Co-designing a Chatbot for and with Refugees and Migrants

ZHIFA CHEN

Co-designing a Chatbot for and with Refugees and Migrants

Zhifa Chen

Master's Thesis
Collaborative and Industrial Design (CoID)
Department of Design

School of Arts, Design & Architecture
Aalto University
2019

Supervisor: Andrés Lucero
Advisor: Mika P. Nieminen & Yichen Lu

Co-designing a Chatbot for and with Refugees and Migrants

Abstract

Aalto University
School of Arts, Design
and Architecture

Aalto University, P.O. BOX 31000, 00076 AALTO
www.aalto.fi

Master of Arts thesis abstract

Author Zhifa Chen		
Title of thesis Co-designing a Chatbot for and with Refugees and Migrants		
Department Department of Design		
Degree programme Collaborative and Industrial Design		
Year 2019	**Number of pages** 166	**Language** English

An information portal, HandbookGermany.de, is developed to support the integration of refugees and migrants into society in Germany. However, the information-seeking process is exhausting, cumbersome, and even confusing if refugees and migrants are not proficient at using web services. In light of this, a chatbot-based conversational service is considered as an alternative to enhance the information-seeking experience. For the purpose of designing products and services for refugees and migrants, a great deal of research proposes employing co-design methods as an effective means.

The overall aim of this thesis is to explore, understand, and define possibilities of improving refugees and migrants' experiences of social integration by proposing an engaging and efficient chatbot solution. Furthermore, this thesis aims to explore the necessity of co-design approach as a critical methodology to develop solutions. Therefore, the main research question in this thesis is how can a co-design approach contribute to designing a chatbot supporting social integration within the context of refugees and migrants.

User experience, problems, and needs are unveiled in depth by listening to migrants and refugees' problems, behaviors, and expectations (i.e., document studies, questionnaires, cultural probes, and expert interviews), and observing how migrants interact with the chatbot (i.e., participant observations and empathy probes). The research findings are then transformed into design questions. The designer, developers, and migrants jointly generate concepts leveraging generative toolkits in co-design workshops. By using surveys, the

Method for the Assessment of eXperience (MAX), and *property checklists*, the resulting concepts are later validated with refugees and migrants.

As *research through design*, this thesis draws three conclusions. Firstly, the co-design approach benefits defining problems in the complex context of refugees and migrants by supporting them in expressing ideas and thoughts. The defined problems can then be converted into design questions that promote the proceeding of the design process. Secondly, the co-design approach helps to develop mature concepts, which lays a foundation for the final design. Thirdly, the utilization of co-design tools plays an essential role in validating and refining the solution efficiently, as they make ideas concrete and visible so that refugees and migrants can easily reflect on them throughout the whole design process.

Keywords co-design, chatbot, conversation design, refugees and migrants

Acknowledgments

First and foremost, my sincere thanks go to Mika P. Nieminen for being one of my advisors and allowing me to join the ERICS project. I'm thankful for the valuable feedback and suggestions I received for this thesis. Furthermore, I would like to thank my co-workers, Panajis Rantala, and Jesse Palo, in the ERICS project, contributing to the facilitation of the user research, co-design sessions, and the user experience evaluation. I also thank Elina Nurmi at International House Helsinki, Mia Hartikainen at Arffman Consulting, and Essi Virtanen at Luona (Refugee Center Espoo) for connecting me with the people to conduct my research and enabling me to meet and interview the migrants and refugees in Finland.

Moreover, I would like to express my gratitude to Andrés Lucero for supervising me and providing constructive advice for this thesis and Yichen Lu, as my advisor, for her patience and always helping me out with my thesis challenges.

Finally, I wish to thank my family and friends in China, for supporting my study and offering me with emotional and moral comfort.

Helsinki, April 2019
Zhifa Chen

Table of Contents

List of Figures

List of Tables

1.
Introduction

1.1 Background

The number of refugees requesting asylum in European countries has increased dramatically since 2014 (Holmes & Castaneda, 2016). Most of the refugees are unfamiliar with their host countries, especially when they first arrive at their destination. As refugees fall short of necessary information about everyday life, they face a variety of barriers, such as language learning, employment, school application, and childcare. These problems affect not only refugees but also many migrants.

As one of the countries receiving the biggest number of migrants and refugees, Germany has taken actions to support the integration of newcomers into society. An information portal, HandbookGermany.de (HBG), was developed to meet such demands. The HBG website offers information on diverse aspects of living (e.g., seeking asylum, renting houses, healthcare, searching kindergartens, applying for universities, and hunting for jobs).

Nevertheless, the information-seeking process is exhausting, cumbersome and even confusing for refugees and migrants if they are not skilled in using web services. In the case at hand, a chatbot-based conversational service is considered as an alternative to enhance the information-seeking experience. Chatbots are easy-to-use, natural and intuitive, and save time and energy (Di Gaetano & Diliberto, 2018). Also, information-seeking tasks can be performed with high efficiency as chatbots eliminate further information that may distract users. By immersing themselves in the conversation, users directly pay attention to every single message instead of having to scan a large number of texts online (Di Gaetano & Diliberto, 2018).

Such a demand for producing chatbot services led to the initiation of the ERICS (European Refugee Information and Communication Service) project, which is partially funded by the European Institute of Innovation & Technology (EIT). The ERICS project aims to develop a chatbot supporting arrivers by providing information from the HBG website to solve their everyday-life issues.

Four diverse actors are contributing to the ERICS project: T-Systems Multimedia Solutions (T-Systems MMS), Deutsches Forschungzentrum für

Künstliche Intelligenz (DFKI), Technische Universität Berlin (TU Berlin), and Aalto University. T-Systems MMS is a company responsible for Handbook-Germany.de portal, which is about to use the chatbot service to be created by the ERICS project. Apart from offering the HBG portal as an initial pilot platform, the commercial launch of the ERICS project will be implemented by T-Systems MMS as well. DFKI is a German research center focusing on artificial intelligence[1] and natural language processing[2]. By using this expertise, they will build the server and the dialog management system[3] and be in charge of the integration and the back-end development of the chatbot. As for TU Berlin, they are leading the ERICS project, and their principal responsibility is to develop crowdsourcing-based solutions for translating and improving the HBG website. Aalto University is responsible for the chatbot interface and user experience design as well as front-end development.

[1] In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals.

[2] A subfield of computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

[3] A component of a dialog system, responsible for the state and flow of the conversation.

1.2 Motivation

This thesis developed out of my great empathy for and interest in refugees and migrants as well as my desire to practice co-design and user experience design within the context of product development in real life. Besides, the gap between co-design approach and chatbot development in the context of refugees and migrants both in industry and academia motivated me to explore the topic in this thesis.

As a ‘migrant’ in Finland, I face a variety of unknowns concerning integration into society, which enables me to empathize with refugees and migrants deeply. Additionally, I am enthusiastic about creating a positive social impact through design, and its social relevance characterizes the ERICS project. These factors motivated me to begin this thesis by joining the ERICS project as a user experience designer.

Another motivation for this study was to gain the opportunity to be a co-design practitioner and learn to collaborate with software developers. The ERICS project allowed me to put the co-design method and tools studied at university into practice by approaching diverse groups. My previous design practices were restricted to the conceptual level and rarely concerned the implementation phase. This thesis represented my first experience of possessing complete authority and responsibility and of independently cooperating with a frontend developer in chatbot development.

1.3 Scope

This thesis focuses on investigating the co-design of human-chatbot interaction within the context of refugees and migrants during chatbot development. It focuses on the collaboration amongst designers, users, and developers on their joint effort to create a conversational user interface.

In the ERICS project, refugees and migrants will be initiating a conversation with the chatbot to seek desirable information regarding everyday life. The topics are mainly about finding an internship, learning new languages, student financing, applying for schools and universities, and vocational training. These topics are subsumed under the “Learn” section on HBG website. However, after the practical design task was completed, the scope of question-answering of the chatbot was extended to cover all the topics in the HBG website. This change is discussed in details in section 4.4 Limitations & Future Research.

By leveraging the user-centered design and co-design methods, this thesis revolves around user interface design enabling natural and efficient dialogues. The intelligence of the chatbot to be developed can be immature at the beginning. As a result, motivating and allowing the users to train the chatbot is also embodied in conversations. However, the research of technology and the database building are excluded in the thesis. The result of this work is a series of artifacts including a chatbot avatar, conversational flows, and the graphical user interface. These artifacts will define a conversational interface for the first pilot chatbot in the ERICS project.

1.4 Objective & Research Question

The objective of the ERICS project is to design a chatbot, with a goal of helping refugees and migrants access information required in their everyday lives. Over the past few decades, user involvement in the design process has become commonplace (Nieminen, 2015). Users are the best experts regarding their own everyday lives and the collaboration between designers and users, or other stakeholders is beneficial for improving the process of generating ideas in product or service development (Kaasinen, Koskela-Huotari, Ikonen, & Niemeléi, 2013; Steen, 2013; Steen, Manschot, & De Koning, 2011). The collective creativity between designers and end-users and other stakeholders in the design process is called co-design (Kleinsmann & Valkenburg, 2008; Sanders & Stappers, 2008; Steen, 2013; Vaajakallio, Lee, & Mattelmäki, 2009).

A great deal of previous research has proposed using a co-design approach, especially when developing products and services for economically or socially marginalized people (Arce, 2004; Hussain, Sanders, & Steinert, 2012). Various studies have assessed the efficacy of leveraging the co-design approach in the context of refugees and migrants. Xu et al. claim the participatory approach can be an excellent fit to refugees and migrants' multi-cultural nature enabling a holistic understanding of them (Xu, Maitland, & Tomaszewski, 2015). By conducting co-design workshops, Daiute (2010) successfully explored refugees' issues in different societies with teenagers affected by the war in former Yugoslavia (Daiute, 2010). In the field of Human-Computer Interaction, Fisher et al. (2016) indicate the adaptation and application of co-design techniques can benefit understanding how refugee and immigrant youth enact as an information intermediary in their communities (Fisher, Yefimova, & Bishop, 2016). However, despite the growing scholarly interest regarding the undertaking of product or service design for refugees and migrants, to date, there is a lack of empirical and theoretical research of chatbot development by applying co-design approach. Therefore, the main research question investigated in this thesis is:

How can a co-design approach contribute to designing a chatbot supporting social integration within the context of refugees and migrants?

1.5 Research Approach

This thesis followed a *research through design* approach, where the co-design process functioned as a form of research contributing to a design activity (Archer, 1995). In a *research through design* process, prototypes can show a hypothesis, as they are produced from investigated research questions (Stappers, 2007). According to Lucero, (2009), the design concepts represented through prototypes can be on the basis of research methods or theories. Knowledge can be generated by designing the artifacts and evaluating the usage or the artifacts themselves. Later, the gained knowledge can contribute to design recommendations, theories or frameworks (Lucero, 2009). Furthermore, the act of designing in the thesis emphasizes the exploratory nature of design research, and the reflection of the design process is able to bring about new knowledge.

As a *research through design* following the ethics of user-centered design and co-design approaches, this thesis began in April 2018, spanning over eight months of user study, co-design, and practical design activities. By integrating theories of *say, do and make* (Sanders, 2003), the *framework of co-design processes* (Sanders & Stappers, 2014), and *joint inquiry and imagination* (Steen, 2013), the design project was iteratively deployed in four main phases: pre-design, generative, evaluative, and post-design (see Figure 1). The logic of the co-design approach is discussed in section 2.1.2. Co-design Approaches and section 3.1.1. The Co-design Approach for the ERICS Project.

In the *pre-design* phase, the quantitative and qualitative research methods were employed. I started with a month-long exploration of the background, design contexts, users, knowledge of co-design and chatbot design. To reveal how a co-design approach should be applied, a literature review was accomplished focusing on clarifying the definition of the co-design paradigm, roles, and competencies of co-design actors, and appropriate co-design tools. Furthermore, I also defined the notion and operating principles of chatbots, justified the necessity of anthropomorphism for a chatbot. Following the research of conversation design, I benchmarked relevant empirical cases, and identified methods to evaluate chatbot experience. Meanwhile, in practice,

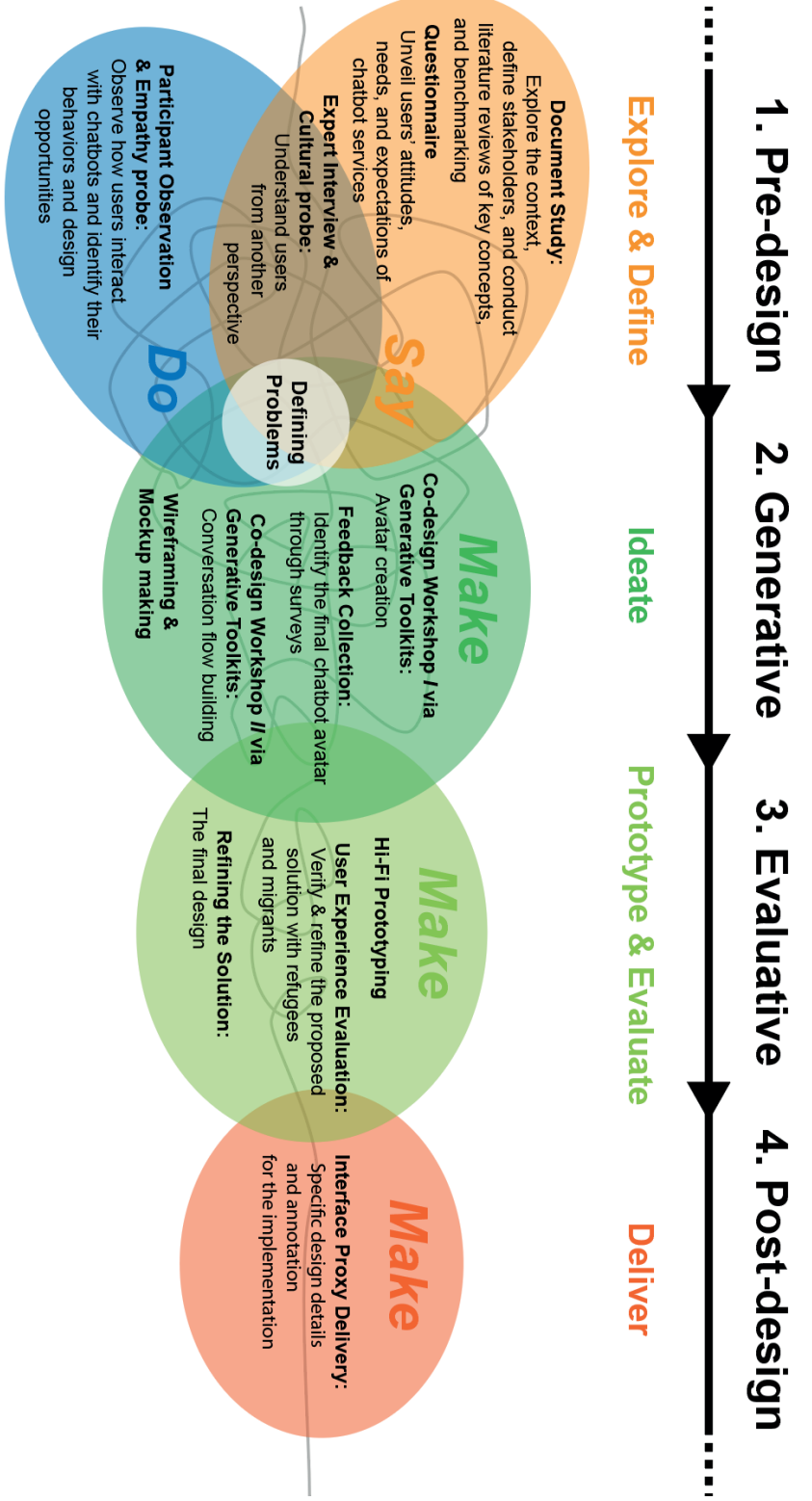


Figure 1. The co-design process in the ERICS project

problems and design questions are defined through user research. With the support of questionnaires and document studies, I listened to what migrants said about their attitudes, demands, and experience of information seeking and chatbot services and revealed refugees' psychological health. I also interviewed two experts from the International House Helsinki (IHH)[4]. Besides, I conducted a participant observation where the users' behaviors, experiences, and mental models[5] were discovered by observing how the users (a refugee and two migrants from Arffman Consulting[6]) interacted with the early prototype of the chatbot. After that, I further explored their thoughts, feelings, and attitudes through empathy probes.

Based on the exploration of *Say* and *Do*, three design questions were defined as followed:

- What constitutes an empathic chatbot avatar that is capable of motivating engagement with refugees and migrants?
- What represents a natural conversation flow between the user and the chatbot in the context of questions answering?
- How can users be encouraged to build chatbot training data by engaging in natural conversation?

In the second *generative* phase, by leveraging generative toolkits, concrete solutions were conceived through two co-design workshops addressing the defined design questions. In the first workshop, five co-design participants (two migrants, one in-house developer, one research assistant and one designer[the author of this thesis, also as a migrant]) created three proposals to the chatbot avatars with different system personalities. After that, the final avatar was selected by thirty refugees from the reception center of Luona[7]

[4] International House Helsinki is a collaborative public service provider between the metropolitan municipalities to help immigrants in aspects such as work and tax.

[5] A mental model is an explanation of someone's thought process about how something works in the real world. It is a representation of the surrounding world, the relationships between its various parts and a person's intuitive perception about his or her own acts and their consequences. Mental models can help shape behavior and set an approach to solving problems and doing tasks.

[6] Arffman Consulting is a private Finnish institution that mostly organizes language and culture lessons for migrants and refugees alike.

[7] Luona is a Finnish social service and healthcare provider.

responding to a survey. In the second workshop, the co-design participants remained unchanged, and they jointly built conversational flows for the chatbot based on their experience and knowledge.

In the third *evaluative* phase, I crafted a prototype demonstrating the defined concept visually. By utilizing the *Method of Assessment of eXperience* (MAX) (Cavalcante, Rivero, & Conte, 2015) and the *property checklists* (Jordan, 1998), the prototype – the design concept – was evaluated by refugees and migrants. This thesis first invited seven migrants at IHH to test the prototype. The next evaluation session occurred at Arffman Consulting with ten refugees participating. The refined chatbot design was completed based on the collected feedback and suggestions of refinement.

Lastly, in the phase of *post-design*, interface proxies clarifying the final chatbot concept were produced to guide and facilitate the frontend developer implementing the chatbot.

1.6 Thesis Structure

The structure of this thesis is organized in the following four chapters: Introduction, Literature review, Co-designing a chatbot, Discussion and Conclusion.

Chapter 1 (Introduction) mainly outlines the background, the motivation, the scope, the objective and research question, the research approach as well as the structure of this thesis work.

Chapter 2 (Literature review) situates the topics in co-design approach, chatbot and user experience evaluation. This chapter clarifies co-design and its relevant tools and design process. It defines a chatbot and interprets the necessity of anthropomorphism in chatbots. Also, it investigates the conversation design and benchmarks a related case. Finally, this chapter identifies the user experience evaluation methods.

Chapter 3 (Co-designing a chatbot) then elaborates the co-design process of the chatbot in the ERICS project. By collaborating with the users (i.e., refugees and migrants) and the frontend developer, this thesis explores the problem and defines the design questions, ideate and prototype the proposal with iterative improvements, and then eventually evaluates and refine it. This chapter ends by showing the design specification facilitating the implementation of the chatbot.

In Chapter 4 (Discussion & Conclusion), the thesis focuses on the reflection of the co-design tools and participants in practice and answers to the research question. The limitations and future research of this thesis are also articulated. Eventually, the chapter presents the implications and conclusion.

2. Literature Review

2.1. Co-design

2.1.1 The Definition of Co-design

The term co-design has been used to refer to the work described in this thesis. Many theoretical and empirical studies describe the concept of co-design. In the field of design, the term co-design refers to proactively involving non-designers in the design process (Mattelmäki, & Sleeswijk Visser, 2011; Sanders, Singh, & Braun, 2018). Non-designers can be end-users, internal team members specialized in non-design disciplines, external stakeholders or anyone affected by design. The central idea of co-design is designing with people, rather than designing for people (Sanders, Singh, & Braun, 2018). To be specifically, Brandt & Eriksen (2010, p. 19) claim co-design is a series of workshops "for sketching and trying out possibilities," where designers and participants congregate. According to Mattelmäki and Sleeswijk Visser (2011, p. 3), co-design "is about users or more generally, people imagining and planning with issues that are not-yet-existing and utilizing the skills that are in the core of professional design competence." Sander and Dandavate (1999, p. 89) argue that co-design facilitates "exchange between people who experience products, interfaces, systems and spaces and people who design for experiencing." This argument is in line with what Mattelmäki and Sleeswijk Visser (2011) mention that co-design emphasize experience-driven mindsets. Similarly, Kleinsmann claims that

> co-design is the process in which actors from different disciplines share their knowledge about both the design process and the design content. They do that in order to create a shared understanding on both aspects, to be able to integrate and explore their knowledge and to achieve the larger common objective: the new product to be designed. (Kleinsmann, 2006, p. 30)

Designers must consider a number of specific factors when they implement co-design activities. For example, Mattelmäki et al. (2011) describe four elements of co-design. Firstly, they emphasize the role of "people that are affected by the design" (e.g., end-users and other stakeholders). In user-centered design, it is essential to hear their voices in the design process. Secondly, they argue that all participants need to collaborate to build an

efficient dialogue via workshop-like activities. Thirdly, they indicate the necessity of using methods and tools to empower participants who are not trained in design to express their experiences and ideas and generate visually tangible artifacts that contribute to the final design. Fourthly, they stress cooperation in which participants exchange thoughts and jointly create solutions. Furthermore, Donetto et al. (2015) claim that co-design is driven by shared ownership in which participants not only 'have a say,' but are also entitled to make decisions during solution development. Sanders et al. (2008) suggest that co-design can occur across the whole design process, and it primarily benefits the early front end of the design development process, such as the phase of idea generation.

