover for MRI, CT and PET-CT scans in direct payment centres The exception is Cardiac CT's which are covered by non-direct payment terms Cover up to €500 is provided in non direct payment centres and this is subject to outpatient excess **Plan B** has full cover for CT, PET-CT and MRI There is no excess for category 1 There is €125 excess for category 2 Refer to your insurer's rules for more information. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":5}
bot	Concerning "non-approved radiology centres", these centres are NOT on an approved list at your insurer **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	sporting injury cover
bot	I didn't understand the benefit you are asking for.
user	/command_list_knowledge{"knowledge_item":"benefits"}

bot	There are 63 benefit categories as follows; a and e, alternative medicine, breastfeeding, cancer, cancer accommodation, cardiac, child healthcare, chiropody, chiropractor, consultant, convalescence, day case, dental, dietitian, emergency, employee assistance, excess, family excess, fast access, fertility, gender, gender reassignment surgery, gp, health screening, healthcoach, hearing, high tech hospital, home nursing, hospital overseas, hrt for gender dyshporia, individual excess, inpatient, liver disease, maternity, occupational therapy, optical, orthoptistry, osteopath, out-patient radiology, outpatient, overseas, parent with child, physio, podiatry, policy limit, post natal, pre natal, price, prices per annum, private hospital, private room, psychiatry, psychology, public hospital, radiology, reflexology, repatriation, routine dental, semi-private room, special procedures, speech therapy, surgery, vaccination,

Participant #3

user	cancer accommodation support
bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment
user	how many nights in a policy year?
bot	Apologies! I didn't understand that last request.
user	maximum number of nights
bot	I'm sorry. I didn't quite understand what you said.
bot	Apologies! I didn't understand that last request.
user	Plan B cancer accommodation support
bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment
user	convalescence
bot	Concerning "Convalescence", **Plan A** has €50 per day for 14 days **Plan B** has €51 per night for 14 nights in semi private and private room accommodation

user	optical
bot	Concerning "Optical", **Plan A** has no cover **Plan B** has cover for eye tests of €30 in each 24 month period Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	vaccinations
bot	Concerning "Vaccinations", **Plan A** has no cover **Plan B** has no cover
user	individual excess
bot	Concerning "Individual Excess", please note that less excess means more money back for you when you claim. **Plan A** has €150 **Plan B** has €100
user	emergency
bot	Concerning "A&E Charge", **Plan A** has €20 cover per visit Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres **Plan B** has €75 for 1 visit (Applies to A&E in Public Hospitals Only) Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit
user	psychology
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "an employee assistance programme", **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	radiology
bot	"radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}

bot	Concerning "Radiology Test Fees", **Plan A** has cover up to €500, subject to out-patient excess **Plan B** has 50% of charges up to a maximum of €850 per year. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits. Click on the button below to see more benefits from the category
user	everyday health expenses
bot	I'm sorry. I didn't quite understand what you said.
user	/command_list_knowledge{"knowledge_item":"benefits"}

bot	There are 63 benefit categories as follows; a and e, alternative medicine, breastfeeding, cancer, cancer accommodation, cardiac, child healthcare, chiropody, chiropractor, consultant, convalescence, day case, dental, dietitian, emergency, employee assistance, excess, family excess, fast access, fertility, gender, gender reassignment surgery, gp, health screening, healthcoach, hearing, high tech hospital, home nursing, hospital overseas, hrt for gender dyshporia, individual excess, inpatient, liver disease, maternity, occupational therapy, optical, orthoptistry, osteopath, out-patient radiology, outpatient, overseas, parent with child, physio, podiatry, policy limit, post natal, pre natal, price, prices per annum, private hospital, private room, psychiatry, psychology, public hospital, radiology, reflexology, repatriation, routine dental, semi-private room, special procedures, speech therapy, surgery, vaccination,
user	gp
bot	Concerning "GP Visits", **Plan A** has 24 hr GP Line 24 Hr Nurseline Three free GP live online consultations **Plan B** has cover for €30 per visit Practise Nurse is covered for 1 visit at €20 Full cover is provided for 12 video calls to your insurer's online doctor
user	health screening

bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits

Participant #4

user	what is cancer accommodation support?
bot	I'm sorry but I don't have a direct answer. You can ask me to list the terms and benefits that I know or click one of the buttons below.
user	/command_list_knowledge{"knowledge_item":"benefits"}