To summarize the discussion above, co-design includes six features (Donetto, Pierri, Tsianakas, & Robert, 2015; Mattelmäki & Sleeswijk Visser, 2011; Sanders & Stappers, 2008):

a) Diversity of participants: the more people with different backgrounds, experiences, interests, and roles that are involved, the more holistic the solution is;
b) Engagement: co-design occurs through a series of design events where designers and people who are relevant to the design outcome meet together;
c) Facilitation with methods and tools: co-design utilizes tools enabling non-designer participants to explore, envision and create in a designerly way;
d) Collaboration: co-design emphasizes joint efforts instead of individual intelligence;
e) Shared authority: in co-design, all the participants explore, express, and develop in an egalitarian way;
f) Spanning the design process: co-design can appear in different stages of the design process, especially the ideation stage.

2.1.2 Co-design Approaches

Prior to demystifying the co-design approach, it is necessary to clarify those terminologies which are used by this thesis to describe the application of co-design. According to Sander et al. (2010), this thesis defines the following concepts in the field of co-design:

a) Tool: a material item applied in co-design events
b) Toolkit: an assemblage of tools that are jointly utilized to achieve a goal
c) Technique: a logic and instruction describing how to use tools and toolkits
d) Method: a strategical combination of tools, toolkits, and techniques for serving a defined purpose
e) Approach: a global mindset and structure guiding the implementation of the research plan.

Based on the *theory of inquiry* by John Dewey, Steen (2013) proposes a co-design approach called *joint inquiry and imagination*, in which designers, users, and stakeholders jointly perceive the problems (*explore & define*), conceive solutions (*ideate*), and then implement and evaluate solutions (*prototype & evaluate*). Firstly, enabling users to express and share their personal experience is critical for uncovering the unknown and identifying a possible problematic situation. Secondly, designers can define and articulate the problem. During the phase of exploring and defining, co-design practitioners perceive the situation and make problems unambiguous by accessing the users' experiences. Following the defined problems, co-design participants explore and develop solutions through conception, in which they imagine and envisage possible ideal situations. The higher the number of different people seeing the issue from diverse angles and backgrounds is involved, the more holistic the solution is. Finally, designers invite end users to experience the solution and evaluate how does it effectively settle the problem before jumping to conclusions. Steen claims that involving people to test the solution and improving the solution based on their feedback also reflects the ethics of co-design.

Similarlly, Sanders and Stappers (2014) formalize a co-design approach

consisting of four phases: *pre-design*, *generative*, *evaluative* and *post-design*. They depict these phases as the convoluted timeline shown in Figure 2. Here, the left black spot illustrates that the design opportunities have been identified, and the right black place indicates that the final design outcome has been utilized. In the first phase of *pre-design*, the designers and co-designers devote themselves to understanding the past, present, and future context of people's experience. Next, in the *generative* phase, all the co-design actors jointly *generate* ideas, concepts, and design proposals. Afterward, they move to the evaluative phrase to verify the hedonic and pragmatic aspects of a product or service through prototypes and evaluations. Eventually, in *post-design*, the defined product or service is released onto the market, and the design team monitors how people experience the product or service in real life. The result can be a new initiation of another design process.

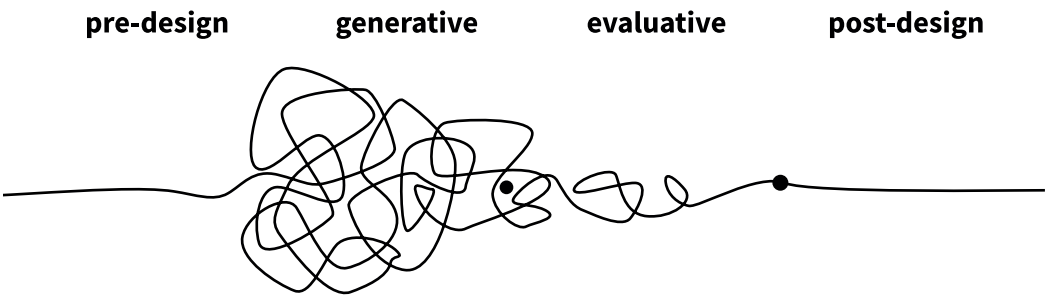


Figure 2. Phases along a timeline of the design process (from Sanders & Stappers, 2014)

As previously discussed, co-design stresses an experience-driven mindset (Mattelmäki & Sleeswijk Visser, 2011). In terms of accessing user experience and empathizing with users in co-design, Sanders (2003) suggests an approach called *say, do and make*, which is divided by different focus and information (see Figure 3). She argues that "explicit knowledge" can be extracted by listening to what people express in words in research activities, such as questionnaires and interviews. At the same time, observing what people do, can lead to generating "observable information or observed experience" (e.g., field research). Uncovering what people 'make' to convey

their thoughts, feelings, and dreams, is beneficial for producing “tacit knowledge,” which cannotbe readily indicated by words (Sanders, 2003, p. 20). The session of ‘make’ is mostly conducted in physical or visual forms, and it plays an important role in exploring and creating design solutions. In the later stage of ‘make,’ designers build the final interactive prototype (Sanders, 2003). It is critical to simultaneously explore all three categories to establish an empathic understanding of the users (Sanders, 2003).

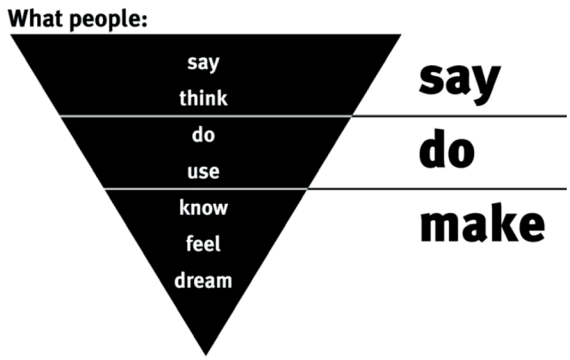


Figure 3. What people say, do and make (from Sanders, 2003)

The co-design approaches mentioned earlier describe different structures of problem-solving taken by co-design participants when confronted with different challenges. However, regardless of the paradigm embedded in those approaches, they should be flexibly deployed and applied according to a specific design context.

2.1.3 Co-design Tools

> *"Co-design is inherently a set of tools for collaborative engagement, i.e., instruments, and competence by the future users to utilize their experiences and creativity for design." (Mattelmäki & Sleeswijk Visser, 2011)*

To empower users to reflect on their own experiences and create design concepts in a designerly way, cultural probes, empathy probes, and generative toolkits are the commonly used tools in the practice of co-design (Mattelmäki, 2005; Mattelmäki & Battarbee, 2002; Sanders, 2003).

Gaver et al. first introduced cultural probes (Gaver, Dunne, & Pacenti, 1999) as a way of exploratory and design-oriented self-documentation method. Cultural probes are visually well-designed packages inviting users to reflect on and convey their feelings, attitudes, and experiences, such as an illustrated postcard for answering questions. Cultural probes can enable designers to be inspired and informed (Gaver, Dunne, & Pacenti, 1999; Mattelmäki, 2005). Lucero et al. (2007) investigate the relation between the results of the cultural probes and the final design concepts and propose six benefits allowing designers to 1) enter an intimate space, 2) discover unexpected uses, 3) gather requirements, 4) look into participants' lives, 5) shift focus, and 6) find inspiration for new concepts.

Similarly, empathy probes collect inspiring data, which illustrates users' internal thoughts. However, empathy probes can achieve a more holistic user understanding through in-depth interviews where users further interpret the data (Mattelmäki, 2005; Mattelmäki & Battarbee, 2002). Mattelmäki (2005) identifies four reasons why designers apply empathy probes in co-design activities: inspiration, information, participation, and dialogue. She claims that probes contribute to enhancing designers' creative thinking and generating new insights. Moreover, probes gather information about user needs, attitudes, and experiences. She also stresses that in co-design, it is critical to empowering users by providing tools to "experiment, observe and potentially record their own experience," as they can be creative people involved in the

design process directly. Finally, probes play an essential role in establishing a conversation and a direct interaction between designers and users.

Generative toolkits consist of a wide variety of components for co-design participants to build design artifacts (Sanders & Stappers, 2014). These artifacts enable participants to access and express their emotions, subjective perspectives, and experience (Sanders, 2000). Generative toolkits enable non-designers to be 'qualified' to participate in the co-design process. Non-designers are capable of visually reflecting on and representing their ideas in a designerly way (Sanders & Stappers, 2014). Generative toolkits also "serve as a common ground for connecting the thoughts and ideas of people from different disciplines and perspectives" (Sanders, 2003, p. 21). Additionally, generative toolkits can support designers to discover "as-yet unknown, undefined, and/or unanticipated user or consumer needs" (Sanders, 2003, p. 21).

According to Sanders and Stappers (2014), the application of probes spans the period from *pre-design* to *generative* research in the co-design process defined by them. They are especially helpful to understand people's experiences because probes function as a means to collect data regarding user experience in the past and present context. Generative toolkits serve as a core vehicle enabling generative research, in which both designers and co-designers can share common languages and exchange ideas to devise the design proposals.

In summary, both probes and generative toolkits are beneficial for building a bridge and a common language between designers and co-designers who are not professionally trained in design. They provide participants a means to express their attitudes, needs, experiences, and ideas explicitly and thus exchange insights with designers.

2.1.4 The Roles and Competencies of Co-design Actors

As noted by Sanders (2003), in a traditional user-centered design process, the researcher collects data and generates insights regarding user needs by observing and interviewing users. After that, the designer, who is often also the researcher, receives all this information and develops solutions while considering technology. In such design processes, the user is regarded merely as a passive information provider.

By contrast, the role of design partners in co-design events is radically changed. Co-design is driven by the belief that users are creative and supportive in the design process (Vaajakallio & Mattelmäki, 2014). Users are qualified to speak for themselves and intensely perceive their attitudes, motivations, and needs. As an “expert of their experiences,” users play a critical role in knowledge development (Sanders & Stappers, 2008, p. 12). In addition, users invited to co-design activities can produce fresh ideas and design proposals by leveraging their diverse competence and creativity (Mattelmäki & Sleeswijk, 2011). Accordingly, people directly included in the design process can become ‘co-designers’ (Sanders & Stappers, 2008). Mattelmäki and Sleeswijk Visser (2011) also emphasize that users involved in the design process can be ‘evaluators’ of ideas, concepts and the final product or service.

In co-design of software development, developers are considered as essential participants. Through an academic discussion about co-design with Yngve Sundblad, Lucero Vera (2007) points out, in his doctoral thesis, that the developers should also be involved in the design of solutions, as they possess the technical knowledge which significantly affects the final product.

As for designers as ‘organizers,’ they need to plan and implement co-design events in which stakeholders purposefully congregate. Design is a specialized discipline which incorporates a wide variety of skills, knowledge, and approaches. This expertise will not suddenly emerge when ordinary people are invited to co-design events (Buxton, 2005). Hence, designers are required

to act as ‘facilitators’ via guiding and providing participants with supportive tools, which enable them to think and make in a designerly way (Sanders & Stappers, 2008; Vaajakallio & Mattelmäki, 2007). Additionally, designers frequently participate in the collaborative process as ‘contributors’ (Mattelmäki & Sleeswijk Visser, 2011). Although designers possess expertise that other actors do not have in the field of design, they still build the consensus about democratic decision making in solution development with other non-designer participants (Hussain, Sanders, & Steinert, 2012; Sanders & Stappers, 2008).

In co-design of software development, users and developers play essential roles as ‘experts,’ ‘co-designers’ and ‘evaluators.’ Designers take steps back and share the authority and ownership with other participants by acting as ‘organizers,’ ‘facilitators’ and ‘contributors.’

In terms of competencies mastered by the co-design participants, Nieminen (2015) proposes a User-Centered Design (UCD) Competency Model (see Figure 4). This model describes twelve of the most relevant competencies for user-centered design in four categories of User Strengths, Soft Skills, Designer Strengths, and Hard Skills:

User Strengths:

1) Subject Domain Experience: (Tacit) knowledge, crafts, and skills gained by (long) experience.
2) Context Availability: Access to real context of use, may be restricted or difficult to arrange.
3) User Cultures, Social Networks, and Practices: Each user groups have their own language and culture that can be difficult to grasp and utilize in a design project.

Soft Skills:

4) Communication: Well-functioning and democratic communications are the most time consuming and critical part of design work.
5) Multidisciplinarity and Collaboration: Multiple points of view increase the eventual impact of design.
6) Motivation and Ambitions: Intrinsic motivation, self-improvement,

professional ambitions, competitive salary, and manageable workload create solid design conditions.

Designer Strengths:

7) User Involvement: Capabilities in selecting the right users and working with them effectively.
8) Problem-solving and Designerly ways: Attitude and determination towards designing a change for the better and having the means to realize it responsibly.
9) Conceptualization, Visualization, and Validation: Skills in creating product concepts, prototyping, and testing.

Hard Skills:

10) Process and Methods: Process, management, and methodological excellence and the skills to adapt them.
11) Technology and Market Potential: Awareness of advances in available technologies and relevant trends at target markets.
12) Subject Domain Knowledge: Knowledge and skills gained by education and certification.

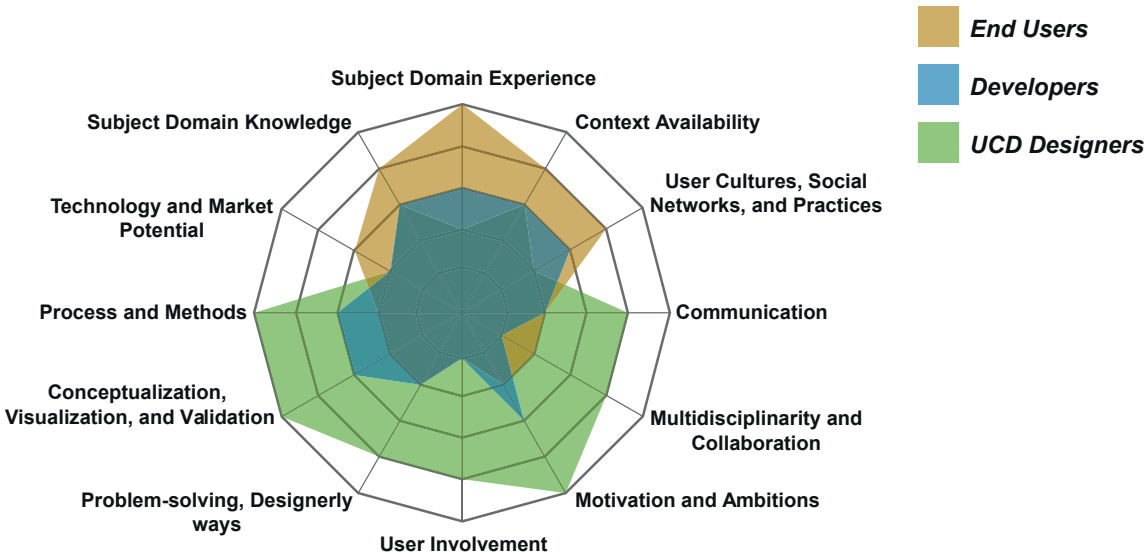


Figure 4. The UCD Competency Model (from Nieminen, 2015)

According to Nieminen (2015), users gain experiential knowledge not only from work tasks and context of use but also from user cultures, networks, and practices, contributing as a resource for design. Subject Domain refers to “the domain of expertise of the target users of the design project”, and Subject Domain Experience refers to “the specialized skills and knowledge that can be gained through personal experience” (user experience). The usage of products is often situated and firmly coupled with the location, time, and other contextual factors surrounding the actual use. Therefore, the context of use is a critical factor in UCD. Also, users understand the culture, to which the product is going to fit. This understanding of culture can be difficult to be obtained by an outside observer, especially when the user group demonstrates behavior or language which can only be accessed by a member of the tightly knit community.

Designers’ competencies consist of the core values in UCD: user involvement, problem-solving in a designerly fashion, and professional skills in conceptualizing, visualizing and testing the design solutions. Designers are trained to identify and engage the right people at the right time in the design process and have the right interaction methods to foster their participation and creativity. UCD is essentially an attempt to discover users’ problems and conceive solutions to improve their lives. Designers can generate design concepts in a concrete and visualized form. They then evaluate the concepts to find the best available solution. Besides, designers also master hard skills gained mainly by training and education, such as managing the design process and employing design methods.

Nieminen also mentions, in software development, developers show up as generalists tightly coupled to projects. They have a good grasp of technology, but lack of direct contacts to users or their contexts is evident. In the meantime, knowledge regarding business opportunities and the subject domain is necessary for UCD, which can be brought by customers, sales and experts.

For a successful UCD, all the design participants should be skilled at communication and collaboration with multidisciplinary attitudes and have

motivation for contributing to the project. Communication is the most time-consuming but essential part of a design project, as it enables effective teamwork and information sharing among all participating stakeholders. Understanding a complex design task requires expertise from different fields. Multiple points of view enhance the impact of the design. The motivation of the design participants in a design project is essential to its success. Motivation can arise from personal interest toward the topic of the project, the satisfaction gained from excellent utilization of one's abilities and skills, or potential to learn new things.

2.2 Chatbot

2.2.1 The Definition of Chatbot

To date, many web-based chatbots have been deployed with purposes, such as information seeking, site guidance, and FAQ (Frequently Answered Questions) answering. These chatbots cover diverse limited domains like customer service, education, website supports, and entertainment. The term, chatbot, indicates a machine system designed to simulate and reproduce an intelligent conversation with users (Angga, Fachri, Elevanita, & Agushinta, 2015; Di Gaetano, & Diliberto, 2018; Huang, Zhou, & Yang, 2007; Shawar, & Atwell, 2005). As a rule, chatbot services are delivered by the multi-turn Question Answering (QA)[8] (Huang, Zhou, & Yang, 2007). Therefore, chatbots are also called conversational agents. Chatbots fall into two main categories: auditory-based and textual-interface-based. During such chatbot services, Sansonnet et al. (2006) claim human users expect chatbots to understand their textual input and then generate appropriate responses. In order to produce the responses, chatbots require natural language processing techniques, dialog management modules, and external knowledge bases (e.g., corpora of data). The natural language processing functions as the basic algorithm[9] to parse the input of texts, and the dialog management modules manipulate the conversational process. However, some chatbots directly capture keywords from the input and respond with the most matching keyword from the corpora. The larger the base of knowledge is, the abler the chatbot is to answer an obvious question (Angga, Fachri, Elevanita, & Agushinta, 2015).

[8] A computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language.

[9] In computer science, an algorithm is an unambiguous specification of how to solve a class of problems. Algorithms can perform calculation, data processing, automated reasoning, and other tasks.

2.2.2 Rationale

According to Quorteroni et al. (2007), the process of Questions Answering (QA) by a chatbot consists of the following five steps (Quarteroni & Manandhar, 2007):

- Firstly, either a chatbot takes the initiative to deliver a greeting message, or the users start the conversation by sending a direct question (The interaction strategy deployed for the chatbot determines that who is leading the dialogue: the user or the system [Cahn, 2017]).
- Secondly, the dialogue management module parses the query to identify whether it is relevant to the preceding questions. If the result is 'no,' then the question is submitted to the QA component. If it is 'yes' but elliptic or anaphoric, the system harnesses preceding queries to develop it with complementary keywords or replace it with corresponding entities. This action aims to create a revised question. Afterward, the refined query is sent to the users to identify whether it is exactly what they mean. The users are required to reformulate their utterance if the refined question is incorrect.
- After obtaining an appropriate query, the chatbot sends the corre sponding answer generated from the QA components to the users.
- Subsequently, the chatbot inquires about whether the users are interested in a follow-up session. They can input a query, if they have more requests.
- Lastly, the chatbot exchanges greetings with users when they want to terminate the conversation.