bot	There are 63 benefit categories as follows; a and e, alternative medicine, breastfeeding, cancer, cancer accommodation, cardiac, child healthcare, chiropody, chiropractor, consultant, convalescence, day case, dental, dietitian, emergency, employee assistance, excess, family excess, fast access, fertility, gender, gender reassignment surgery, gp, health screening, healthcoach, hearing, high tech hospital, home nursing, hospital overseas, hrt for gender dyshporia, individual excess, inpatient, liver disease, maternity, occupational therapy, optical, orthoptistry, osteopath, out-patient radiology, outpatient, overseas, parent with child, physio, podiatry, policy limit, post natal, pre natal, price, prices per annum, private hospital, private room, psychiatry, psychology, public hospital, radiology, reflexology, repatriation, routine dental, semi-private room, special procedures, speech therapy, surgery, vaccination,
user	plan A cancer accommodation
bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment
user	optical cover
bot	Concerning "Optical", **Plan A** has no cover **Plan B** has cover for eye tests of €30 in each 24 month period Glasses and contact lenses are covered for 75% of charges up to €100 in each 24 month period
user	Compare hearing benefit
bot	I didn't understand the benefit you are asking for.
bot	Concerning "a hearing test", **Plan A** has no cover **Plan B** has no cover
user	hearing
bot	Concerning "a hearing test", **Plan A** has no cover **Plan B** has no cover
user	list of term

bot	There are 28 terms which I can explain as follows; agreed charges, alternative medicine, cardiac procedures, convalescence, day case, ECG, excess, family excess, gender dysphoria, health coach, high tech, individual excess, inpatient, MRI, out patient excess, outpatient, outpatient policy limit, partner assistance, PET-CT, pre-existing condition, procedures other than cardiac and special, repatriation, schedule of benefits, semi private, shortfall, special procedures, surgery not available, technologically advanced hospitals,
user	/command_list_knowledge{"knowledge_item":"benefits"}
bot	There are 63 benefit categories as follows; a and e, alternative medicine, breastfeeding, cancer, cancer accommodation, cardiac, child healthcare, chiropody, chiropractor, consultant, convalescence, day case, dental, dietitian, emergency, employee assistance, excess, family excess, fast access, fertility, gender, gender reassignment surgery, gp, health screening, healthcoach, hearing, high tech hospital, home nursing, hospital overseas, hrt for gender dyshporia, individual excess, inpatient, liver disease, maternity, occupational therapy, optical, orthoptistry, osteopath, out-patient radiology, outpatient, overseas, parent with child, physio, podiatry, policy limit, post natal, pre natal, price, prices per annum, private hospital, private room, psychiatry, psychology, public hospital, radiology, reflexology, repatriation, routine dental, semi-private room, special procedures, speech therapy, surgery, vaccination,
user	individual excess
bot	Concerning "Individual Excess", please note that less excess means more money back for you when you claim. **Plan A** has €150 **Plan B** has €100
user	accident and emergency
bot	Concerning "A&E Charge", **Plan A** has €20 cover per visit Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres **Plan B** has €75 for 1 visit (Applies to A&E in Public Hospitals Only) Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit
user	which plan is better for A&E

bot	Concerning "A&E Charge", **Plan A** has €20 cover per visit Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres **Plan B** has €75 for 1 visit (Applies to A&E in Public Hospitals Only) Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit
user	psycology
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Psycho-Oncology Counselling", **Plan A** has no cover **Plan B** has €50 cover per visit for 10 visits. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"psychology"}
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "an employee assistance programme", **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	pyschiatry
bot	"psychiatry" is a category containing 3 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Outpatient Mental Health Care", **Plan A** has €50 per visit for 12 Mental Health Consultations in the Dean Clinic subject to outpatient excess Also €50 per visit for 12 Eating Disorder Consultations in Lois Bridges subject to outpatient excess Also full cover for 24/7 Telephone Counselling Support Face to face/Video/Phone Counselling is covered for 5 visits per issue In addition there is full cover for 1 session per issue for the following support services legal, consumer & health advice, career guidance, life coach & mediation **Plan B** has cover for Mental Health Therapy of €75 per visit for 12 visits; A Clinical Psychologist is covered for €20 per visit for 12 visits under the Alternative Practitioners benefit. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"psychiatry"}
bot	"psychiatry" is a category containing 3 individual benefits. Click on any button for more information.
user	cardiac cover
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "The Blackrock Clinic, the Mater Private and the Beacon Hospital, certain Cardiac Procedures", **Plan A** has full cover in the Beacon Hospital with once-off €300 excess per family per year Reduced cardiac list coverage in the Mater Private and Blackrock Clinic with no excess Ask your insurer for details **Plan B** has 90% cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"cardiac"}
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"cardiac"}
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}

bot	Concerning "approved radiology centres", these centres are explicitly named on an approved list at your insurer **Plan A** has has full cover for MRI, CT and PET-CT scans in direct payment centres The exception is Cardiac CT's which are covered by non-direct payment terms Cover up to €500 is provided in non direct payment centres and this is subject to outpatient excess **Plan B** has full cover for CT, PET-CT and MRI There is no excess for category 1 There is €125 excess for category 2 Refer to your insurer's rules for more information. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"cardiac"}
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":4}
bot	Concerning "non-approved radiology centres", these centres are NOT on an approved list at your insurer **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	consultants
bot	"consultant" is a category containing 3 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Consultant Fees", **Plan A** has €51 per visit **Plan B** has cover for €60 per visit A Second Opinion Service is covered Also 50% cover for a Consultant at your insurer's clinic including orthopaedic, oral maxillofacial and sports medicine . Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"consultant"}
bot	"consultant" is a category containing 3 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}

bot	Concerning "Delivery Consultant Fees", **Plan A** has full cover up to agreed charges incurred on day of delivery for participating consultants **Plan B** has full cover up to agreed charges incurred on day of delivery for participating consultants. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"consultant"}
bot	"consultant" is a category containing 3 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}
bot	Concerning "Radiology Consultants' Fees", **Plan A** has full cover subject to outpatient excess **Plan B** has €60 cover per procedure. Click on the button below to see more benefits from the category

Participant #5

user	Which plan is best for cancer support
bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment
user	Is there a limit to the number of nights each year
bot	I'm sorry. I didn't quite understand what you said.
bot	Apologies! I didn't understand that last request.
user	Is there a limit to the claims on Plan B
bot	I'm sorry! I don't have that information. Please contact your insurer.
user	Do I get any other accomodation support benefits with plan B
bot	I'm sorry! I don't have that information. Please contact your insurer.
user	Which plan is better for convalescence
bot	Concerning "Convalescence", **Plan A** has €50 per day for 14 days **Plan B** has €51 per night for 14 nights in semi private and private room accommodation
user	Can I make multiple claims in a year
bot	I'm sorry. I didn't quite understand what you said.

bot	Apologies! I didn't understand that last request.
user	Am i limited to one claim per year
bot	Apologies! I didn't understand that last request.
user	Can I claim more than once
bot	Sorry! I didn't catch which benefit you asked for.
bot	Apologies! I didn't understand that last request.
user	What does convalescence cover?
bot	Concerning "Convalescence", **Plan A** has €50 per day for 14 days **Plan B** has €51 per night for 14 nights in semi private and private room accommodation
user	Which plan is better for overseas care
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":6}
bot	Concerning "overseas hospital bills", **Plan A** has €100,000 cover **Plan B** has €100,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":7}
bot	Concerning "surgical procedures overseas which are also in Ireland", **Plan A** has full cover in the EU subject to approval **Plan B** has €100,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "Companion Repatriation Expenses", **Plan A** has €1,000 cover **Plan B** has €1,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}

bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}
bot	Concerning "Expenses For Companion", **Plan A** has no cover **Plan B** has €1,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":8}
bot	Concerning "surgical procedures overseas which are not found in Ireland", **Plan A** has full cover in the EU subject to approval **Plan B** has €100,000 cover. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":4}
bot	Concerning "Gender Reassignment Surgery", **Plan A** has cover up to €100,000 subject to prior approval **Plan B** has cover for medically necessary surgeries subject to pre-approval up to the level of cover on the plan. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"overseas"}
bot	"overseas" is a category containing 8 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":5}
bot	Concerning "Repatriation Expenses", **Plan A** has €2,000,000 cover **Plan B** has full cover. Click on the button below to see more benefits from the category
user	vaccinations

bot	Concerning "Vaccinations", **Plan A** has no cover **Plan B** has no cover
user	Is there any vaccination cover available
bot	I'm sorry! I don't have that information. Please contact your insurer.
user	Individual excess
bot	Concerning "Individual Excess", please note that less excess means more money back for you when you claim. **Plan A** has €150 **Plan B** has €100
user	Do specific procedures carry different excess
bot	I didn't understand the benefit you are asking for.
bot	Concerning "Family Excess", please be aware that excess applies to the entire family unless stated differently. Less excess means more money back for you when you claim. **Plan A** has €250 **Plan B** has €100 excess per person
user	Which plan is better for accident and emergency
bot	Concerning "A&E Charge", **Plan A** has €20 cover per visit Consultations & treatment Up to €150 in your insurer's approved minor injuries & illness centres Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres **Plan B** has €75 for 1 visit (Applies to A&E in Public Hospitals Only) Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit
user	Are there any other A&E benefits
bot	I don't have enough information for that request. Please ask your insurer for details.
user	Psychology
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Psycho-Oncology Counselling", **Plan A** has no cover **Plan B** has €50 cover per visit for 10 visits. Click on the button below to see more benefits from the category

user	/ask_pbi_or_benefit_tag{"benefit_tag":"psychology"}
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "an employee assistance programme", **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	radiology
bot	"radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits. Click on the button below to see more benefits from the category

user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":2}
bot	Concerning "Radiology Consultants' Fees", **Plan A** has full cover subject to outpatient excess **Plan B** has €60 cover per procedure. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":3}
bot	Concerning "Radiology Test Fees", **Plan A** has cover up to €500, subject to out-patient excess **Plan B** has 50% of charges up to a maximum of €850 per year. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}
bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":4}
bot	Concerning "approved radiology centres", these centres are explicitly named on an approved list at your insurer **Plan A** has has full cover for MRI, CT and PET-CT scans in direct payment centres The exception is Cardiac CT's which are covered by non-direct payment terms Cover up to €500 is provided in non direct payment centres and this is subject to outpatient excess **Plan B** has full cover for CT, PET-CT and MRI There is no excess for category 1 There is €125 excess for category 2 Refer to your insurer's rules for more information. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"out-patient radiology"}