Based on the process of QA above, Quorteroni et al. summarize the dialogue move sets from the system and the user (Quarteroni & Manandhar, 2007) (see table 1):

Table 1. The dialogue move sets from the system and the user (from Quarteroni & Manandhar, 2007)

User Move	Description	User Move	Description
Greet	Conversation opening	Greet	Conversation opening
Quit	Conversation closing	Quit	Conversation closing
Ask questions	Users ask questions	Ask questions	Answer
Acknowledge	Acknowledgment of previous utterance, e.g. 'Thanks.'	Acknowledge	Acknowledgment of previous utterance, e.g., 'Ok.'
Clarify request	Clarification request	Follow up	Proposal to continue session

2.2.3 Anthropomorphism

In the chatbot context, anthropomorphism indicates assigning human-like characteristics to non-human agents (Seeger, Pfeiffer, & Heinzl, 2017). In light of this concept, anthropomorphic chatbots are digital representations of machine conversation systems giving users an illusion that they are conversing with a humanoid being.

According to Nan et al. (2006), the necessity of considering anthropomorphism as an essential factor in chatbot design attributes to the following two reasons:

1) Anthropomorphizing chatbots contributes to making the interaction between the user and the chatbot engaging significantly. They are able to increase users' trust towards information source and thus boost perceived source credibility[10]. High source credibility is critical for the HBG website aiming to deliver trustworthy and unbiased information to newcomers in Germany.

2) Nan et al., argue that the presence of a web-based anthropomorphic chatbot allows more human communication information to be transmitted, which can lead to strong social presence[11]. It is almost axiomatic that such social presence is capable of shortening the physical distance between the users and the information source provided by the chatbot. As a result, users can have a perception of engagement with the help of "interpersonal interactions" and tend to trust the information source and then enhance the perceived source credibility of the website (Nan, Anghelcev, Myers, Sar, & Faber, 2006, p. 616).

[10] A computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language.

[11] In computer science, an algorithm is an unambiguous specification of how to solve a class of problems. Algorithms can perform calculation, data processing, automated reasoning, and other tasks.

2.2.4 Conversation Design

According to *Actions on Google*, conversation design is defined as a design language based on human conversation. In chatbots, with considering user needs and technological limitations, conversation flows are built for achieving a particular goal (“What Is Conversation Design,” n.d.). In other words, conversation design is about teaching computers to be fluent in human conversation and its conventions. Three principles suggested to be taken into consideration when designing the conversation for chatbots: 1) start with what humans do, 2) adapt to technical limitations, and 3) leverage technical strengths (“What Is Conversation Design,” n.d.). Language patterns of human have evolved over the past hundred thousand years, and breaking the patterns is almost impossible and required to pay a heavy price. Therefore, chatbots are suggested to adapt to the communication system that users learned first and knew best so as to create an intuitive and frictionless conversational
experience.

In some way, chatbots can exceed human capabilities. For example, they possess exhaustless patience, and the interlocutor does not need redundant responses such as filler words or other formulaic languages (e.g., ums and ahs). In that case, leveraging the technical strengths can enhance the conversational experience (“What Is Conversation Design,” n.d.). To date, however, chatbots still fall short of human capabilities. Chatbots sometimes cannot correctly understand the user intent within a specific context. Consequently, the conversation design for chatbot requires adaption to technical limitations.

Conversation design processes

As stated by *Actions on Google*, a conversation design process includes three steps (see Figure 5):

a) gathering requirements,
b) high-level design, and
c) detailed design.

Although the process is illustrated linearly, in practice, these are discrete steps, and they are unnecessary to be fully completed before the next is begun. “To borrow software development terms, the conversation design process should be more ‘agile,’ and less ‘waterfall,’ with considerable testing and iteration did throughout. ” (“Conversation Design Process,” n.d.). Firstly, the user is required to be defined. Who are they? What are their goals? What is their context? Considering technical limitations, level of effort, and timeline, what use cases will be supported? After defining the user and the context, a persona representing the brand and mission can be generated. Subsequently, it is time to start working on the sample dialogues. Sample dialogues are the key to designing the conversation, as they give a quick, low-fidelity sense of the ‘sound-and-feel’ of the interaction. Besides, it illustrates the flow that the user will experience without technical distractions of code notation, complex flow diagrams, and recognition-grammar issues. Instead of designing with a screen in mind, designers should focus on the spoken conversation first. As a result, it is beneficial to maintain the thread of the dialogue and create a graphical interface that is suitable for conversation. As soon as some sample dialogues are generated, designers can start testing and iterating on the designs. In the last phase of the process, the detailed design will be specified. The detailed design includes ensuring the features adequately covers the long tail[12] of ways that a conversation can deviate from the most common paths. To account for this, handling for errors and other unlikely or uncommon scenarios will be added. Finally, designers scale the design to help users wherever they are (“Conversation Design Process,” n.d.).

[12] A computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language.

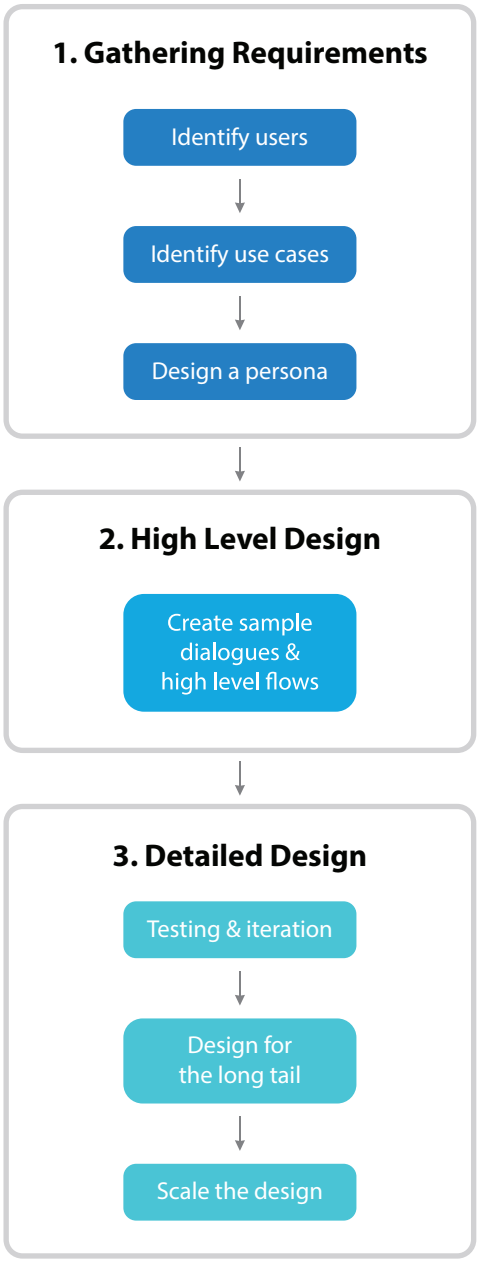


Figure 5. Flowchart of the conversation design process by Actions on Google

Conversational components

Conversational components constitute a prompt. Prompts are the core of the conversational interaction and should be designed for every turn in the dialogue. The display prompts will be presented on devices with screen output. Actions on Google defines the most relevant conversational prompts ("Conversational Components," n.d.):

- **Acknowledgments** (reassure the user that they have been heard and that the chatbot is keeping track of the conversation)
- **Apologies** (possess a transitional social or phatic function)
- **Commands** (indicate actions the user can take)
- **Confirmations** (give users feedback on how their input was understood)
- **Discourse markers** (not only empowers users to correct mistakes immediately, but it also reassures them in a socially and conversationally appropriate way by establishing common ground)
- **Endings** (let the user decide when the conversation should be finished)
- **Errors** (include No-Match and No Inputs)
- **Greetings** (introduce the chatbot itself and make an excellent first impression by showing value)
- **Questions** (get the user to continue the conversation)
- **Suggestions** (help the user answer a question or discover new features)
- **Chips** (is an auto-suggestion button that users can tap to take actions fast and efficiently)

The full guideline of the usage of those conversational components is displayed in Appendix 1: The most relevant conversational components by Actions on Google. In the conversation design process, prompt writing plays a vital role in both high-level and detailed design.

Enhancing the conversational experience

Psychologically speaking, users expect a chatbot to master human-like conversational etiquette. Jain et al. (2018) discover, in the introductory phase, users await the chatbot to say “hi” or “how are you” firstly. This act is crucial, as it starts to build the first impression for the chatbot. Similarly, in the concluding phase, users also expect the exit experience where the chatbot can say “goodbye” even though they may not respond to the chatbot’s farewell. Besides, Jain points out that the users’ expectations on the capabilities of a chatbot can be significantly increased because of their cognitions of a system allowing free interactive patterns with natural language. To reduce the gap of expectation, the users prefer receiving an explicit clarification about what the chatbot is capable of doing. More specifically, the chatbot is required to explicitly interpret its capabilities with examples not only in the initial introduction but also in scenarios where low engagement or failures of dialogues occur. To improve the interaction efficiency, Jain suggests leveraging chips to help users save more time on reply. A chip reduces the interaction cost, which enhances the users’ willingness to be involved in the conversation.

For reinforcing user engagement, *Actions on Google* suggests using user-focused text to keep the conversation on track instead of making the chatbot persona the center of attention. Also, prompts should be informative, concise, short and simple so as to make the chatbot accessible to people of all background and enable users to focus on the conversation quickly. Somewhat, jargon and legalese are not recommended because of the risk of eliciting misunderstanding and mistrust (“Language,” n.d.).

In summary, the user experience of interacting with a chatbot can be enhanced through four actions:

1) adapt to conversational etiquette,
2) clarify capabilities at the start,
3) utilize chips to increase efficiency, and
4) write user-centered and straightforward prompts.

2.2.5 Related Case Study

As the ultimate goal of this thesis is to design a conversational user interface of an anthropomorphic chatbot, existing relevant empirical cases are seen as valuable references and supports. When selecting the case study, I took the following criteria into account:

a) the chatbot is text-based and uses a keyboard as the way of input;
b) the chatbot focuses on a specific domain with specialized knowledge;
c) the chatbot functions as an information provider.

Based on these standards, I conducted a benchmarking on the Tinka chatbot (https://www.t-mobile.at/), which has gained the best chatbot accolade among DAX[13] firms ("T-Mobile Austria wins consumer experience award," 2018). With artificial intelligence, Tinka is a dialogue system acting as a virtual assistant on the T-Mobile Austria website. Tinka is able to provide customers with information about Telekom products, prices and contracts, addressing issues in a chat-like conversation (Jan, 2017). I investigated Tinka in three aspects: visual appearance and personality, graphic user interface design, and interaction design.

Visual appearance and personality

Tinka appears on customers' web interface with a two-dimensional still graphics depicting a young woman with long hair (see Figure 6). The visualization focuses on the upper part of Tinka's body including the head, the torso (shoulders and arms) and occasionally the hands. She is a 459-year-old alien from a satellite city, Arkayn, on the planet Gliese, but she has been living on Earth for more than 27 years (Jan, 2017). Tinka devotes to work as a professional, patient, and humorous customer service agent on the T-Mobile Austria website. In an exception to dedicating herself to solve the customer' problems, Tinka also tells jokes and brain teasers.

[13] The DAX (Deutscher Aktienindex (German stock index)) is a blue chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange.

Figure 6. Jan, F. M. (2017, May). July 8, 2018, TINKA AS FACEBOOK CHATBOT. Retrieved from https://www.welove.ai/en/blog/post/tinka-facebook-chatbot-en.html

Graphic user interface design

The color scheme of the Tinka's chat interface primarily adopts dark violet conveying a broad sense and impression of the outer space, which is consistent with her identity traits (an alien from another galaxy) (see Figure 7). Its slight partial pink responds gracefully to the color theme of T-Mobile's Brand logo as well as the host website. The overall color design of the chat interface is fairly harmonious to fit into the context of the T-Mobile Austria website. As for the layout, the icon of Tinka locates on the website' upper right corner (see Figure 8). When one starts the conversation with her, the built-in chat window extends to a quarter of the whole webpage far right (see Figure 7). It can even be enlarged to the entire window if one would like to concentrate on the dialogue wholeheartedly.

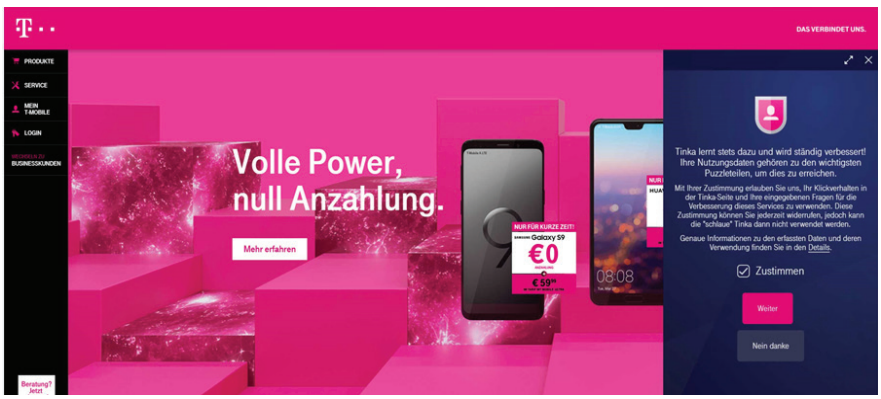


Figure 7. The color of the chat window of Tinka on the T-Mobile Austria website [Digital image]. (n.d.). July 8, 2018, Retrieved from https://www.t-mobile.at/

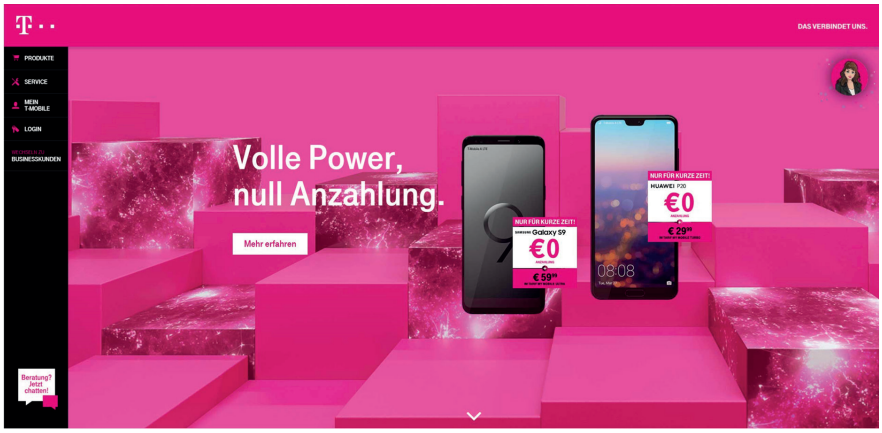


Figure 8. The location of the icon of Tinka on the T-Mobile Austria website [Digital image]. (n.d.). July 8, 2018, Retrieved from https://www.t-mobile.at/

Interaction design

Tinka catches users' attention and engages with them by greeting and telling her name and role proactively when users enter the website (Figure 9, top-left). After users initiate the dialogue, Tinka clarifies some topics, to which she is capable of answering. In addition to allowing users to ask an open question, she also uses chips to recommend some actions (Figure 9, bottom-left). Once Tinka starts to respond, a typing indicator appears, which implies Tinka is typing currently. As for rating the message, Tinka allows users to assess all the response. Users can hover over the prompts and then either select the pop-up thumb up button implying positive feedback or the thumb down button indicating negative feedback (Figure 9, top-right). If one gives a thumb down, then Tinka continues to ask the reason why that answer dissatisfies the user. By utilizing chips ('helpful,' 'relevant,' 'friendly,' and 'precise') as a quick survey, users can quickly report which aspects should be improved (Figure 9, bottom-right).

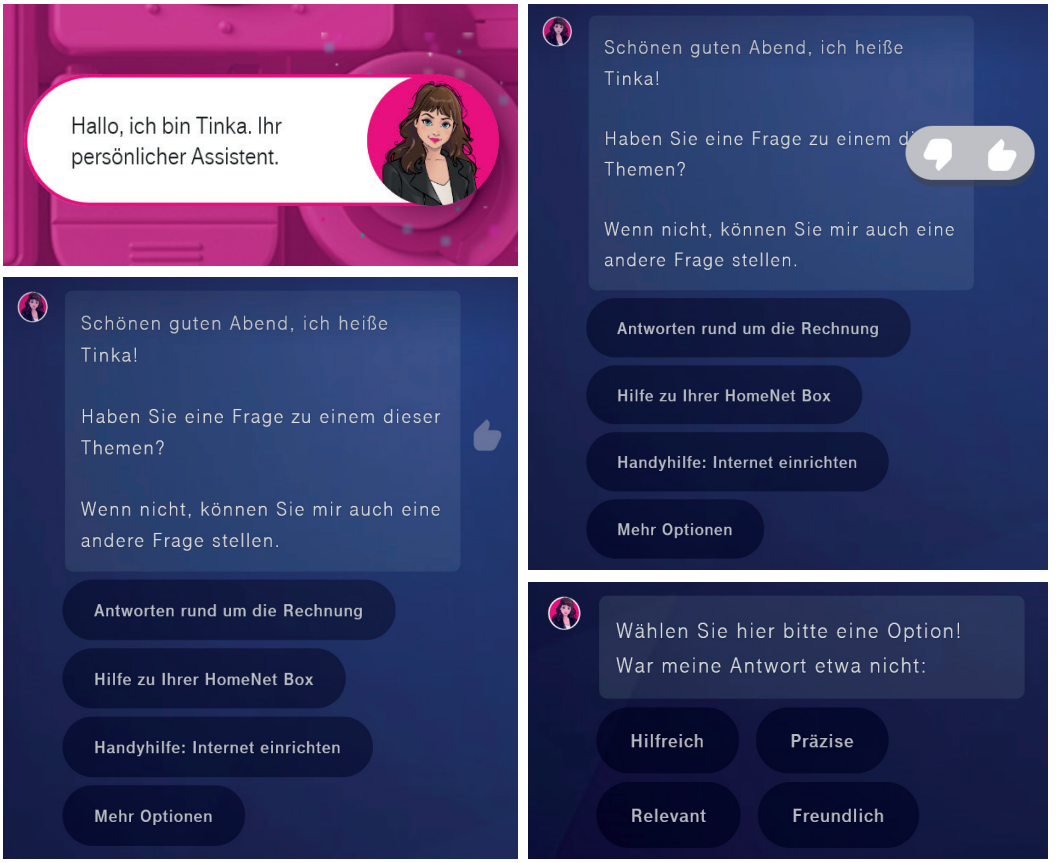


Figure 9. The user interface of Tinka on the T-Mobile Austria website [Digital image]. (n.d.). July 8, 2018, Retrieved from https://www.t-mobile.at/
The proactive greeting of Tinka (top-left); **Rating messages with thumb-up/down buttons** (top-right); **Tinka suggesting topics to be asked and offering option buttons** (bottom-left); **Tinka inquiring the areas of improvement for prompts** (bottom-right).

Findings

Many valuable findings could be taken into consideration in the succedent chatbot design. In terms of visual appearance and personality of the chatbot avatar, this thesis suggests building the avatar according to the qualities of the company brand and the context where it serves. Users much prefer the chatbot with personality

matching its domain. For instance, professionally behaving is expected in a new chatbot, while casual and funny characteristics are compelling in a shopping chatbot (Jain, Kumar, Kota, & Patel, 2018). The representation of chatbot avatars should create engagement with users psychologically through a means of visualization. As regards the graphic user interface, the overall color scheme of the chat interface is proposed to be consistent with the host website to create a seamless and harmonious visual experience.

The initial interaction can start with a proactive greeting to draw users' attention before the conversation begins. Next, a welcome prompt is an excellent means to set expectations of capabilities of the chatbot and provide essential instruction of conversing with the chatbot. These two moves are able to shape the first part of future relationships between chatbots and users (Casey, 2018). During the conversation, using chips can avoid dialogue failures due to users' incorrect input or faulty artificial intelligence. Besides, chips also help users reduce the effort of typing and save more time for them.

2.3 User experience evaluation

2.3.1 The Definition of User Experience

User eXperience (UX) has become a critical concept for understanding and designing the quality in the usage of software products (Vermeeren, Law, Roto, Obrist, Hoonhout, & Väänänen-Vainio-Mattila, 2010). As stated by ISO 9241, UX is interpreted as “a person’s perceptions and responses that result from the use or anticipated use of a product, system or service.” Vermeeren et al. (2010) further define UX as the integration of the user’ beliefs, preferences, feelings, emotions, and physical reactions into a dynamic and evolving whole before, during and after interacting with an object. In software development, usability is another concept frequently mentioned. In line with ISO 9241, usability is defined as the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.”