bot	"outpatient radiology" is a category containing 5 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":5}
bot	Concerning "non-approved radiology centres", these centres are NOT on an approved list at your insurer **Plan A** has no cover **Plan B** has no cover. Click on the button below to see more benefits from the category
user	sporting injury cover
bot	I didn't understand the benefit you are asking for.
user	/command_list_knowledge{"knowledge_item":"benefits"}
bot	There are 63 benefit categories as follows; a and e, alternative medicine, breastfeeding, cancer, cancer accommodation, cardiac, child healthcare, chiropody, chiropractor, consultant, convalescence, day case, dental, dietitian, emergency, employee assistance, excess, family excess, fast access, fertility, gender, gender reassignment surgery, gp, health screening, healthcoach, hearing, high tech hospital, home nursing, hospital overseas, hrt for gender dyshporia, individual excess, inpatient, liver disease, maternity, occupational therapy, optical, orthoptistry, osteopath, out-patient radiology, outpatient, overseas, parent with child, physio, podiatry, policy limit, post natal, pre natal, price, prices per annum, private hospital, private room, psychiatry, psychology, public hospital, radiology, reflexology, repatriation, routine dental, semi-private room, special procedures, speech therapy, surgery, vaccination,

Participant #6

user	Is there a limit on the amount of time I can stay at a hospital overseas? Is there an excess on the initial amount that i have to pay before I am covered?
bot	I didn't understand the benefit you are asking for.
bot	Apologies! I didn't understand that last request.
user	Which plan is better for overseas hospital??
bot	Concerning "overseas hospital bills", **Plan A** has €100,000 cover **Plan B** has €100,000 cover
user	Is there an excess that I have to pay??

bot	I didn't understand the benefit you are asking for.
bot	Apologies! I didn't understand that last request.
user	Does either plan have an excess?
bot	"excess" is a category containing 2 individual benefits. Click on any button for more information.
user	What is covered on overseas hospital on plan A?
bot	Concerning "overseas hospital bills", **Plan A** has €100,000 cover **Plan B** has €100,000 cover
user	which plan is better for hearing??
bot	Concerning "a hearing test", **Plan A** has no cover **Plan B** has no cover
user	does any plan cover hearing?
bot	Concerning "a hearing test", **Plan A** has no cover **Plan B** has no cover
user	which plan is better for family excess?
bot	Concerning "Family Excess", please be aware that excess applies to the entire family unless stated differently. Less excess means more money back for you when you claim. **Plan A** has €250 **Plan B** has €100 excess per person
user	which plan is better for a semi-private room in private hospital??
bot	Concerning "a semi-private room in a private hospital", **Plan A** has coverage with once-off €300 excess per family per year No excess applies to 5 orthopaedic procedures (hip and knee replacement) in 12 private hospitals Check with your insurer for more details **Plan B** has full cover but 90% cover for certain cardiac and special procedures in the Hermitage and the Galway Clinic
user	which plan s better for psychology??
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Psycho-Oncology Counselling", **Plan A** has no cover **Plan B** has €50 cover per visit for 10 visits. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"psychology"}
bot	"psychology" is a category containing 2 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}
bot	Concerning "Psycho-Oncology Counselling", **Plan A** has no cover **Plan B** has €50 cover per visit for 10 visits. Click on the button below to see more benefits from the category
user	which plan is better for cardiac?
bot	"cardiac" is a category containing 4 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Health Screening", **Plan A** has Full cover for Mammograms and Dexa scans in certain centres Women's Cancer Screening is covered up to €30 Men's Cancer Screening is covered up to €20 Full cover for 'HeartBeat' cardiac screening with no excess Home testing kits are covered up to €20 Health Screening up to €150 every 2 years is paid directly to centre **Plan B** has cover for Lifestages Screening Programme of €335 per screen and heart check and cancer check every 2 years in your insurer's Medical Centre There is full cover for Mammograms and Dexa Scans every 2 years There is no excess applied for the above The Joint Care Programme provides Joint Care Screening to assess your mobility levels carried out by a Physiotherapist employed by Spectrum Health This is full covered every 2 years There is full cover for 3 visits with no excess for Joint Care Physiotherapy carried out by a Physiotherapist employed by Spectrum Health subject to referral from a Joint Care Screening The Medfit Cardiac Care programme is covered for €250 every 2 years There is Urgent Cardiac Care benefit of €250 per year The Medfit Cardiac Rehabilitation programme is covered for€250 per year Full cover is provided for genetic testing for specified genetic mutations in an approved clinic and €125 cover for initial consultation Follow on preventative treatment is covered up to the level of cover provided in the Inpatient section of benefits. Click on the button below to see more benefits from the category
user	which plan is better for hospital stays and operations covered?
bot	I didn't understand the benefit you are asking for.
user	which plan is better for operations covered?
bot	Concerning "Convalescence", **Plan A** has €50 per day for 14 days **Plan B** has €51 per night for 14 nights in semi private and private room accommodation
user	which plan is better for hospital stays
bot	"hospital stay" is a category containing 12 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":1}

bot	Concerning "Cancer Accommodation Support", **Plan A** has no cover **Plan B** has €100 for 1 night's accommodation for each treatment. Click on the button below to see more benefits from the category
user	/ask_pbi_or_benefit_tag{"benefit_tag":"inpatient"}
bot	"inpatient" is a category containing 12 individual benefits. Click on any button for more information.
user	/select_number_pbi{"number_pbi":8}
bot	Concerning "a semi-private room in a public hospital", **Plan A** has full cover **Plan B** has full cover. Click on the button below to see more benefits from the category