The boundary between UX and usability is inexplicit. Some studies take the stance that UX is subsumed within the satisfaction component of usability (e.g., Bevan, 2009), whereas some claim that UX is an umbrella term for all user behavior including usability (e.g., Vermeeren, Law, Roto, Obrist, Hoonhout, & Väänänen-Vainio-Mattila, 2010). This thesis adopts the later view in chatbot design and evaluation. As defined above, UX contains all the user’s emotions, physical and psychological responses that occur before, during and after usage. It involves the effectiveness and efficiency of interacting with the chatbot. Additionally, Kocaballi et al. (2018) point out, in measuring UX in conversational interfaces, it is necessary to go beyond usability and explore more about subjective qualities, such as user satisfaction (Kocaballi, Laranjo, & Coiera, 2018).

Given all mentioned so far, UX plays an essential role in software development, especially in product evaluation. Task performance measured through usability testings is not all in UX, and it is also imperative to assess how users perceive the application.

2.3.2 Methods for Validating Chatbots

Through UX evaluations, designers can determine the validity the usage as well as perceive the value of the chatbot. The term, validity, refers to the extent to which any measuring instrument verifies what it claims to measure (Carmines & Zeller, 1979, p. 17) and that the instrument anticipates something useful (Cook, 2004, p. 206). Besides, UX evaluations allow designers to identify how the chatbot should evolve in the future to meet users' varying expectations (Moreno, Seffah, Capilla, & Sanchez-Segura, 2013).

According to ISO 9241-11, in software products (e.g., chatbots), the UX can be evaluated in the two aspects: pragmatic (task performance) and hedonic (user perception). From the pragmatic perspective, the effectiveness of products should be assessed in many aspects such as accuracy, task completion, and chatbot intelligence (Tony, 2017). More specifically, Jordan (1998) proposes *property checklists* in which five important properties of a product should be validated: performance, features, usability, aesthetics, and size.

As for the hedonic perspective, Walker et al. propose a framework called Paradigm for Dialogue System Evaluation, in which subjective factors are estimated such as ease of usage, clarity, naturalness, friendliness, robustness, and willingness to use (Walker, Litman, Kamm, & Abella, 1997). This thesis emphasizes chatbot design from the perspective of user experience design and interaction design instead of the angle of technology development. Therefore, the scope of UX evaluation is formed within the area of graphical user interface and human-chatbot interaction.

Nevertheless, designers or researchers cannot always implement a productive UX evaluation process where users precisely define their emotions, feelings, and attitudes of the verified chatbot (Miles, Greensmith, Schnadelbach, & Garibaldi, 2013). Some evaluation methods may be too superficial or heavy to generate valuable insights (Miles, Greensmith, Schnadelbach, & Garibaldi, 2013; Vermeeren, Law, Roto, Obrist, Hoonhout, &

Väänänen-Vainio-Mattila, 2010). The users evaluating the product may fail to remember interpreting essential informaltion, if they do not have clear guides and facilitation (Cavalcante, Rivero, & Conte, 2015). Therefore, it is necessary to apply an appropriate UX evaluation method which stimulates and guides the users to share their thoughts and feelings regarding using the chatbot.

The *Method for the Assessment of eXperience* (MAX) is a proven UX evaluation method covering both the hedonic and pragmatic aspects. As a post-use UX evaluation method, MAX facilitates the users to assess four aspects (emotion, ease of use, usefulness, and intention to use) with a set of cards displaying avatars and a board. With the intensity ranging from 1 to 4, each avatar on the MAX card demonstrates the user's possible reactions to the evaluated product. To guide the users, MAX presents the following four questions on the board as a prompt: 1) "what did you feel when using it?" (emotion), 2) "was it easy to use?" (ease of use), 3) "was it useful?" (usefulness), and 4) "do you wish to use it?" (intention to use) (Cavalcante, Rivero, & Conte, 2015).

3.
Co-designing a Chatbot

3.1 Explore & Define

3.1.1. The Co-design Approach for the ERICS Project

By reflecting on and integrating the co-design approaches (discussed in section 2.1.2 Co-design Approaches): *Four phases of co-design processes* (Sanders & Stappers, 2014), *joint inquiry and imagination* (Steen, 2013), and *say, do and make* (Sanders, 2003), a co-design process is deployed in the ERICS project (see Figure 15). This co-design process is composed of four stages, being the *pre-design* phase, the *generative* phase, the *evaluative* phase, and the *post-design* phase. These stages are interdependent and ideally conducted in an iterative process.

The *pre-design* phase is the stage in which designers explore and define the problems by listening to what the users (i.e., refugees and migrants) and other stakeholders say about experiences of seeking information on everyday life. The problems can be further investigated through observations where designers identify how the users interact with chatbots and then reveal their behaviors, mental models, and possible design opportunities. After that, the defined problems can be transferred to concrete design questions for the following concept generation. The *generative* phase entails co-design participants (i.e., designers, users, and developers) using co-design toolkits to conceive possible chatbots in workshop-like events. Once the concrete concept is formed, designers build the prototype. They then evaluate the design proposal with users and jointly improve the final solution in the *evaluative* phase. Eventually, in the phase of *post-design*, designers communicate with developers and provide them with interface proxies and annotations facilitating the implementation of the chatbot.

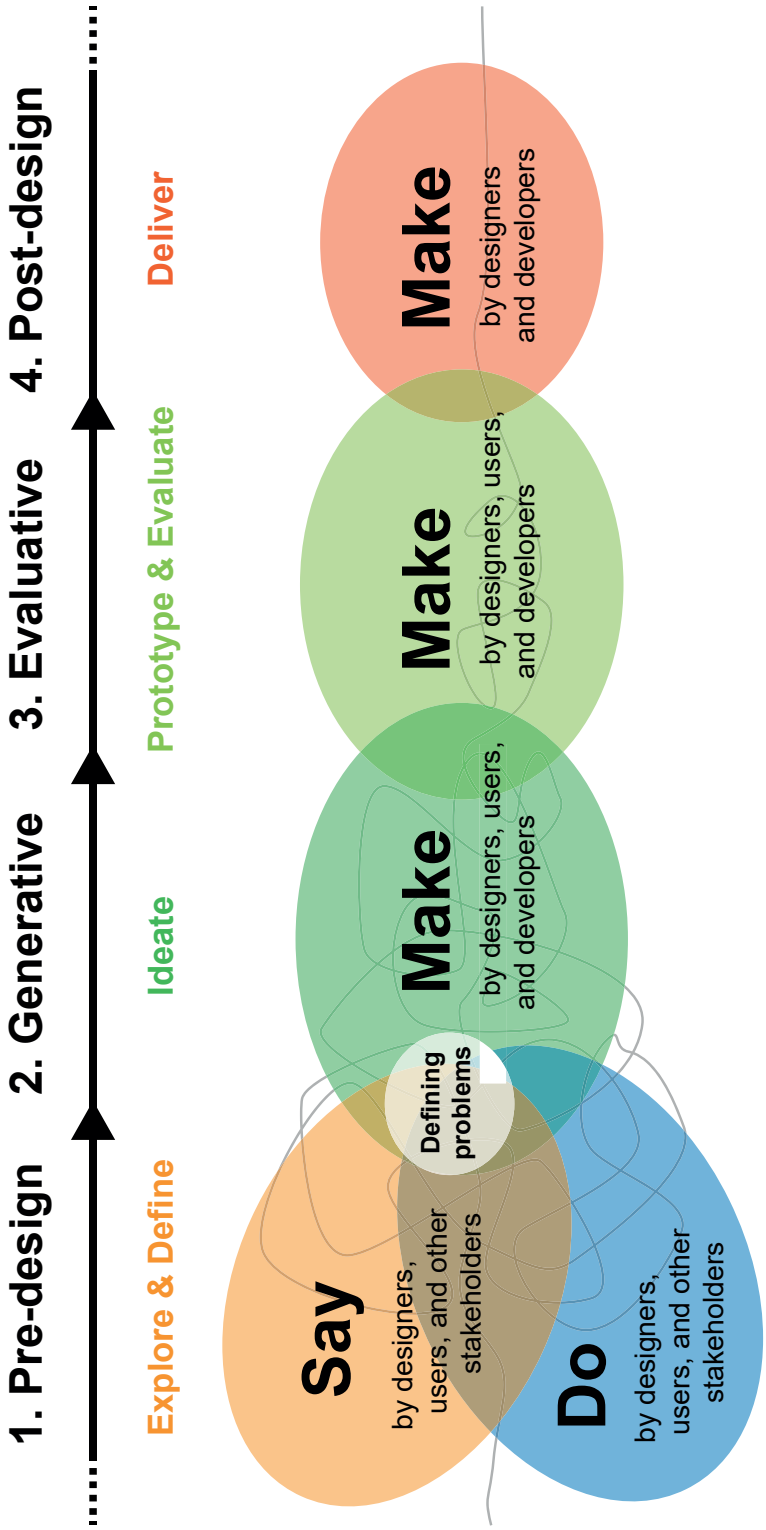


Figure 10. The co-design approach in the ERICS project

3.1.2 The Defined User Groups in Co-design

The goal of the HBG website is to provide information about social integration for newcomers in Germany incorporating refugees and migrants. According to the Convention Relating to the Status of Refugees, the definition of a refugee is a person who,

> owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it. (U. N., 1951, vol. 189, p. 137)

The term migrant can be defined as “any person who lives temporarily or permanently in a country where he or she was not born and has acquired some significant social ties to this country” (“Migrant/Migration,” n.d.). Migrants can be people coming to a new country for work, education or reunion with family.

However, I decided to investigate and co-design with refugees and migrants in Finland rather than Germany for the following reasons:

a) the geographic restriction significantly prevents the Helsinki-based design team from reaching the end-users (i.e., refugees and migrants) in Germany;
b) lack of resources hardly leads to no possibility of conducting the field studies and co-design sessions in Germany;
c) Despite Finland varies from Germany politically, culturally, and economically, refugees and migrants share common demands of integration in economic, health, educational and social contexts (Martínez-Solimán, 2016).

3.1.3 Questionnaire

To gain “explicit knowledge” of user experience (Sanders, 2003), I conducted an online questionnaire regarding migrants’ problems to successful information seeking. Those participants were people who came to Finland for education, work, and migration. The questionnaire explored the following key areas:

a) the manner in which the participants sought information to support their integration into society;
b) the participants’ emotional and psychological status throughout the whole process of their assimilation into society;
c) participants’ expectations towards the chatbot to be designed in the ERICS project.

Besides, relevant literature on refugees’ mental health issues was investigated to support this session.

Participants

The recruitment occurred in May 2018. The participants were approached by utilizing two mediums. First, the invitation was sent as a personal email containing detailed information about the ERICS project and the goal of the questionnaire to 30 migrants who came to Finland for work and migration. Second, the invitation was posted to the Aalto International group on Facebook. The questionnaire targeted migrants with a maximum of three years of living experience in Finland. The launch page for the questionnaire is included in Appendix 2: Chatbot Design Survey for the ERICS project 2018. In total, 29 participants (age 20-49) from 15 different countries outside Finland (nine of those counties were European) took part in the questionnaire (see Figure 16). The overall sample consisted of 16 males (55.2%) and 13 females (44.8%). Thirty-one percent of participants had been living in Finland for less than a year, and almost half of respondents (44.8%) reported that they had been residing in Finland for more than one year but less than two years. The remaining 24.1% of participants had lived more than two-years in Finland.

Data analysis

Table 2 displays the information channels by which participants sought help to solve problems in everyday life. To obtain supports, most of the respondents approached their friends who had experience regarding the same issue or relevant knowledge. For example, Participant 2 (P2) explained that it was "natural to seek help from friends as a human being . . . Friends always show caring and love, which gives me warmth and belongingness in a new environment". More than sixty percent of the informants posted questions in online migrant communities such as Facebook groups, Wechat groups, and online forums, and nearly half of the informants decided to seek information independently by web retrieval.

Table 2. Approaches by which participants seek information about everyday life

	%
Friends	72.4
Community on social media, such as Facebook group	65.5
Web retrieval	55.2

Table 3 displays the factors obstructing the successful gathering of information. Searching the right information through the right channels was reported as the biggest problem by most of the participants. The language barrier was seen as another major obstacle preventing them from gaining information. For instance, P5 reported that "especially with work, most of the information is in Finnish, which is understandable but inconvenient." Additionally, there were some less common negative factors, such as the inefficiency of finding information, the invalidity of the information source, and slow response time when requesting information. For example, participants complained that "with too much information, it is quite slow and overwhelming to dig out desirable information online" (P9), and what is worse is that "some information is incorrect" (P12). Although friends were considered a good source of information, occasionally it took "a while to get the required information" (P1).

Table 3. Factors preventing gaining information on everyday life smoothly

	%
The unknown of information channels	62.0
Language barriers	58.6
Inefficiency of digging out information	27.5
Invalidity of the information	13.8
Latency of getting information	6.9

On the other hand, 86.2% claimed a social network (e.g., support from friends or migrant communities) was helpful to acquire information (see Table 4). Almost half of the respondents considered language-aid tools and services as an enabler for information seeking. One-fifth of participants thought that consulting through public services (24.1%) was also beneficial for finding information on everyday life.

Table 4. Factors facilitating information seeking

	%
Social network (e.g. friends & community)	86.2
Language support (e.g. Google Translate)	51.7
Public service (e.g. customer service in Tax office)	24.1

In terms of psychological status, most of the participants reported that they suffered from mental health problems due to social exclusion, vocational setbacks, and family disconnection. 6.9% of respondents reflected they had faced discrimination and prejudice directly or indirectly. As a result, feelings of isolation and helplessness can be generated and then impair their psychological wellbeing. Isolation and helplessness can be further aggravated if migrants lack social integration skills, such as communication and employment skills. For example, P5 mentioned that her inability to speaking Finnish prevented her from “engaging with the neighborhood” and from “obtaining lots of decent jobs.” Moreover, 41.4% suffered from being separated from their family both physically and psychologically. Also, discomfort stemming from the cultural difference between their original country and host country is another negative experience. This cultural difference can prevent the receiving society

from successfully accepting migrants in a variety of aspects such as employment, social status, and integration (Kirmayer, Narasiah, Munoz, Rashid, Ryder, Guzder, ... & Pottie, 2011). During the process of social integration, all these conflicting feelings and perceptions can lead to depression and anxiety disorders (Pumariega, Rothe, & Pumariega, 2005).

Given the diverse languages, culture, and stress factors associated with migration and resettlement, it is challenging to identify general feelings and thoughts from different migrants and refugees. However, there are some universal or similar emotional states. As noted by Tribe (2002), in post-migration, refugees suffer from psychological and practical problems, such as racism, stereotyping by the host community, unknown cultural traditions, family loss or separation (Tribe, 2002). Those issues of refugees were consistent with the problems reflected by the participants (i.e., migrants).

The questionnaire also explored the expectations of the chatbot to be designed. In terms of personality traits (see Table 5), the majority of the respondents preferred a chatbot being friendly and efficient. Meanwhile, more than seventy percent of them reported being energetic and reliable is an ideal characteristic in a chatbot. Sixty-two percent hoped that the chatbot makes them feel reassured, and more than half of the informants expected the optimism exhibited by the chatbot. Some participants suggested patience, humor, and politeness can be necessary characteristics in the chatbot.

Table 5. Personality traits of the expected chatbot

	%
Friendly	86.2
Efficient	82.8
Energetic	79.3
Reliable	75.9
Reassuring	62.0
Optimistic	55.2
Patient	41.3
Funny	31.0
Polite	17.2

Table 6 displays the top three possible embodiments of the chatbot: human (31%), animals (27.6%), and visualized artificial intelligence (17.2%). For example, P1 explained that it felt "natural to talk with a human . . . in chatbots, users should also have a natural conversation. If the chatbot is embodied as a 'human,' then the preconceived impression is naturalness already before using the chatbot". However, some other participants held different views that "'virtual' people may be too dystopian and even scary" (P23). Users know it is not "a human on the other side." Using a human photo somehow indicated "the chatbot may have equivalent human intelligence, which can give the wrong expectation of the performance of the chatbot" (P7). In addition, utilizing a human image may be "unfriendly for the refugees to some extent" because "refugees would not have to flee to a new country if they were not persecuted by some 'bad' people" (P14). Instead, "a characteristic animal avatar" can be more "interesting and friendly" (P11). An image such as an icon for artificial intelligence can be an option, too.

Table 6. Embodiment of the expected avatar

	%
Human	31.0
Animal	27.6
Visualized artificial intelligence	17.2
Alien	13.8
Robot	6.9
Item	3.5

As for the gender of the chatbot, a male persona was expected by the biggest number of participants, followed by a female persona. Nearly one-fifth of the respondents were fond of androgyny (see Table 7). Table 8 showcases the degree of maturity of the expected chatbot. Sixty-nine percent agreed with designing an adult-like chatbot while only 3.4% chose a chatbot acting like a teen. The middle age and the senior were respectively supported by the same number of participants (13.8%).

Table 7. Gender of the expected chatbot

	%
Male	41.4
Female	37.9
Androgynous	20.7

Table 8. The degree of maturity of the expected chatbot

	%
Teen	3.4
Adult	69.0
Middle age	13.8
Senior	13.8

Findings

By analyzing the participants' troubles, I again confirmed that a chatbot is an appropriate alternative to address problems preventing refugees and migrants from succeeding during the process of requesting information on everyday life. The professional editorial team behind the chatbot database can guarantee the reliability and trustworthiness of information for the newcomers. Through the chatbot, refugees and migrants are able to receive answers in a timely way efficiently.

It is evident that refugees and migrants' expectations toward the chatbot are closely in accord with their mental health issues. During the process of social integration, they are confronted with practical and psychological issues (e.g., communication difficulties, discrimination, prejudice, acculturation, and isolation). Hence, they reflect a mental model in which they regard 'friends' and 'communities' as primary and reassuring supporters.

The emotional bonding between the chatbot and refugees and migrants can be conducive to enhancing the user experience. More specifically, the amity, inclusiveness, and enthusiasm reflected by the embodiment and personality of the chatbot can help to relieve refugees and migrants' depression and anxiety disorders. The participants expect the chatbot to have certain traits, such as friendliness, efficiency, energy, reliability, the ability to reassure and optimism. They also recommend the chatbot be depicted as an adult with visualization options such as human, animal, or concretized artificial intelligence.

3.1.4 Expert Interview & Cultural Probe

With the support of cultural probes, I conducted two expert interviews with office-bearers from International House Helsinki who supported refugees and migrants in aspects of their daily lives, such as registering the Population Information System and managing taxes. Through expert interviews, I gained facts about how the office-bearers perceived refugees and migrants and how they communicated with refugees and migrants coming to seek help. To obtain a holistic picture of the conversational process between the agents of public services and refugees and migrants and to inspire the design, I also applied cultural probes (Gaver, Dunne, & Pacenti, 1999) in the interviews. According to Lucero et al. (2004), a familiar and natural environment where the participants doing the probe assignment allows them to feel at ease and relaxed. The actual probe was implemented in their offices. This flexibility can benefit data collection. However, due to the limited time of the participants contributing to the study, I decided to reduce the workload of the cultural probes and integrate it into the interviews.

Participants

Two office-bearers from International House Helsinki were contacted for this study (see Figure 11). One was a male representative (age 39) of tax administrations with ten-year experience of managing taxation. His responsibility was to help customers to understand the tax system and collect information on employment to establish a database. The other one was a female customer service agent (age 42) from the local register office with more than six years of working experience. She was responsible for helping refugees and migrants to register in the Finnish Population Information System.

Procedure

The interviews were conducted face-to-face at the participants' premises (i.e., the office at International House Helsinki) in May 2018. The sessions were planned for an hour. In the first ten minutes, the interviewer (the author) explained the purpose of the session, including the focus of the project and how the current sessions fit within the big picture of co-designing a chatbot.

Besides, the interviewees agreed to sign the consent form after being explained the confidentiality policy and how the data gathered would be used. The actual interview started after the 10-minute introduction and lasted 30 minutes. The interviews were semi-structured, which were able to reduce some areas or topics but also discover unexpected stories. The interviews explored three specific aspects:

a) problems preventing refugees and migrants from smoothly handling bureaucratic affairs;
b) frustrations that public service agents face when answering refugees and migrants' questions;
c) reflections on the conversation between public service agents and refugees and migrants.

In the last 20 minutes of the interviews, a cultural probe – the exploratory card of conversation flows – was given to the interviewees to reflect on a general process of dialogues between them and their customers (i.e., refugees and migrants) (see Figure 12). The task was designed as ambiguously as possible by answering the question: what is the conversational flow between you and your customers. The participants were reminded to reflect before, during and after the conversation. All the sessions were audio-recorded. Pictures were also made during the session to capture specific aspects of the probes that the participants were creating.