Appendix: Reflections on the Implementation

This research project consisted of a case study in implementing a chatbot for a complex domain using open source software. The process and experience of implementation is considered a result. There were many disparate problems solved related to coding, language understanding and dialogue. This chapter reflects on these problems and their solutions.

Python Libraries

Installation of Rasa into a python virtual environment proved very useful in managing dependencies [55]. When Rasa released version 3.0 of their software it was possible to make a fairly smooth migration from Rasa 2 to Rasa 3 since the new version could be installed in a separate environment on the same machine. The Rasa action server is well architected to easily support custom code and custom modules. This includes code for execution of policy actions and for validation of extracted entities before filling slots. Python libraries such as Fuzzywuzzy for Levenshtein matching proved effective and were also easy to install and use.

Entity Extraction

Rasa lacked capabilities which might reasonably have been expected, for example enumerated types. One of the entity types in the domain, *benefit_tags*, was a finite list of about sixty elements including “cancer”, “gp” and “dental”. An element of this list needed to be extracted from utterances such as “*What cover do you provide for **cancer**?*” The final stage of entity extraction had to be custom coded. At first the Rasa lookup tables functionality was utilised [56]. It was discovered that this only provided hints to the NLU component and did not provide the desired extraction of enumerated types.

Custom python code was written to calculate Levenshtein distance on the keys of the *benefit_tag* and *pbi* dictionaries in the ontology. This was done to identify the most probable element of the enumerated type. The python fuzzywuzzy library was used to good effect here.

During the first survey execution it was discovered that participants often add extraneous words into their utterances which are extracted by Rasa NLU as part of an entity, for example a plan benefit item (pbi). This leads to incorrect matching with keys in the dictionaries of *benefit_tags* and *pbis*. For example, a participant was asked to choose the best plan for optical. The participant asked the bot, “Which plan is better for **optical policies**?” Rasa NLU then extracted “optical policies” as an entity of type *benefit_tag*. The best match using Levenshtein distance from python’s fuzzywuzzy library was “public hospital” with a confidence of 0.65. This is a completely different benefit to what the user intended to ask about.

A method was devised to prevent incorrect entity extraction by creating a *junk* entity type. This type meant the NLU extracted certain words before they could be extracted as part of a

benefit_tag or *pbi* entity. Junk words included the following; "benefits", "policies", "care", "options", "cover", "best", "problems". As testing continued, more *junk* words were discovered.

The NLU was trained with these *junk* entities as below;

```
which plan has a better [range](junk) of [hospitals](benefit_tag)
which plan is better for [cancer](benefit_tag) [benefit](junk)
which plan has a better [cardiac](benefit_tag) [care](junk)
which plan has the [best](junk) [optical](benefit_tag) [policy](junk)
how is [cardiac](benefit_tag) [covered](junk)
```

On occasion, Rasa still extracted junk words as part of a *benefit_tag* when there was an unusual expression of intent in a user utterance. An extra layer of protection was added with a python method, *remove_junk_words*, This was invoked in the custom Rasa validator methods called before the slots were filled.

Forms

When a participant intended to view a benefit category they were presented with a list of items (*pbis*) to choose from. This was implemented as a Rasa form. A single form was created, *number_pbi_form*, and its definition in domain.yml is included below with an explanation

`number_pbi_form:`
`  required_slots:`
`  - number_pbi_slot`
`  - text_pbi_slot`
	A participant could select a particular *pbi* in two ways;
	1) by stating its name
	2) by choosing a number for its position in the list of pbis in a category
	
	The validators for the corresponding slots would set both slots when either one was set. For example, the first pbi in the category “cancer” could be set by typing the number “1” or by entering the text “Aids and accessories for cancer”

Rasa has good functionality out of the box to enable a user to pass transparently through a form without entering any input. This is the case when the slots have already been filled by previous dialogue. This scenario is expressed in the following story in the *stories.yml*

`- story: selecting a pbi by its name`
`  steps:`
`  - intent: state_pbi_selected`
`  - action: number_pbi_form`
`  - action: action_state_plan_benefit_values`
	At any point in the dialogue a participant could simply state “What cover do you have for aids and accessories for cancer?”
	