Figure 11. Interviews with an office-bearer from International House Helsinki

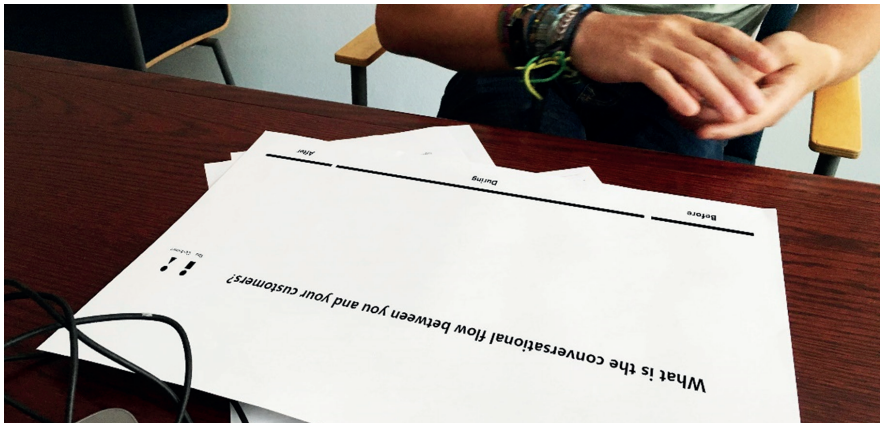


Figure 12. An interviewee is using a cultural probe (an exploratory card of conversation flows) to reflect on a process of information service for refugees and migrants

Data Analysis

After conducting all the interviews, the interviewer (the author) made transcripts (including photographs) and an initial categorization of the data within 24 hours after the interview. The analysis team consisted of the interviewer (the author) and one developer who also worked as a research assistant in the ERICS project. Through the latent qualitative content analysis (Graneheim & Lundman 2004), the transcripts of interviews were investigated. Firstly, the transcripts were read to acquire an overall picture of content regarding the goal of this study. Next, the transcripts were scanned again to achieve immersion while jotting down the reflections. After that, each interview text was divided into meaning units, of which condensed meaning units were extracted out. Eventually, those condensed meaning units that have similar or relevant meaning and content were merged and abstracted into subthemes. These subthemes were then connected with an overall theme (see Table 9 for the process and the result). As for the cultural probes, the data from the exploratory cards was transcribed, and comments were number and color-coded for each participant. The illustrations on the cards were developed and scanned, as they provided information on the written content of the probes. General findings were formulated based on the interpretation of the probes.

Table 9. Content analysis of transcribed data from interviews. Examples of meaning units, condensed meaning units, subthemes and theme

Meaning unit	Condensed meaning unit	Subtheme	Theme
"They [refugees and migrants] do not speak very well English, Swedish or Finnish. Usually, they come without any translators. Also, they find all the application very hard to fill because of lots of information and steps."	Incapacity of speaking English, Swedish or Finnish preventing the success of business handling	Language barriers preventing smooth communication and decreasing working efficiency	The ways of transmitting information affect the efficiency and effectiveness of social integration services
"The language problem between myself and the customers makes things slow and unproductive. There were some difficult situations where we had to turn the customer away and told them to return with someone who can translate for them."	The language gap decreasing working efficiency		
"When making applications, sometimes they [refugees and migrants] would call someone, but the conversation would go through the phone, and it might not be good as you don't know how would the middle person bring the information you gave."	Uncertainty about the validity of information delivered by Intermediators	Indirect ways of information delivery reduce the effectiveness	
"They [refugees and migrants] did not get their registrations done in time, because they either have not brought the necessary paperwork or haven't visited the magistrate, which is mandatory for all."	Registration is beaten back by incomplete material	Unknown of information results in the failure of public service	
"Returning Finnish citizens get impatient with the service because they do not understand what has to be done and assume that things are already done when they are not."	Encumbered services caused by the unknown of the situation lead to negative emotion and feeling	Impeditive business handling caused by the unknown of information generates negative emotions between interlocutors	The efficiency of transmitting information influences interlocutors' emotion
"I get annoyed when the customer is annoyed. This is often a result of missing paperwork. Often the customers understand why everything is the way it is once explained."	Reciprocally negative emotional influence between interlocutors		
"Sometimes they [refugees and migrants] would call someone as the middleman, and the conversation will go through the phone. It is not so good because you don't know what this person will bring the information you gave. I feel frustrated because my information does not go to the customer clearly."	Get frustrated by the middleman because of the uncertainty of accuracy of information delivered	Indirect information delivery risking negative emotion	

Table 9 (continued).

Meaning unit	Condensed meaning unit	Subtheme	Theme
"Customers asked a lot of questions that are related to the area which is not my expertise, and I could not answer. I tried to figure out. Mainly I googled it, or I tried to appoint the person to the right directions."	Limited knowledge failing to answer all the questions	Call for a comprehensive knowledge base to cover the topics about social integration	The necessity of information retrieval supports in digital services
"The Magistrate's website is not very user-friendly. To find something, you need to know exactly what you are looking for."	Inefficient information seeking on the Magistrate's without explicit questions or keywords indicating an intention	The need for constructively forming relevant topics or keywords in information retrieval	
"Sometimes half of what I have said is forgotten, and I have to explain it again. There is no way to return to the conversation that we had."	Repetitive clarification because of no records of conversational history	The conversational history efficiently supports information seeking	

Findings

The content analysis revealed three main themes and eight associated subthemes (see Table 9). The participants gave concrete examples of their experiences of interacting with refugees and migrants in customer services, and they recalled and visualized the conversational flow with refugees and migrants.

According to their feedback, I discover the means of transmitting information greatly affects the efficiency and effectiveness of customer services of social integration, such as languages, direct or indirect communication, and information source. Inefficient information seeking and dialogue can lead to negative emotions of refugees and migrants (e.g., frustration and impatience). Therefore, refugees and migrants need information retrieval supports such as suggestions of questions or keywords accurately clarifying their intentions and retrospective conversational history.

Through the cultural probe, the participants reflected on a normal dialogue process where refugees or migrants came to request information services. Figure 13 illustrates a conversation flow of refugees or migrants consulting on taxation. The dialogue starts with a greeting and ends with a blessing. Before processing the request, the tax service agent clarifies the refugee or migrant's intention by asking questions. When service representative is dealing with the request, the refugee or migrant occasionally looks at him to seek signs indicating they are still involved in the service. This process reveals the refugees and migrants' mental models that they need immediate feedback to immerse themselves in a seamless conversational experience.

Extension of the User Group. In addition to refugees and migrants remaining the most prominent user group, workers at bureaucracies holding accountable for migration and social integration can be seen as users of the chatbot-service. The interviewees mentioned refugees and migrant sometimes would ask questions which were out of their expertise or to which they were uncertain about the answer, they had no other good information source, but the web retrieval:

> *"Customers asked a lot of questions which did not belong to my area of expertise, and I could not answer. I tried to figure out. Mainly I googled it, and then I tried to appoint the person to the right directions. But the whole process was quite time-consuming." (Interviewee 1)*

Simple Languages in Chatbot Services. According to the interviewees, the most prominent problem was the language barrier. Most refugees and migrants were incapable of speaking English very well, and they had limited lexicons. Interviewee 2 reported that "simple words are encouraged when communicating with refugees and migrants." Thereby, I suggest the chatbot to use accessible and understandable wording when dialoguing with users.

Examples of Questions by Chatbots. Interviewee 2 pointed out that sometimes refugees and migrants had troubles with expressing their intention precisely. For example, it may take a while for them to search for specific information on the website if they do not know how to form the questions correctly. Hence, relevant question examples and quick question suggestions based on the user's input can enhance the efficiency of task performance and user satisfaction:

> *"Sometimes I needed to ask more than two questions to figure out the real intention of the customer. " (Interviewee 2)*

Conversational History in Chatbots. As stated by interviewee 2, the customers – refugees and migrants – could repeatedly ask for preceding information, as they could not remember such a large amount of given information along with the conversation going. To improve the convenience of requesting information and the accessibility of information in chatbots, I propose a function of saving the chat history if users prefer:

> *"Sometimes they forgot half of what I have said. It is understandable that lots of information could not be easily fixed in their memory, and there is no way to return to the conversation that we had." (Interviewee 2)*

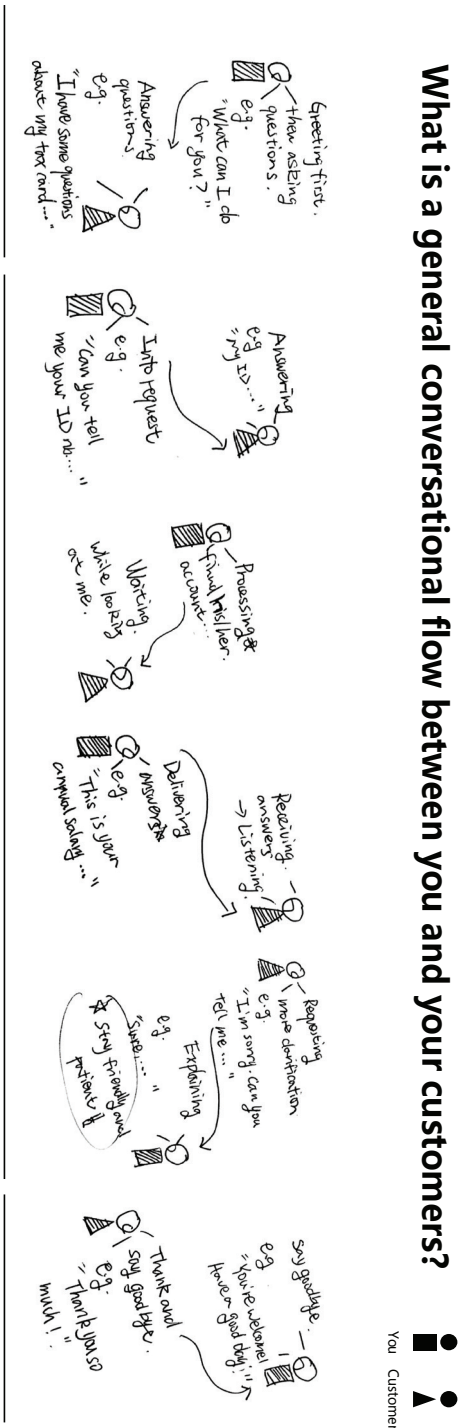


Figure 13. The result of the cultural probe task: what is a general conversational flow between you and your customers

3.1.5 Participant Observations & Empathy Probes

The purpose of observation is to explore user experience in using chatbot – understanding user behaviors and mental models. I employed participant observation to investigate how refugees and migrants interact with a very early chatbot demo created by developers to experiment with the QA system in the ERICS project. This current pre-design chatbot prototype was seen as an opportunity to study the human-chatbot interaction (see figure 14). Participant observation allows a researcher to gain explicit and tacit knowledge of user interaction by taking part in social events (Musante & DeWalt, 2010). During the observation, interaction with participants was avoided entirely, and I merely acted as a 'bystander.' Through the participant observation, I identified how the users initiated a conversation and how they reacted to the chatbot's prompts and task performances. Also, I produced written accounts and descriptions – field notes – to represent participants' interaction with the chatbot. Not only did I record all the details and actions of the users by video recording but I also documented the events and objects significant to my thoughts and reflections.

Following the participant observation, I used empathy probes (Mattelmäki, 2005) in which those participants received probe artifacts and assignments to help them reflect on how they felt about the chatbot demo. Personal interviews with the informants were then conducted based on their reflections. Referring to other probe studies (Mattelmäki & Battarbee, 2002), the probe kit incorporated illustrated cards with open questions and a sheet of stickers (see Figure 15). Illustrated cards can be a capable medium to ask these sorts of questions, as they are informal and friendly modes of communication (Mattelmäki & Battarbee, 2002). The cards contained images demonstrating a particular interface of the pre-design chatbot and questions:

a) How do you like the chatbot avatar?
b) How would you feel about the pre-design chatbot's way of talking?
c) How much do you like the chatbot interface?
d) How much are you engaged in the conversation opener by the pre-design

chatbot?

e) How do you feel about each particular message?

f) How much are you motivated to rate the chatbot?

g) How much are you motivated to improve the answer?

h) How do you feel about the conversational closure?

The questions concerned participants' attitudes towards the pre-design chatbot's avatar, tone, conversational flow, and graphic user interface. Oblique wording and evocative images were used to open a space of possibilities, allowing participants as much room to respond as possible. To facilitate the participants communicating their feelings, I provided a sheet of stickers with cartoon characters and faces presenting different emotions and ideas.

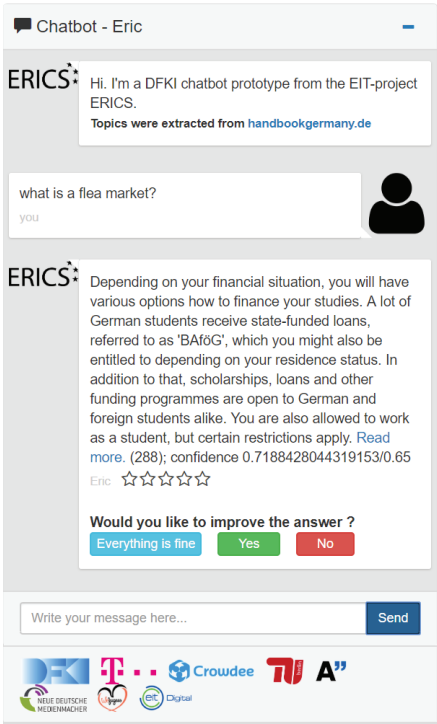


Figure 14. The chat interface of the pre-design chatbot prototype in the ERICS project

Figure 15. Empathy probes: the illustrated cards with open questions and stickers finished by the participants in the ERICS project

Participants

One refugee and two migrants who are skilled in using web services from Arffman Consulting were recruited for this study. They have been living in Finland for less half of a year. The participants varied in their education (colleges, vocational-technical schools), background (cleaner, waiter, developer), age (between 25 and 29), and gender (one male, two female).

Procedure

The sessions were organized at the participants' premises (i.e., Arffman Consulting) in June 2018. The sessions were planned for 90 minutes, and the whole process was recorded via video. In the first 15 minutes, I explained the goal of the session, including the background of the ERICS project, the competency, and limitation of the pre-design prototype, and how the current sessions benefit co-designing a chatbot. The actual observation started after the introduction and lasted 25 minutes (see figure 16). In the next 20 minutes, the participants were given the probes incorporating eight illustrated cards with an open question and a sheet of stickers. In the last 30 minutes, each participant attended a personal interview, in which the result of the probes was discussed. The interviews were intensive and focused on the reflection of the created illustrated cards representing the participants' experiences of using the prototype. Before the interviews, the probes were collected and reviewed. The highlighted area of interest, doubts and relevant critical themes on the cards functioned as a starting point for more in-depth and reflective discussion.

Figure 16. One migrant was experiencing the very early chatbot demo in the ERICS project

Data analysis

The data was analyzed by using Affinity diagram. Affinity diagramming is a technique utilized to organize, analyze, and make sense of large amounts of rough, unstructured, and seemingly different qualitative data (Hartson & Pyla, 2012). According to Lucero (2015), an affinity diagramming process consists of four stages: creating notes, clustering notes, walking the wall, and documentation.

In the first stage, interpretation sessions occurred within 48 hours after the participant observation and empathy probes. The interpretation team consisted of the author of this thesis and a research assistant in the ERICS project. Both interpretation team members created notes utilizing sticky notes while reviewing the corresponding video from the session and the resulting artifacts produced by the empathy probe. Each member created between 60 and 100 affinity notes per participant. Each note was first color-coded to identify the participant by using a different color. In the second stage, we assigned meaning to the data collected through interpretation. The sticky notes were used to analyze several rounds of interpretation. The first part of the analysis consisted of creating a framework diagram. The goal was to identify the different steps in the process of question answering by chatbots, including phases before, during, and after the conversation. We scanned the notes and produced a common table for each participant, individually connecting the contents of the notes to each of the formerly mentioned phases. The table was then discussed, disclosing the differences and similarities between analysts in the interpretation, and resulting in more general findings. In the third state, we conducted two rounds of discussions over the course of two days to build the affinity diagram (see Figure 17). We read the notes individually and started grouping notes, creating clusters slowly, which later led to categories. A shared understanding of issues regarding the interaction between the users and the chatbot was built, which formed categories that were naturally revealed. In the end, these categories were processed into more general findings.

Figure 17. Analyzing data gained from the participant observation and the empathy probe through affinity diagramming

Findings

Through the participant observation, I discover proactive instruction and guide from the chatbot can be beneficial for chatbot navigation, thereby increasing the conversational experience. Smooth navigation can be achieved by providing a clear interpretation of the chatbot capability, explicit guidance in rating messages, and a useful follow-up to the failure of answering questions. The participants are dissatisfied with the welcome message, “Hi. I’m a DFKI chatbot prototype from the EIT-project ERICS”, as they were perplexed about the chatbot’s intention and capability. Instead of asking questions directly, they then start the dialogue by saying ‘hi.’ They are expecting the chatbot to provide essential tips and instructions for optimal use. For instance, one participant murmured to himself, “I don’t know what question you (the chatbot) can answer.” The participants show eager desires in the chatbot offering damage controls (e.g., other approaches to obtain the answers) after the chatbot fails to respond to the request and apologizes. One participant even makes an extremely negative decision, “if the bot cannot answer my question or take actions to solve that problem, this will be my last time using it.” The participants rate the message, only when their questions are correctly or awkwardly answered. Two participants do not notice the five stars rating, which demonstrates the necessity of specific clarification and guide in rating messages.

Asking users to give concrete suggestions to refine the answers can be a burdensome task. Unfortunately, the participants consider it is unreasonable to be requested to improve the answer. For example, a participant reported, “if I know the right answer to the question that I am going to ask, then why should I come to you (the chatbot)?”.

Besides, a chatbot can improve the quality of interaction and user satisfaction by applying social etiquette in real life (e.g., greetings and farewells). It is worth noting that all the participants terminate the conversation by directly closing the chat window, which is not in line with the human-to-human interaction pattern of ending the conversation with clear signs described in the previous expert interviews (see section 3.1.4 Expert Interview & Cultural Probe).

The empathy probe reveals that emotional engagement through the chatbot avatar and the chat interface is seen as a critical factor for creating the great user experience in chatbot services. Participants select stickers of sad emojis and robots to express their impressions of that chatbot prototype. The monogrammed icon of the chatbot avatar cannot let themselves immerse in the human-like conversation. Instead, they expected a visual embodiment such as “a kind creature or even a cute robot.” Moreover, participants reported that they perceive the prototype as “a software application” rather than a chatbot with a particular personality and a vivid voice. Compared to the “cold and forum-like” interface, a “warmer and colorful” chatbot interface can be more compelling. These dehumanized determinants (e.g., monogrammed icons and machine-like personality and voice) psychologically prevent them from engaging in the ‘conversation.’

Furthermore, a short, concise, and explicit message can enhance the efficiency of information transmission and the chatbot training tasks required from the users (e.g., rating the answer). The participants complain that the information in “each message sent by the chatbot was too chaotic and overwhelming,” which gives rise to the hardship of focusing on the main point. In combination with their low motivations for training the bot, participants usually refuse to rate the bot, unless they are extremely satisfied or

dissatisfied with the given answer. Participants choose confusing emojis when the conversation starts because there are no instructions and guides for using the chatbot. This demand for instructions and guides of usage is consistent with what I found and conjectured in the participant observation. One participant explains the reason why the chatbot should clarify the capabilities of the chatbot in the beginning by making a metaphor. She considered the user as a ‘student’ with questions in a specific field and the chatbot as a ‘tutor’ with an unknown professional background, “How can a student get his or her questions answered correctly if he or she does not know the tutor’s areas of expertise?”.

As for the behavior that the users close the conversation without any sign of farewell, one of the participants explains, “it is unnecessary to say goodbye to the bot when the conversation is over if the bot is not ‘smart.’” She, however, agrees that the chatbot proactively giving a warm farewell can enhance the satisfaction of the service and the exit experience.

3.1.6 Problems & Design questions

Through the session of say (questionnaire and culture probe) and do (participant observation and empathy probe), I discovered refugees and migrants' demands of engaging with the chatbot psychologically and emotionally. They suffer many mental issues such as the difficulty of acculturation and isolation throughout the process of social integration. An empathic human-like chatbot avatar promoting natural conversations can be conducive to user experience and user engagement. It has been shown that human-like chatbots suiting to their domain contribute to user trust for the content of the website, and thereby increase the retention of users (Jain, Kumar, Kota, & Patel, 2018; Nan, Anghelcev, Myers, Sar, & Faber, 2006; Seeger, Pfeiffer, & Heinzl, 2017). Psychologically speaking, anthropomorphic agents are in line with fundamental human needs to be socially related to other humans (Seeger, Pfeiffer, & Heinzl, 2017). Besides, a study by Reeves and Nass claims that humans treat computers as social entities with different attitudes and social rules. Computer-users regard some computers as 'experts' and others as 'generalists,' and they interact with the computer in a human-to-human way (Reeves & Nass, 1996). All the above leads to the first design question:

What constitutes an anthropomorphic chatbot avatar that is capable of motivating engagement with refugees and migrants

The users decided to abandon the pre-design chatbot prototype because it acts 'just like a bot,' i.e., the templates it employs in the dialogue are rigid and unnatural. It prevents users from engaging in human-like conversations. For the purpose of successfully completing the task, the essential instructions and guides of the chatbot are expected by the users. During conversations, the ideal path is that the users ask questions which are correctly parsed and answered by the dialogue. However, in a free-form messaging interface, some conversational flows inevitably are not modeled, which can give rise to a dialog failure. The system, for instance, is either uncertain about how accurately it understands the query or it is incapable of responding to its database (Jain, Kumar, Kota, & Patel, 2018; McTear, Callejas, & Griol, 2016).

For enhancing user satisfaction, to address all the dialogue errors or dead-ends with appropriate responses is an essential consideration in chatbot design. Based on this, the second design question is defined:

What represents a natural conversation flow between the user and the chatbot, which enhances the efficiency of question answering

Whether or not chatbots appear intelligent is contingent on the quality of information they have access to (Cahn, 2017). Consequently, it is imperative to invite users to confirm the aptness and usefulness of answers provided through the QA system. In the research of say and do, however, users present low motivation to provide feedback. In addition, there is a strong possibility that the question captured by the QA system may be uncertain or ambiguous because of the technological limitations. As a result, the system typically generates a list of ranked responses or answers with low confidence scores. In such cases, collecting training data on the answers preferred by users is critical for improving the bot's accuracy in real-time, as the training data improves the quality of the corpora and complements algorithm development. This leads to the last design question:

How to motivate users training the chatbot while engaging in a natural conversation

3.2 Ideate

3.2.1 Chatbot Avatar Creation - Co-design Workshop I

I set up the first co-design workshop to explore and create possible solutions for the first design question: what constitutes an anthropomorphic chatbot avatar that is capable of motivating engagement with refugees and migrants. A chatbot requires, in addition to robust artificial intelligence, personalities to build an impression that the chatbot exists as an 'intelligent being' and to become convincing in the user's eyes. Therefore, it is critical to include several psychological characteristics in a chatbot, such as "personality traits, biographical facts, and expressed emotions" (Kuligowska, 2015, p.10). This workshop planned to devise chatbot avatars that comprised both visual embodiment and personality.

Participants

Initially, I intended to invite the same three participants from the empathy probes study to participate in the workshops. For different reasons, none of them could be available for the co-design activity. In that case, I invited four new participants (two migrants and two in-house developers) who agreed to take part in the workshop. In the workshop, I served as a facilitator and a design contributor. The workshop participants varied in their nationality (India, Italy, China, Finland), age (between 24 and 31), and gender (1 female, 4 male). All the three migrants including the author had less than two years of residence in Finland, and the two developers were originally from Finland.

Procedure

The co-design workshop was conducted in June 2018 at Aalto University. The sessions were planned for a total of two hours. The sessions consisted of four parts:

1) Introduction (10 minutes)
2) Brainstorming personality qualities for the chatbot (40 minutes)
3) Visualizing the chatbot avatar (30 minutes)
4) Sharing and discussing outcomes (20 minutes)

Introduction. To create a comfortable and relaxed atmosphere, participants first introduced to each other. After that, they familiarized themselves with the background of the ERICS project and the objective of the workshop: how we might design an anthropomorphic avatar that emotionally connects with refugees and migrants in Germany. In this context, it was crucial to clarify that the avatar should be created on the basis not only of the participants' experience of social integration in Finland but also of their perceptions of living in Germany.

Brainstorming personality qualities for the chatbot. The participants were required to brainstorm what personality qualities they would like the chatbot to exhibit when talking to it. The participants first produced a list of adjectives of personality traits (see Figure 18). During this session, data about the expected personality traits and visual embodiment of the chatbot from the previous questionnaire (see section 3.1.3 Questionnaire) was fed in the form of design artifacts (i.e., questionnaire findings in the form of cards). This action aimed to provide the participants with further inspiration and a firmer basis for their opinions. Next, they voted and discussed to reduce the list to the 4–6 most essential personality characteristics for promoting engagement between the chatbot and the user.

Visualizing the chatbot avatar. The participants visualized an avatar manifesting the selected traits. The visualization was followed by a story-making where participants created a narrative for the visualized avatar. The stories described the chatbot's characteristics, including background, appearance, age, gender, interests, a tone of voice, manners of interaction, as well as personality traits.

Sharing and discussing outcomes. Lastly, each participant gave a short presentation of his or her created avatar and story, and then they discussed and voted for the top three avatar personas.

Figure 18. Co-design participants are brainstorming personality traits of the chatbot in the ERICS project in co-design workshop 1

Generative Toolkits

A generative toolkit was applied by the participants to create 'artifacts' expressing their thoughts, feelings, and ideas both visually and verbally. The toolkit contained a board on which a large number of visual and tangible components could be arranged and juxtaposed in different ways. According to the guideline of persona creation by Actions on Google ("Create a Persona," n.d.), the board was designed with consisting of four parts or tasks (see Figure 19):

1) "brainstorm as many personality traits that you expect the chatbot to possess as possible,"
2) "discuss and piece six personality traits together to form a whole chatbot personality,"
3) "visualize the avatar based on the created chatbot personality, " and
4) "make up a story for the avatar."

This board functioned as a connection amongst all the sessions in the workshop. In turn, the components covered a variety of tools and representational materials such as pens, markers, scissors, colored sticky notes, printed cards with words and phrases describing personality, and inspiring images of different creatures. Those images comprised diverse ethnic groups, living creatures, and forms of representation (e.g., 2D, 3D, and cartoon) enabling the participants to become inspired or build associations (see Figure 20). As for the visualization of the chatbot avatar, the participants

could directly use the images, cut and reassemble pieces of images, or draw their own avatars.

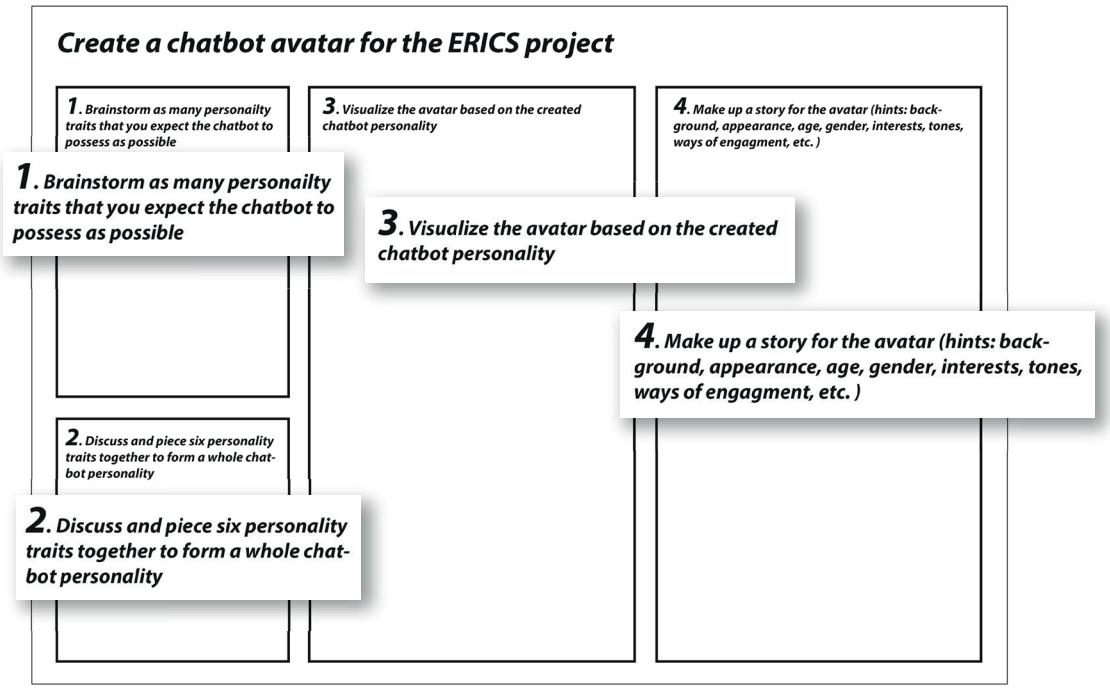


Figure 19. The board for creating a chatbot avatar in the ERICS project

Figure 20. Co-design participants are creating the chatbot avatar through the generative toolkit in co-design workshop 1

Findings

Personality traits. The participants jointly define six critical characteristics for the chatbot: reliable, efficient, friendly, reassuring, energetic and optimistic. They claim that the chatbot, as an information portal with serious topics concerning social integration, is required to ensure reliability. The chatbot should respond to users' requests in an efficient fashion. Even though the chatbot may fail to answer the query, it should still provide error handling removing the users' dissatisfaction, i.e., other alternatives to obtain the answers. The participants see these features as fundamental requirements. Therefore, they expect the chatbot to exhibit reliable and efficient traits through visual embodiment or task performance. In order to obtain a delightful experience, they wish the chatbot to be friendly and reassuring. According to their explanations, friendliness is not only a critical quality in regular customer services but also the desired factor enabling them to psychologically feel safe as a newcomer in a new host country. Moreover, the participants report the process of social integration can be extremely demanding, which unavoidably causes frustration and agitation. In that case, if the chatbot can reassure them in expressions and tones of voice, it will be helpful to enhance their conversational experience. By answering a series of questions, the participants reveal their intent of utilizing the chatbot service. They desire to avoid being exhausted physically and psychologically in the information-seeking process and then save more energy to manage social integration successfully. Also, the chatbot is expected to eventually improve their optimism for leading a new life in the receiving county. Hence, being energetic and optimistic are the qualities that they would like to see from the chatbot as well.

It is noteworthy that the structure of the expected experience of the chatbot is in line with what Hassenzahl, Diefenbach, and Göritz (2010) claim: pragmatic quality as hygiene factors and hedonic quality as motivators for user experience. As stated by Hassenzahl et al. (2010), pragmatic quality refers to an assessment of a product' ability to achieve particular "do-goals" (e.g., to receive an answer from the chatbot) and is similar to a general understanding of usability. Hedonic quality is an assessment regarding a product's ability to

generate “pleasure in use and ownership,” which fulfills “be-goals” (e.g., to be stimulated) (Hassenzahl, Diefenbach, & Göritz, 2010, p. 5). In terms of user experience, pragmatic quality functions as hygiene factors, which removes barriers and dissatisfaction as well as dampen the negative effect. Hedonic quality performs as motivators driving users’ intention to use the product and generate positive affect and experience (Hassenzahl, Diefenbach, & Göritz, 2010). Figure 21 illustrates the hierarchy of the expected chatbot personalities in the framework of pleasure-driven pragmatic-and-hedonic user experience. Being reliable and efficient is seen as the chatbot’s pragmatic qualities preventing users’ dissatisfaction happening. On the other hand, being friendly, reassuring, energetic and optimistic is the hedonic aspect where users can be motivated to utilize the chatbot service and, thus, obtain a pleasurable and meaningful user experience. More specifically, the chatbot ultimately intends to create pleasure for the users through optimistic and energetic personality traits.

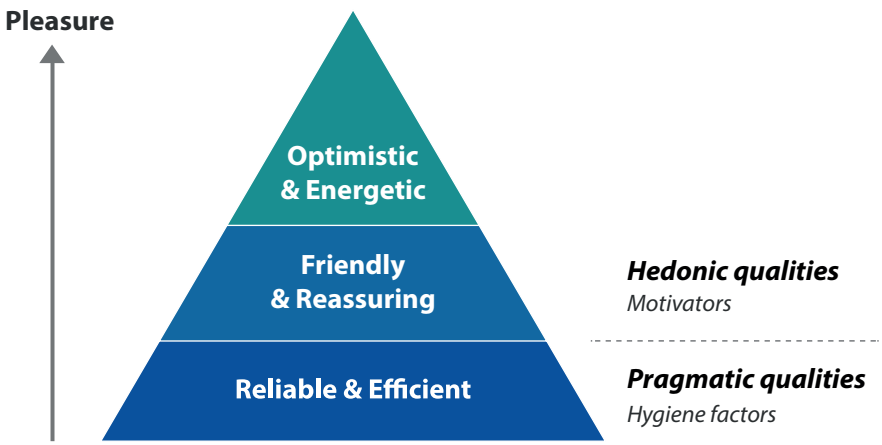


Figure 21. Hierarchy of expected chatbot personalities

Avatars. This workshop produced the final three proposals for the chatbot avatar. The decision was made by voting and considering the findings of the previous quantitative research about migrants’ expectations of the chatbot in the ERICS project.

The first candidate is a pigeon-like avatar, Eike (see Table 10). Eike is a gentle city-born messenger communicating information and delivering peace and love to refugees and migrants. Eike knows about living in Germany and is happy to share it with anyone coming with questions. In a soft and friendly voice, Eike aims to soothe worries and assure information seekers of a secure, peaceful and bright future. The voice of Eike is familiar and friendly for users. Eike communicates with a folksy tone that might be used with an old friend. Eike uses contractions like “don’t” instead of “do not” because that is how most real humans speak to one another in everyday conversations.

Table 10. The first candidate of the chatbot avatar

Example of avatar	Example of dialogue	Chatbot move	Message
	Hello Coo there! How can I assist you on this fine day? How can I learn German? Just a second, I know exactly what your’re looking for! You can find German language resources online from www.handbookgermany.de You can also try to enroll on German language courses at www.germancourses.de Thank you! No problem, I’m here to help, coo!	Greet	“Coo there! How can I help you on this great day?”
		Confirm	“Yay, I think I found exactly what you were looking for!”
		Clarify the request of rephrasing	“Sorry, I didn’t quite catch your question. Could you rephrase that?”
		Acknowledge	“Hmm, I couldn’t find anything about that. ’
		Acknowledge failure and follow up	“Our system is experiencing some problems. Our staff is already fixing it. Hang tight!”
		Clarify the request of rating	“Coo, please rate my messages to improve our service greatly!”

The second candidate, Niklas, is created based on a persona of human (see Table 11). Niklas is a German researcher of cultural science, whose passion is to find elegant solutions to cultural problems. Supporting refugees and migrants’ integration benefits Niklas’ research in Germany. As the leader of the research group, Niklas’ knowledge is extensive and precise. However, this is his first time leading a research group. As a result, there is still much room for him to learn. Friendliness is Niklas’ primary tool in creating a comfortable

and safe environment for newcomers. With passion and carefulness, Niklas is very enthusiastic about helping all who need information about everyday life. Niklas carefully pushes the user in the right direction and then waits for a reaction that can be recorded and hopefully replicated. According to Niklas, every interaction can be considered for user research, if it is conducted correctly.

Table 11. The second candidate of the chatbot avatar

Example of avatar	Example of dialogue	Chatbot move	Message
		Greet	"Hey there, friend! :)"
		Confirm	"Excellent, this will surely help us!"
		Clarify the request of rephrasing	"Oh my, I don't think that can be done. Let's try again differently!'
		Acknowledge	"Oops! I can't find anything like that. :("
		Acknowledge failure and follow up	"Uh oh, that experiment failed! We have to start over. Please re-open this window."
		Clarify the request of rating	"I think we did great today. How about you? How do you think I did?"

The final candidate is Erics who is a friendly and helpful technological being (see Table 13). Erics seeks to learn about its home, Germany. Erics always acts in a patient and friendly way. It keeps learning by teaching newcomers to integrate into society. Erics' voice is clearly synthetic, yet friendly and polite. Erics focuses on the task at hand, and it gives careful explanations to make sure that the user understands what to do next.

Table 12. The third candidate of the chatbot avatar

Example of avatar	Example of dialogue	Chatbot move	Message
		Greet	"Hello, how can I help you?"
		Confirm	"Great, we're making progress!"
		Clarify the request of rephrasing	"I'm sorry, I couldn't complete that request. Can you say that again differently?"
		Acknowledge	"Seems like we can't anything related to that, perhaps we should try something else."
		Acknowledge failure and follow up	"I'm sorry, something went wrong, and I have to restart. Please open this chat again!"
		Clarify the request of rating	"Thank you for letting me help you. Please tell me what you thought of me. It helps me learn!"

3.2.2 Feedback Collection – Identification the Final Chatbot Avatar

In order to achieve a convincible design decision where the chatbot avatar is able to engage with the broader masses of the users emotionally, the three avatar proposals generated in the first co-design workshop are assessed by refugees through a quick on-site survey (see Figure 22). The physical survey is distributed in the Finnish refugee reception center, Luona. The survey mainly explores which avatar candidate is most preferred by the refugees and why. In the validating questionnaire, the respondents can give free form feedback and comments on the avatar candidates. In total, 30 respondents (age 18-46) mainly from Afghanistan, Iraq, and Somalia participated in the survey. The overall sample contained 21 males (70%) and nine females (30%).

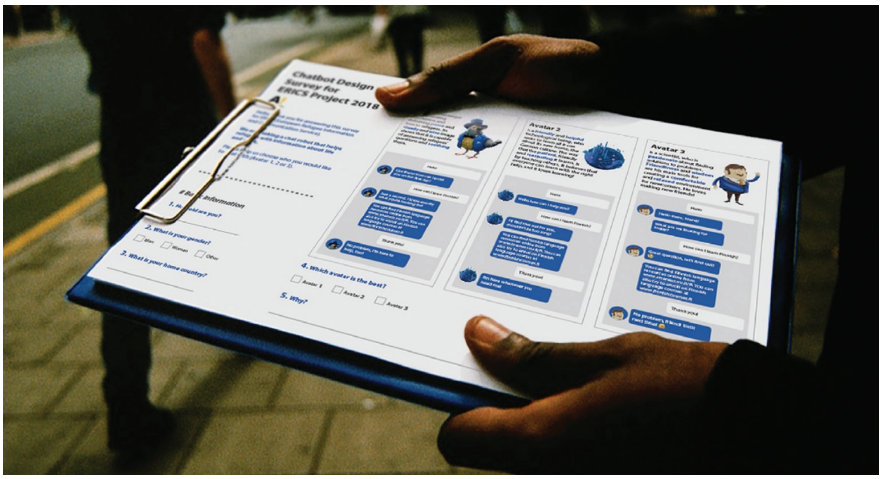


Figure 22. A quick on-site survey to identify the final chatbot avatar

Findings

Table 13 demonstrates the result of the survey. Most of the respondents vote for Eike, while a small number of the participants fancy Erics and Nicolas. According to the respondents, Eike is “more friendly,” “funnier,” and “cuter,” which can enhance their pleasure. Besides, a pigeon-like avatar looks more “peaceful” and “relaxing,” and psychologically, they can feel “safe” in the

conversations. The respondents report that both Eike and Niklas can evoke an emotional engagement, but Niklas' way of speaking is too "casual" and "loose," which does not appropriately suit the domain and the topics. Those respondents picked Erics mainly because it is "cool." In light of this result, I decided to create a chatbot avatar based on the persona of Eike.

Table 13. Avatars preferred by the respondents (refugees)

	%
Social network (e.g. friends & community)	86.2
Language support (e.g. Google Translate)	51.7
Public service (e.g. customer service in Tax office)	24.1

Final avatar design - Eike

Two requirements are taken into consideration when creating a chatbot persona. First, the avatar requires visual characteristics representing German culture. According to Pumariega et al. (2005), the cultural transition is recognized as an essential segment of social integration for refugees and migrants. To help refugees and migrants adapt to a new culture, classic anthropology and social sciences propose assimilation where the individual renounces his original culture and identifies with the culture of the receiving country. Consequently, the chatbot avatar, as a messenger delivering information in Germany, is necessary to reflect German culture. Second, the visual embodiment character should exhibit defined personality traits. The visualization of the avatar (e.g., the face and gestures) provides important secondary communication besides the explicit expression, such as words (Pelachaud, 2000). Thus, the avatar is expected to visually showcase its personality, which serves as a part of user experience creation.

Persona. Based on the integration of the outcome of the co-design workshop, the feedback from the survey, and the defined requirements of avatar creation, I created the final avatar image, Eike (Eike is a German name) (see Figure 23). Eike is a gentle city-born messenger pursuing peace in the neighborhood. Eike knows all there is to know about living in a German city and is happy to share his knowledge with anyone who comes to seek it. In a

soft and friendly voice, he aims to soothe worries and assure information seekers of a secure, peaceful and bright future. Eike always has a positive look, which makes refugees and migrants feel hopeful of living and working. Eike's open and stable posture gives him a scholarly and curious image, conveying that he is capable of providing answers to refugees and migrants' questions. Large and friendly eyes convey a feeling of welcomeness. His sharp clothing and stylish Tyrolean hat[14] give him a sophisticated but folksy look. The cartoon style lets people know that the information seeking experience could be cheerful and funny, and the same goes for life. The blue and yellow in Eike echo with the color scheme of European Flag but also the host website's palette (i.e., the HBG portal), which visually and psychologically indicates a harmonious feeling. Through his appearance, Eike is going to demonstrate the following brand traits:

1) Reliable but not patronizing
2) Efficient but not hasty
3) Friendly but not pushy
4) Reassuring but not sheltering
5) Optimistic but not dismissive
6) Energetic but not light-headed

Figure 23. The visual embodiment of Eike the chatbot

[14] The Tyrolean hat, also Bavarian hat is a type of headwear in what is now part of Germany.