	Rasa NLU would recognise the intent as *state_pbi_selected* with the appropriate *pbi* entity. The validator for the *pbi* slot would set both slots in the form. Rasa dialog

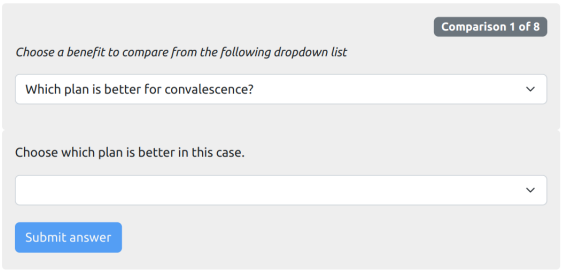
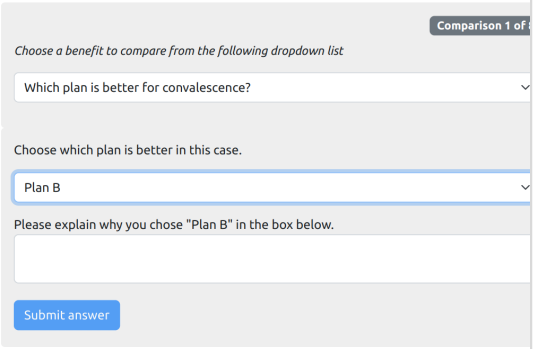
	policy would then invoke the next action to state the value of the *pbi* for both plan A and plan B

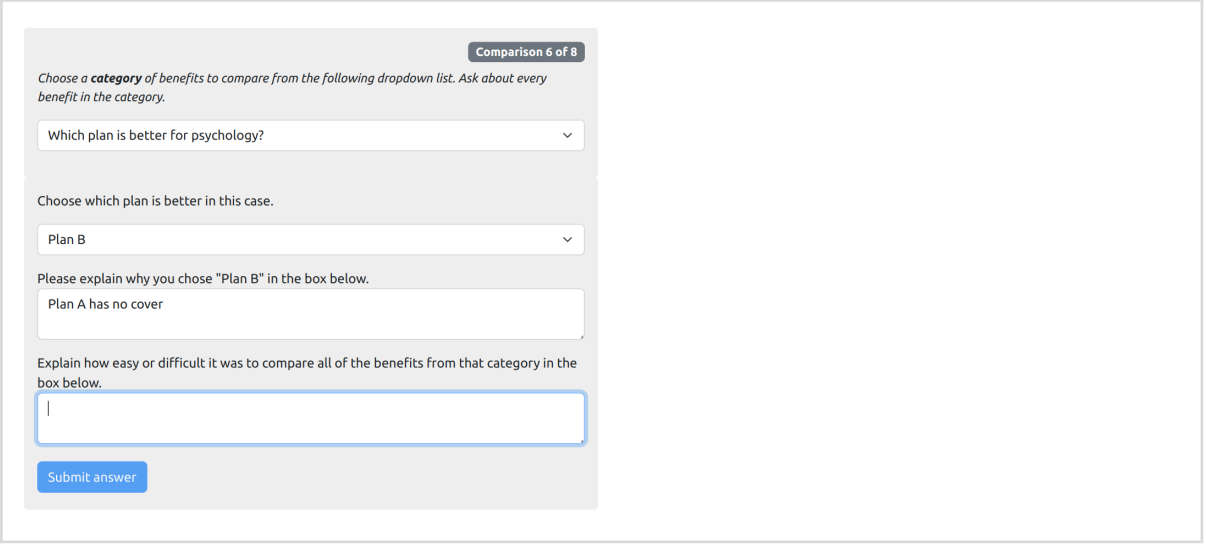
Rasa forms proved to be problematic at times, in part because of the different ways in which users could enter and leave the form. Scenarios arose where the Rasa policies predicted incorrectly that the next action should be to enter the form. In those cases the validator for the *pbi* slot was called but the slot value was *None*. A workaround for this problem was reached by hard coding an exit from the form by returning a python value of *None* for the *requested_slot* in the validator method. The solution seemed obtuse but it worked.

Custom Quiz Application

Since the domain under question was complicated, it was desired to separate out the participant's difficulty with the domain from their difficulty with the chatbot in so far as that separation was possible. A custom quiz application was created for this purpose.

The participant was empowered with a choice between two comparison questions at each stage. Once the question was selected, the chatbot was shown in an iframe where it could be interrogated. The participant was able to focus on that single aspect of the domain.
A three or four step process was followed for each question;

<table>
<tr>
<td>1. A single domain question is selected
 
</td>
<td>2. The participant enters their answer
3. The participant is prompted to explain the answer. This checks their understanding of both the original question and the response.
 
</td>
</tr>
<tr>
<td colspan="2">4. In the case of a category question, a further question is posed to determine how the participant coped with category navigation.</td>
</tr>
</table>



The explanations given by participants were often illuminating. Simple statements included "*Plan A has no cover*" as a reason for choosing "*Plan B*". This showed understanding of the question and the response. Another statement "*I do not wear contacts or glasses so it is not needed monthly for me*" indicated that this particular participant did not understand that they were asked to evaluate the plans on an objective basis and not on a personal basis.