Voice. Eike speaks in a welcoming, friendly and lively voice that users would expect from a host of a grand urban festival. Serving refugees and migrants with everything he has learned, Eike happily responds when he is confident of completing a request. His voice has urban vibes, which reflects Eike's background in a busy city life. Although Eike is used to a busy life, he wants to soothe and help people take it slow and immerse themselves in the experience of their daily lives. Eike's voice directs users to take actions with determination, but without pressing them too hard. Encouragement and reassurance are what Eike wants to help users to perceive. Eike is proud of his identity being a pigeon. With a cheerful "coo," Eike reminds people of that he is a symbol of love and peace. However, Eike just started working for that new job. He tends to work together with the user to find solutions to their problems and ensure that the user feels like their questions are taken seriously and are answered with great enthusiasm. He is aware of the possibility of failing to satisfy the users, yet he is always sincere to apologize and is humble to learn and improve himself.

3.2.3 Conversation Flow Design – Co-design Workshop II

The next workshop was planned and conducted to resolve the remaining two design questions:

- what represents a natural conversation flow between the user and the chatbot, which enhances the efficiency of question answering;
- how to motivate users training the chatbot while engaging in a natural conversation.

Accordingly, the objective of this workshop was to co-design a conversation flow covering the well-worn path most users will follow as well as the long tail of paths that remain, i.e., the low-confidence and failure paths. In the meantime, motivating users to rate the answer had to be taken into account in the conversation.

Participants

Given that participants of the first workshop were very productive and passionate about this project and chatbot design, they were invited again to this workshop as co-designers (see section 3.2.1 Chatbot Avatar Creation – Co-design Workshop I). I used the same construct of working in promoting the workshop, in which I still worked as a facilitator and design contributor.

Procedure

The co-design workshop was carried out in June 2018 at Aalto University. The sessions were planned for two and a half hours. The sessions consisted of five parts:

1) Introduction (10 minutes)
2) Familiarizing with the chatbot persona (10 minutes)
3) Brainstorming user personas (25 minutes)
4) Sample dialogue writing (45 minutes)
5) Sketching high-level conversation flow (30 minutes)

Introduction. This workshop commenced with an introduction of the objective of this workshop and the context of the chatbot to be designed (e.g., the technical capabilities and critical use cases). Clarifying the background information of the ERICS project was omitted, as the participants were already familiar with it.

Familiarizing with the chatbot persona. Later, they familiarized themselves with the determined chatbot persona incorporating the visual representation and personality (see more information in Section 3.2.2 Feedback Collection – Identification of the Final Chatbot Avatar). It was essential for the participants to understand this system persona representing the brand and mission before they wrote sample dialogues for it.

Brainstorming user personas. After that, the participants were split into two groups of two. They were asked to brainstorm user personas based on their knowledge about refugees and migrants or their own stories and experiences (see Figure 24).

Sample dialogue writing. With the support of generative toolkits, participants started to create sample dialogues in a role-playing fashion where one played Eike, and the other one performed as the user persona they had devised. Each group was required to create three sample dialogues for the primary – high confidence – path, the alternative – low confidence – path, and the failure path respectively.

Sharing and discussing outcomes. Eventually, by sketching a high-level conversational flow[15] on a whiteboard, the participants presented and discussed the sample dialogues to abstract the flow and logic of the conversation.

[15] A high-level conversational flow is the overall conversational design covering the question-answering system 's architecture and flow. It describes the relationship between various modules and functions of the question-answering system.

Figure 24. The participants are creating user personas in co-design workshop 2

Generative Toolkits

To facilitate the participants to create the conversational flow, they employed a generative toolkit to improve their 'design abilities.' This toolkit comprised two small packages (see Figure 25). The first package was used to produce user personas, containing persona sheets, photos of diverse people, pens, sticky notes, scissors, and double-sided tapes. Following the template of user persona creation by *Actions on Google* ("Create a Persona," n.d.), the persona sheet was devised with three sections:

1) "who is the user?";
2) "what are the user's goals?";
3) "what is the user's context?".

The second package supported the sample dialogue writing including a pile of tip cards and conversation cards. The tip cards illustrating each conversational component's definition, examples, and scenarios of usage guided the participants to leverage conversational components (see more information in Session 2.2.4 Conversation Design), and the conversation cards functioned as a tool for participants to write user utterances and chatbot prompts.

Figure 25. The package of the generative toolkit for creating conversational flows

Contexts

Due to the limitation of the database at that moment, the chatbot is planned to focus on specific domains with use cases such as language learning, vocational, and job application as well as school and university application. The six major categories for the refugees and migrants' questions are defined:

1) language learning (e.g., "how can I learn the German language?"),
2) internship application (e.g., "how can I find an internship position?"),
3) vocational training (e.g., "what qualifications do I need to fulfill to apply for school-based vocational training?"),
4) school (e.g., "how does the German school system work?"),
5) university application (e.g., "how can I enroll in a German university?"), and
6) student finance (e.g., "what kind of scholarships are there?").

The format of displaying information was written in plain text, and the length of each answer varied from 14 to 113 words.

As for the technical capability of the system that the chatbot relies on, it is not intelligent enough to correctly understand and process the users' request with considering current contexts at the beginning of the chatbot development. Consequently, the chatbot is deployed with utilizing command-driven Natural language Processing and a one-turn way of question answering. When the

chatbot fails to parse the user's query precisely and offer answers with high confidence, it asks the user to confirm by selecting a preferred answer from a short list of possible candidates which are strongly relevant to the received query.

User persona creation

Before writing dialogues, it is necessary to have a clear picture of who is communicating. By building the user personas, the participants can clearly understand the users they are designing for in aspects of demography, such as their needs, experiences, behaviors, and goals. The participants devised two user personas representing a refugee and an economic migrant respectively (see Figure 26). Table 14 and 15 display the detailed information of the defined personas.

Figure 26. The created user personas by the co-design participants

Table 14. The user persona 1 (migrant)

Who is the user?	Sarah, 25, is a fresh graduate of Liberal Arts and a dedicated traveler. Sarah traveled a lot when she was younger and met his current boyfriend during her backpacking trip to the Far East. After a year of dating long distance, she decided to leave the US for Germany.
What are the user's goals?	Sarah appreciates culture, good food and social equality in Germany. She is motivated by exploring new experiences and meeting new people in the new country. Sarah wants to learn German so that she can have no troubles in visiting rural regions of Germany.
What is the user's context?	Sarah is browsing relevant information about language learning online at home. She happens to log in to the HBG website.

Table 15. The user persona 2 (refugee)

Who is the user?	Yassir, 35, is currently unemployed. Yassir arrived in Germany from Syria in mid-2015. Yassir appreciates work, family and his hobbies. During his spare time, he plays soccer with his two sons and teaches them history. He is interested in Islamic history and reads a lot. His family and new knowledge motivate him.
What are the user's goals?	The 35-year-old quickly realized that finding work would be harder than he thought. Even when having a teaching license in Syria, he couldn't find a job even from a café. To get a proper job, Yassir would like to take part in a vocational training programme to gain new skills.
What is the user's context?	One of Yassir's friends tells him the HBG website where he could find information about vocational training. He then logins to the site through his phone.

Sample dialogue writing

Through role-playing in pairs, the participants first produced the most common conversational path where users successfully had their questions answered. One group member played Eike and wrote display prompts exhibiting the defined personality traits, whereas the other one represented the created user persona approaching Eike with a particular goal. The participants wrote down the prompts on the conversation cards and then put them on a whiteboard to simulate the actual usage on a chat interface as much as possible (see Figure 27). This procedure and the method were repeated when generating the sample dialogues for the low-confidence path and the failure path. Through discussion and refinement of all the sample dialogues, they jointly produced the final three proposals.

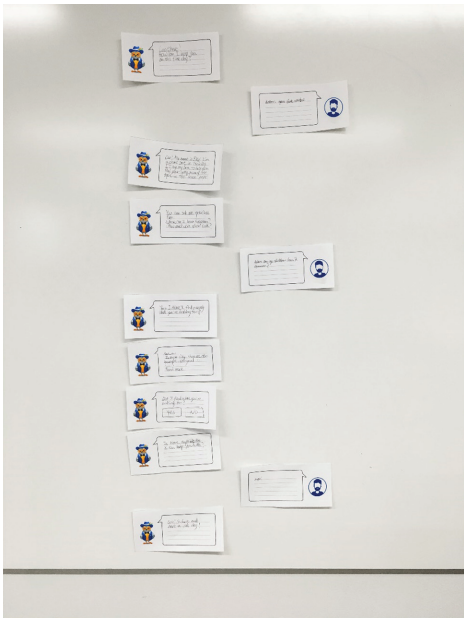


Figure 27. A simulated conversation between Eike and a user through the conversation cards

The high-confidence path (see Table 16). In the phase of pre-conversation, initially, Eike gives a greeting, “Coo there! How can I help you on this fine day?”. This greeting not only exhibited the nature of Eike’s character as a pigeon but also conveyed that he is eager to help. After the user opened the chat window, Eike starts the conversation with an introduction, “Coo! My name is Eike! I’m a Chatbot in training. I’ll try my best to help you find your way around the topics in the ‘learn’ page!”. This prompt is sent to the users for the following intentions:

1) give a brief self-introduction,
2) set expectations of capability, and
3) define the scope of the database.

Then Eike provides the user with some suggestions (i.e., examples of questions) for optimal use. Next, once Eike possesses high confidence to generate an answer to the user’s query by sending a confirmation, “Yay, I think I found exactly what you were looking for!”. This move reassures the user in a socially and conversationally appropriate way and helps carry the thread of the conversation forward by maintaining the context. In terms of the rating, the in-house developer suggests having a pull-out Likert scale rating when the user hovers over the prompt (not only just the answer) sent by Eike, which can help to identify improper prompts as well as develop and improve the chatbot holistically. The Likert scale can measure the user’s attitude towards the given message in terms of direction (by ‘agree or disagree’) and intensity (by ‘strongly’ or not). It contains the following options: strongly agree, agree, neither agree nor disagree, disagree, strongly disagree.

Nevertheless, enabling users to rate each message can hardly guarantee motivating users to assess the answer. When conversing with chatbots, users are frequently cooperative, and they prefer to follow instructions proposed by the chatbot (“Learn About Conversation,” n.d.). In light of this, Eike sends a follow-up question after delivering the answer, “Did I find what you were looking for?”. The user is able to easily respond by tapping the chips with “yes” or “no.”

In the last phase of conversation, the user's intent to closure a dialogue is not steadily explicit to the chatbot and sometimes even not apparent to the user himself or herself (Huang, Lasecki, Azaria, & Bigham, 2016). In that case, Eike sends a follow-up question, "Is there anything else I can help you with?". If the user gave an unambiguous response with "yes" or "no," Eike will conclude the conversation is over. However, if the user abandons the Eike without giving a hint, then Eike will assume the conversation has ended after waiting for a certain period. Ultimately, to create exit experiences for the user, Eike gives a farewell, "Coo! So long, and have a nice day!".

The low-confidence path (see Table 17). Even for a robust chatbot, there is still room for risks that the chatbot cannot correctly understand or interpret the user's query in context. As a result, the chatbot becomes less confident in question answering. In that case, Eike is honest with his limitation of database or intelligence and then provides a list of the top three answers which are incredibly relevant to the given topic. He sends a prompt like "Sorry, I am not sure I had a proper answer, but I found something relevant. Please select the one you prefer.". This move not only supports the user to make their question answered but also trains the chatbot in terms of increasing the question-answer pairs.

The failure path (see Table 18). The failure path occurs when the chatbot fails to understand the user query and complete the task by providing corresponding correct answers. In an extreme scenario where the user can not even find a favored answer from a suggested list of possible answer candidate, he or she is allowed to tap the "none-of-these-above" chip to seek further support. Following the confirmation of failure, Eike apologizes and asks the user to rephrase his or her question, "Sorry, I didn't understand that. Could you say that differently?". At the same time, a chip written with "Help, I'm stuck!" can be shown to the user, too. This chip of further supports is able to direct the user and the undefined question to the staff of the HBG website, which functions as an alternative solution for the system error.

Table 16. The initial high-confidence path

Speaker	User utterance / Spoken prompt
Eike	“Coo there! How can I help you on this fine day?”
User	Open chat window <chips>
Eike	“Coo! My name is Eike! I’m a Chatbot in training. I’ll try my best to help you find your way around the topics in the “learn” page! ” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	“You can ask me questions like: Where can I learn German? How much does school cost?” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
User	“Where can my children learn German?”
Eike	“Yay, I think I found exactly what you were looking for!” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	“In some cities, there are, for example, bilingual kindergartens ... remember to speak to your kids in the language that you are best at. Read more.” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	“Did I find what you were looking for?” Yes/No <chips>
User	“Yes”
Eike	“Is there anything else I can help you with?” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
User	“Nope” (or no rep ly)
Eike	“Coo! So long, and have a nice day!” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>

Table 17. The initial low-confidence path

Speaker	User utterance / Spoken prompt
Eike	"Coo there! How can I help you on this fine day?"
User	Open chat window <chips>
Eike	"Coo! My name is Eike! I'm a Chatbot in training. I'll try my best to help you find your way around the topics in the "learn" page! " Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	"You can ask me questions like: Where can I learn German? How much does school cost?" Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
User	"Tell me about all the types of vocational training."
Eike	"Sorry, I am not sure I had a proper answer, but I found something relevant. Please select the one you prefer." Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	"Answer 1:..." <chips> "Answer 2:..." <chips> "Answer 3:..." <chips> "None of these above! <chips>
User	Answer 1
Eike	"There are three different types of vocational training: the dual vocational training ('Duale Berufsausbildung'), the school-based vocational training ('Schulische Ausbildung') and dual vocational degrees ('Ausbildungsorientiertes duales Studium')." Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	chips "Did I find what you were looking for?" Yes/No <chips>
User	"Yes"
Eike	"Is there anything else I can help you with?" Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
User	"Nope" (or no reply)
Eike	"Coo! So long, and have a nice day!" Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>

Table 18. The initial failure path

Speaker	User utterance / Spoken prompt
Eike	“Coo there! How can I help you on this fine day?”
User	Open chat window <chips>
Eike	“Coo! My name is Eike! I’m a Chatbot in training. I’ll try my best to help you find your way around the topics in the “learn” page! ” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	“You can ask me questions like: Where can I learn German? How much does school cost?” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
User	“Tell me about all the types of vocational training.”
Eike	“Sorry, I am not sure I had a proper answer, but I found three relevant answers. Please select the one you prefer.” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	“Answer 1:...” <chips> “Answer 2:...” <chips> “Answer 3:...” <chips> “None of these above <chips>
User	“None of these above.”
Eike	“Sorry, I didn’t understand that. Could you say that differently?” Help, I’m stuck! <chips>
User	“Help, I’m stuck!”
Eike	“Please contact our staff via Facebook, Email or Phone.” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
Eike	“Is there anything else I can help you with?” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>
User	“Nope” (or no reply)
Eike	“Coo! So long, and have a nice day!” Did this message help you? <a pull-out when hovering over the message> :C - :(- :I - :) - :D <chips>

High-level conversation flow creation

After the participants finished the sample dialogue creation, they started to summarize a high-level flow and logic of the conversation on a whiteboard. By doing this, it became explicit that the dialogue structure was multidimensional—that is, variation in the dialog was dependent on the three of the following:

1) whether the Q&A system was confident to answer the user's query,
2) whether the Q&A system had a matched answer to the undefined query,
3) whether the user was paraphrasing the ill-defined or undefined query.

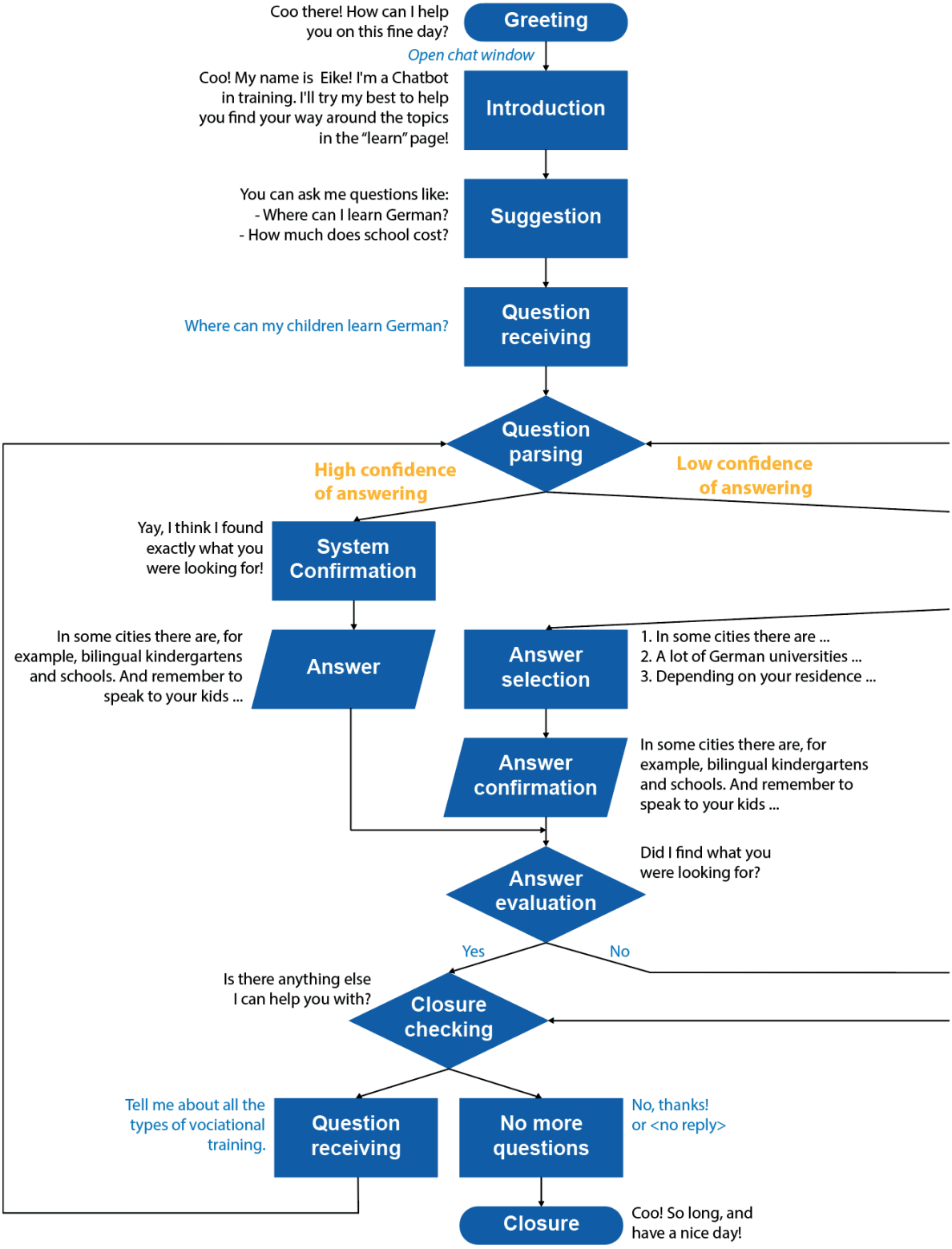
This multidimensional dialogue structure led to the initial high-level conversation flow showing the functionality covered in the sample dialogues above (see Figure 28). Fundamentally, this conversation flow accounts for the following paths of interaction:

a) answering user queries with high confidence,
b) offering options of topic-related answer candidates to remedy the possible negative experience caused by the low confidence of answering the ill-defined or undefined query,
c) paraphrasing the query that is ill-defined or undefined by the QA system,
d) providing follow-ups for the failure of question-answering.

The skeleton of the conversation flow primarily consists of the path of high-confidence question answering. This main conversational interaction starts with Eike's greeting, self-introduction, and suggestions of queries that the user may want to ask. Once Eike responds to the successfully defined user query, he immediately asks the user to evaluate the given answer to improve the quality of the database, i.e., the question-answer pairs. If the user gives negative feedback, then he or she will be directed to the failure path. Besides, the user can utilize a Liker-scale rating to verify each particular message sent by Eike. After one successful round of question answering, Eike checks out whether the conversation was over. If it is continuing, the user will operate another round of question answering. But if the dialogue is terminated, then Eike will give a farewell to create exit experience for the user.

The structure of the conversation flow is replenished by designing the long tail, namely, handing as many cases as possible where the thing can go wrong. The conversation flow contains multiple logical decision point and error and rejection handling. More specifically, Eike provides a list of top three possible answers relevant to the user's query when the query is challenging to understand or hard to interpret by Eike.