The experiment results showed that participants managed well with categories of two items but struggled to compare categories of four items.

Comparison Questions

Questions posed to participants were carefully chosen to expose the complicated nature of the domain. The question difficulty increased slowly to prevent overwhelming the participant.

Question type	Example	Participant reaction
Easy	Which plan is better for cancer accommodation? Plan A: no cover Plan B: €100 for 1 night's accommodation for each treatment	The difference was obvious since one plan had no cover at all. This was understood by all participants.
Medium difficulty	Which plan is better for individual excess? Plan A: €150 Plan B: €100	The answer was determinate but some participants were confused by the concept of excess and thought that higher excess was better.
Providers used different parameters	Which plan is better for accident and emergency? Plan A: €20 cover per visit; Consultations & treatment	Some participants were confused because the items listed by each provider were not directly comparable. The providers used different parameters to state the

Question type	Example	Participant reaction
	Up to €150 in your insurer's approved minor injuries & illness centres; Fracture Clinic Up to €40 per visit in your insurer's approved minor injuries & illness centres Plan B: €75 for 1 visit (Applies to A&E in Public Hospitals Only). Also full cover at your insurer's clinic for unlimited visits with a €50 excess per visit.	benefit value. One participant stated *"I don't like the idea of (up to) in plan A as it is vague"*
A category with 2 items	Which plan is better for psychology? Note that the participant had to navigate 2 benefits in this category	Most participants successfully navigated the category of 2 items
A category with 4 items	Which plan is better for cancer? Note that the participant had to navigate 4 benefits in this category	In the first survey execution most participants failed to navigate the category. In the second survey execution the use of buttons helped to improve the navigation

The utterances seen from the users exposed new intent expressions as they interrogated the chatbot.

Intent Expressions

After each test iteration, further ways of expressing intents were discovered. In the following examples, “X” represents a benefit such as ‘cancer’

Initial intent expressions	Additional intent expressions
Which plan is better for X Tell me about X What cover do you have for X Is X covered	How is X covered How much is X for plan A and plan B Which plan has a better range of X What is the X in plan A and plan B Is X included X

Note that the participants often simply stated the name of the benefit, “X” with no surrounding words at all and expected the chatbot to understand.

These additional expressions were used to generate better NLU training examples and contributed to decreasing the percentage of requests which were misunderstood by the chatbot from 28% to 12%.

Debugging with Rasa

Troubleshooting and debugging with Rasa was often time-consuming with some good results as described in the following sections.
Logs:

Turning on full logging in the Rasa logs leads to information overload. Custom debug statements in the actions and validators often proved to be more useful.
Events table:
The events table records all events which change the dialog state. It is very useful for debugging the next action predictions of policies. However, this was not mentioned in the Rasa training videos [56]. It was discovered later in the project development cycle when an sqlite database was configured as the tracker store. At that point the value of the events table was discovered. A sample action event is shown below which illustrates a scenario where the TED policy was the winner. It includes the confidence in the prediction of the policy.

```
{
    "action_text": null,
    "confidence": 0.9874375462532043,
    "event": "action",
    "hide_rule_turn": false,
    "metadata": {
        "model_id": "2e4df560aa99404cbc49d10bbfdda48b"
    },
    "name": "number_pbi_form",
    "policy": "TEDPolicy",
    "timestamp": 1678121735.0554073
}
```

It would be useful for fine tuning the dialog manager if the predictions of the losing policies were also displayed by Rasa.
Forum:
The Rasa community forum is a place to find answers to common Rasa problems [57]. However, many of the answers relate to previous versions of Rasa and are not applicable and can be counter productive.

Multimodal Dialogue

This project delivered two modes, chat and voice. Beta testing was done for both modes. It was amazing to see how much more excited the beta users were with the voice implementation. The twilio voice communications platform was integrated with Rasa to support the voice channel. The voices provided by twilio were very human-like and engaging. Voice threw up some new challenges for the dialogue system including the following

- dropped words
- utterances misunderstood by the speech-to-text service
- speech timeouts

Call recordings helped to highlight and also troubleshoot the challenges raised by voice. Rasa is well architected in so far as most of the code can be shared between voice channel and chat channel implementations. At any stage of an action, custom behaviour can be applied for a particular channel e.g. buttons are displayed for the chat channel but not for voice. The chatbot experiments were performed using chat only.

Twilio voice recording:
It was desirable for beta testing to record voice calls for analysis. There is no setting in the Twilio management dashboard to automatically record calls. Instead the Rasa twilio connector had to be modified to read the CallSID identifier from the incoming Twilio request for each new voice call. Then the python asyncio event loop was used to send an asynchronous http post request to twilio with the account credentials and the CallSid. This action initiated recording of the call.
Twilio speech timeout:
There is a configurable timeout in the connector that tells Twilio how long to wait for the user utterance. When this is set to a number of seconds there can be a very noticeable lag in the conversation responses from Rasa. The most natural setting is to set the timeout to "auto". However this has a knock on effect of causing Rasa to often repeat the prompt which can be very annoying for users. It can make the voicebot sound idiotic. A reprompt fallback phrase was used to overcome this problem.
Twilio voice connector characters:
Twilio sends a full stop (period) at the end of every user utterance. This caused Rasa to sometimes misinterpret the intent or entity. The connector was modified to always remove this trailing full stop using python code.
Affordances:
When implementing chat, it can be a pitfall to only implement chat and text and ignore other affordances. Following the first chatbot experiment it was realised that category navigation was a failure for most users. The original chatbot widget supported only text so a more comprehensive widget was sourced which also supported buttons and images. The use of button affordances in category navigation helped to increase the participants' ability to navigate categories in the second chatbot experiment.

Appendix Voice Channel Implementation

Voice capability was incorporated using the Twilio communications platform. Twilio was connected to the chatbot framework via the Rasa twilio channel connector. The source code of the connector was modified to support recording of calls. This was tested with beta users.

Twilio Phone Number

An Irish twilio virtual phone number was maintained for beta testing to avoid call charges for participants. The phone number used Twilio infrastructure including speech-to-text and text-to-speech services.

Rasa Twilio Connector

The Twilio connector was configured for a very human like voice provided by Amazon, "`Polly.Matthew-Neural`"

```
twilio_voice:
  initial_prompt: "hello"
  assistant_voice: "Polly.Matthew-Neural"
  language: "en-GB"
  reprompt_fallback_phrase: "I didn't get that could you repeat?"
  speech_timeout: "1"
  speech_model: "enhanced"
  enhanced: "false"
```

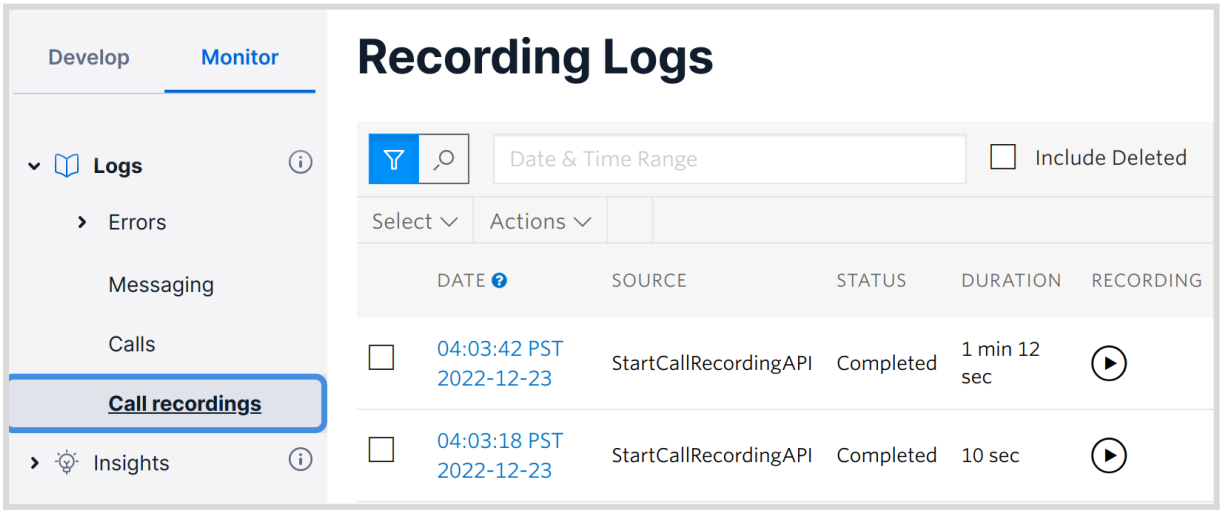
The Rasa twilio connector was modified as follows;

- Add recording capability so that participant voice calls to the chatbot could be recorded for later analysis
- Remove the trailing full stop sent by twilio. This interfered with entity recognition by Rasa NLU

Every time a message was received in the Rasa server indicating that a twilio call had been made by a user, then an async call was made to the Twilio api server to initiate call recording as shown in the code sample below.

```
    async def _twilio_record(self,call_sid):
        logger.warning('_twilio_record call_sid='+str(call_sid))
        if not self.record:
            return
                                                                        url=
f"https://api.twilio.com/2010-04-01/Accounts/{self.account_sid}/Calls/{call
_sid}/Recordings.json"
        obj={}
        await asyncio.sleep(1)
                                    response=requests.post(url,    json    =
obj,auth=HTTPBasicAuth(self.account_sid,self.auth_token))
        logger.warning('_twilio_record response='+str(response))
```

Call recordings were then available in the Twilio management dashboard as shown in the following image.



Ngrok

Ngrok , ingress-as-a-service, was a very useful solution that enabled rapid development of the connector to Twilio from the developer's local machine. Once the developed code was robust it was then deployed to Google Compute Platform (GCP).