Furthermore, Eike offers a path of redefining the unmatched query, where the user is proposed to paraphrase his or her question. If the user is unwilling to do that or stuck in the failure of question answering, Eike admits failure and provided him or her with a follow-up, in which the user can reach out to the human assistance.


Coo there! How can I help you on this fine day?
Greeting
Open chat window
Coo! My name is Eike! I'm a Chatbot in training. I'll try my best to help you find your way around the topics in the "learn" page!
Introduction
You can ask me questions like:
- Where can I learn German?
- How much does school cost?
Suggestion
Where can my children learn German?
Question receiving
Question parsing
High confidence of answering
Low confidence of answering
Yay, I think I found exactly what you were looking for!
System Confirmation
In some cities there are, for example, bilingual kindergartens and schools. And remember to speak to your kids ...
Answer
Answer selection
1. In some cities there are ...
2. A lot of German universities ...
3. Depending on your residence ...
Answer confirmation
In some cities there are, for example, bilingual kindergartens and schools. And remember to speak to your kids ...
Did I find what you were looking for?
Answer evaluation
Yes
No
Is there anything else I can help you with?
Closure checking
Tell me about all the types of vociational training.
Question receiving
No more questions
No, thanks!
or <no reply>
Closure
Coo! So long, and have a nice day!

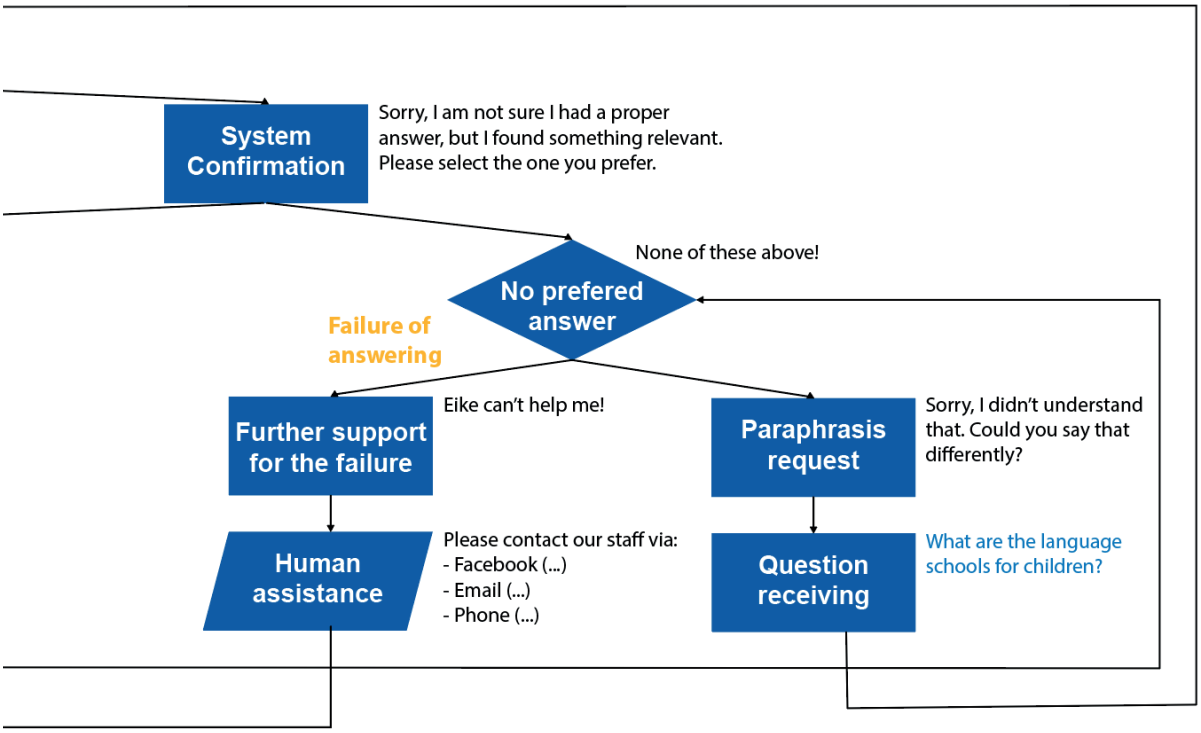


Figure 28. The initial high-level conversational flow

3.2.4 Graphical User Interface & Interaction Design

Following the defined visual representation of the chatbot avatar and the determined conversational flow, through wireframes and mockups, I started to visually detail the user interface including layouts, color schemes, iconography, visual style, and interaction patterns.

A wireframe is a low-fidelity form for demonstrating structure, content, and functionality on a page. To be specific, it is a layout of a page, which displays what interface elements exist. Before the visual design and content are added, it is highly beneficial in the early development to establish the basic structure of interface ("What Is Wireframing," n.d.). Also, a wireframe helps to provide a visual understanding of a page and the logic of interaction (e.g., navigation). As opposed to wireframes representing a product's structure, a mockup reveals the appearance of the interface. A mockup supports making decisions on a user interface's color schemes, visual style, typography, and iconography.

Layouts

Given that the chatbot is required to be embedded without rebuilding the HBG web service, Eike is set in the form of a built-in window on the website. Eike's icon locates in the lower right corner on the webpage, which avoids interfering with the user browsing the website's content (see Figure 29, top). The dialog window pops up on the right side when the user commences the conversation (see Figure 29, bottom).

The dialog window consists of four parts: 1) the chat-window bar, 2) the display area of dialog, 3) the typing area, and 4) the close button. Eike's name and image are presented in the chat-window bar, which conveys whom the user is talking to. In the conversational display space, Eike's prompts are arranged on the left side, while user utterances are placed on the left. Furthermore, the dialog box has an arrow-like corner pointing toward the interlocutor indicating who is the sender of that message.

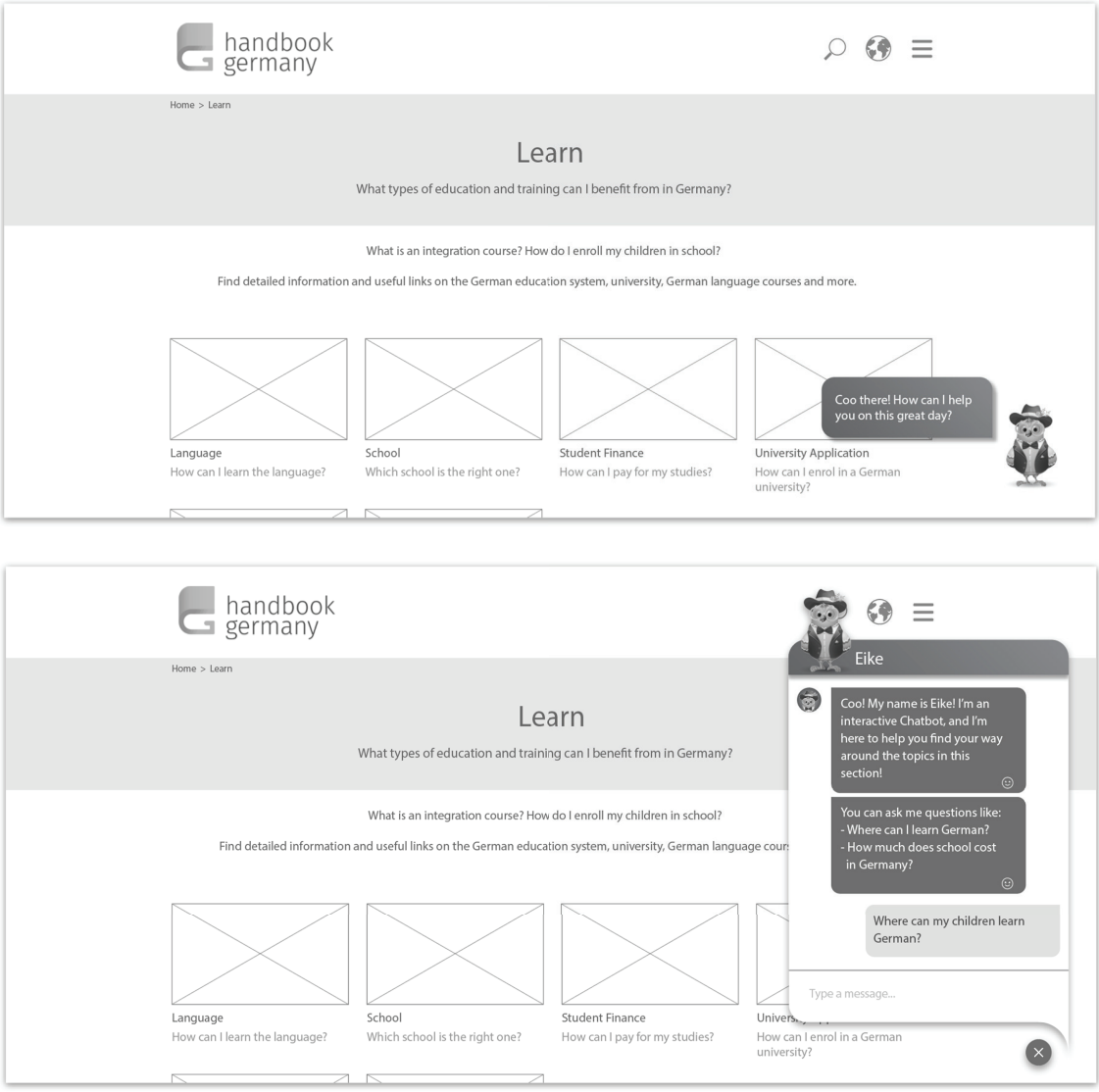


Figure 29. The wireframe of the user interface
The avatar of Eike on the HGB website (top); **The chat window** (bottom)

Color usage and palettes

In line with the principles of the color system by Material Design ("The Color System," n.d.), the color scheme follows four principles:

1) hierarchical,
2) legible,
3) expressive, and
4) consistent.

First, the assigned color indicates the interactivity, relatedness, and level of prominence of a particular interface component (significant elements should stand out the most). Second, texts and critical elements (e.g., icons) meet legibility standards when they appeared on the colored backgrounds across all screen. Third, the interface manifests brand colors at memorable moments so as to reinforce the brand's style. Last, the color is utilized consistently in the chatbot so that certain colors always mean the same thing, even if the context changes. The consistency of the color palette is beneficial in shaping the user behaviors and their mental models, which can enhance the effectiveness of the chatbot.

The color palette consists of blue, yellow, and grey (see Figure 30). Blue is the primary color whereas yellow and grey are part of the secondary palette. In the chat interface, the primary (blue) and the secondary color (yellow) are selected to represent the brand, the HBG portal (see Figure 31). To be specific, blue is utilized on the chat-window bar and the system prompts. However, to create a contrast between them, the light blue is chosen for the system prompts while the dark blue is used on the chat-window bar. For the user, the chatbot's responses are more significant compared to their own messages. Therefore, I assign grey to the user's dialog box, which draws less attention. In regards to the yellow, it is used for buttons and interactive areas. For the purpose of enabling the text to appear clearly and legibly against the colors behind them, the white text presents on the blue dialog box, and the black text fits the yellow buttons and the grey dialog box.

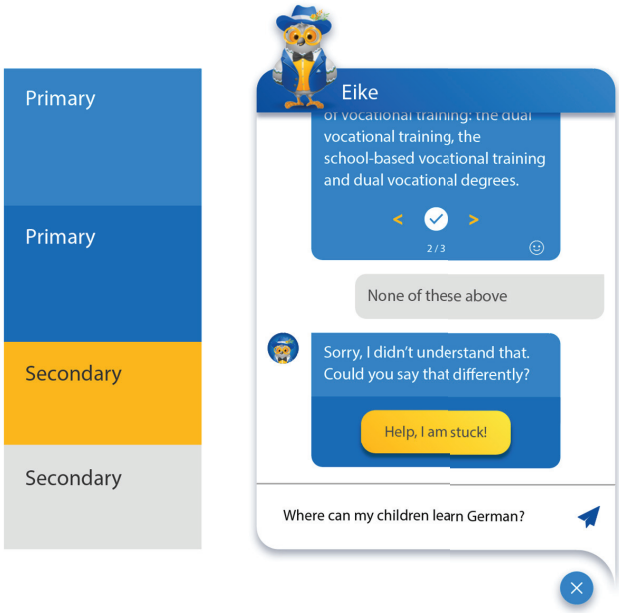


Figure 30. The color palette of the chat interface

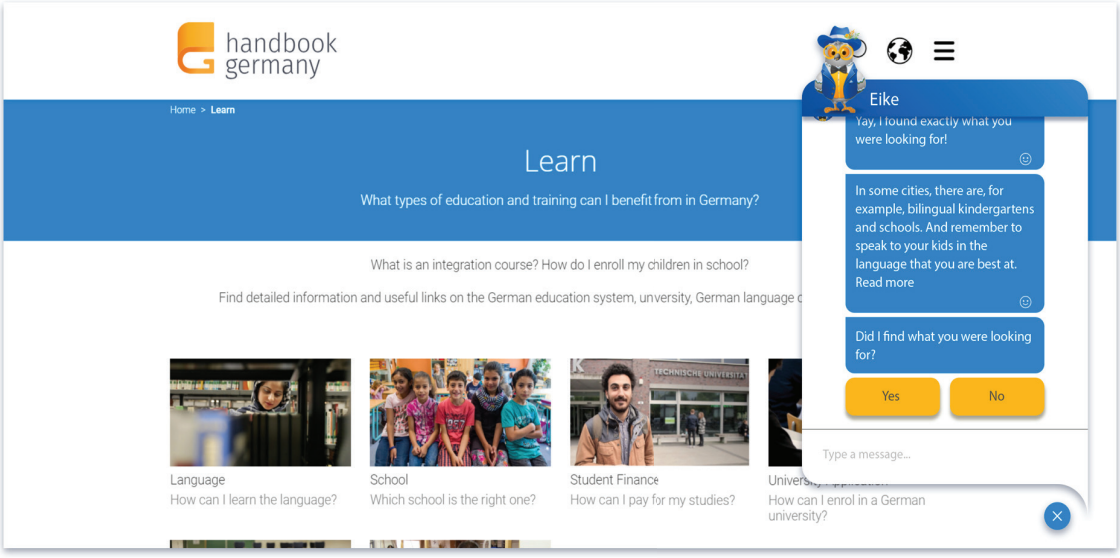


Figure 31. The color scheme of the chat interface in line with the context (the HBG website)

Iconography

Icons help identify actions and provide information. When designing the icon, I take understandability and legibility into account. In terms of rating message, a smiley icon was utilized as an indication giving feedback about feelings towards a particular prompt (see Figure 32, top-left). These icons are ambiguous so as not to arouse any negative emotions among users about gender or race. Once the user starts the evaluation, Eike presents five different emojis representing five attitudes: strongly agree, agree, neither agree nor disagree, disagree, strongly disagree, respectively (see Figure 32, bottom-left). I leverage a color gradient from red (strongly disagree) to green (strongly agree) to reflect emotional reactions. According to Mammarella et al., color has impacts on cognitive processing such as affective connotation and emotional responses. For example, red is commonly associated with a negative personal experience whereas green is linked to positive and relaxing experiences (Mammarella, Di Domenico, Palumbo, & Fairfield, 2016). After the user evaluates the message, the color of the rating icon changes, in order to confirm that the user's action has been accomplished. Also, it indicates his or her attitude towards the given answer (see Figure 32, top-right). The close icon is designed to be simple and classic the user, and it is reduced to its minimal form expressing essential characteristics, i.e., termination (see Figure 32, bottom-right).

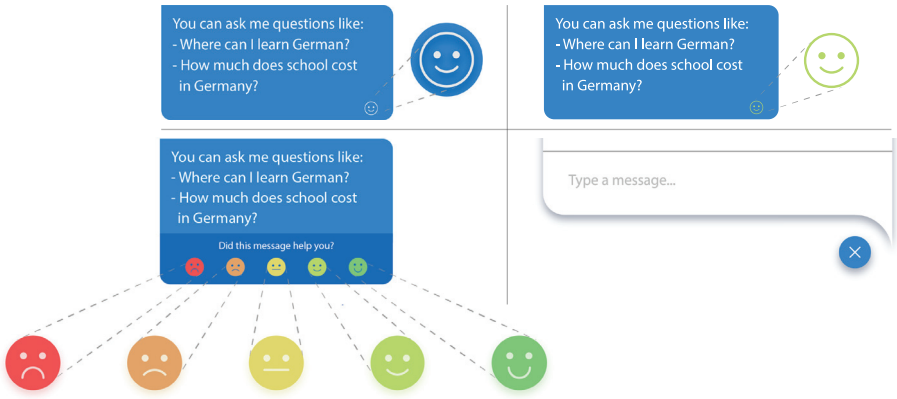


Figure 32. The iconography of Eike the chatbot
The rating prompt (top-left); **The icon indicating the prompt was rated** (top-left); **The rating chips** (bottom-left); **The close button**(bottom-right).

Mockups depicting the conversational paths

Figure 33 illustrates the path of high confidence in the dialogue, and the low confidence path and failure path are represented respectively in Figure 34 and Figure 35.

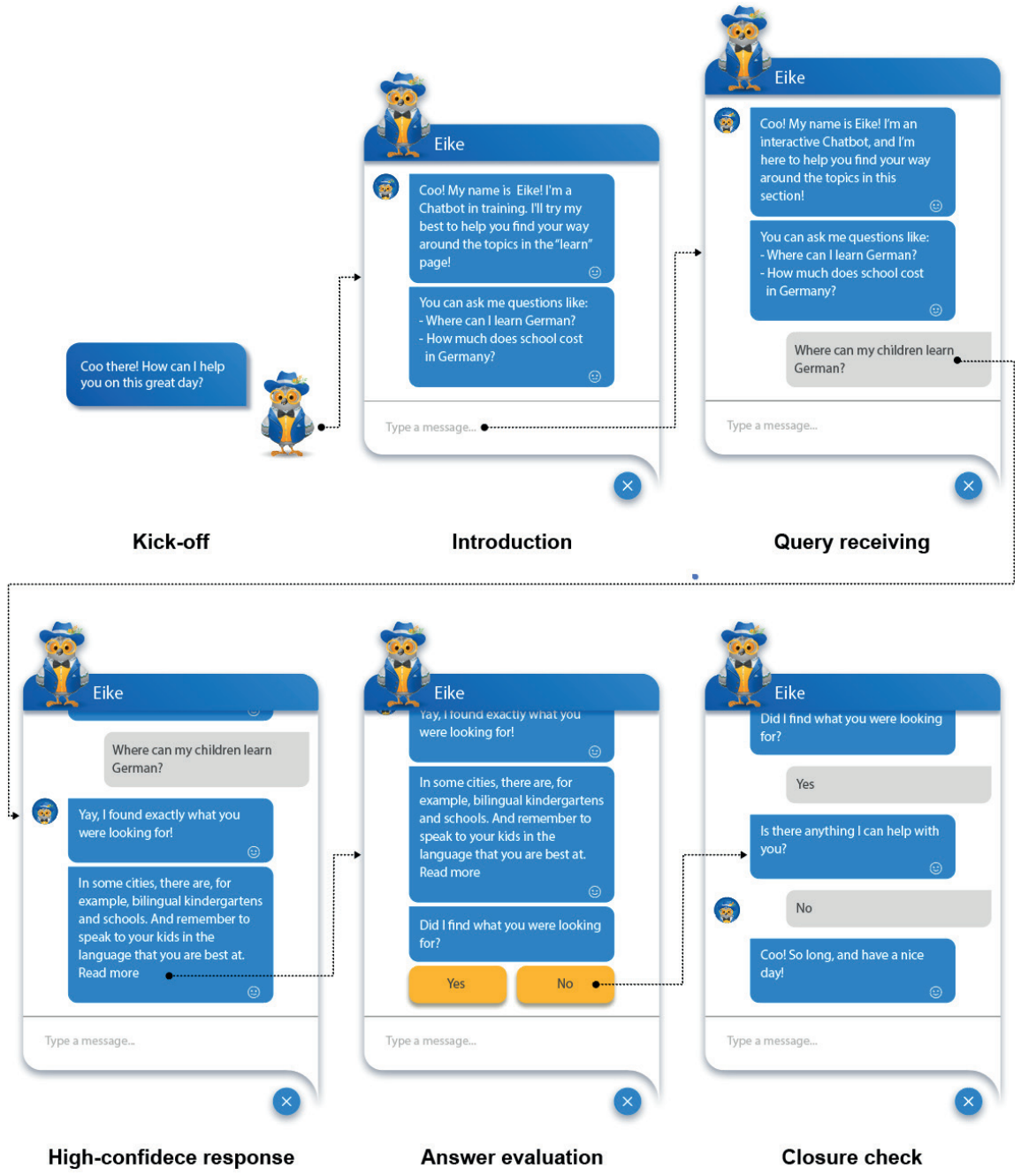


Figure 33. The initial mockups illustrating the high-confidence path

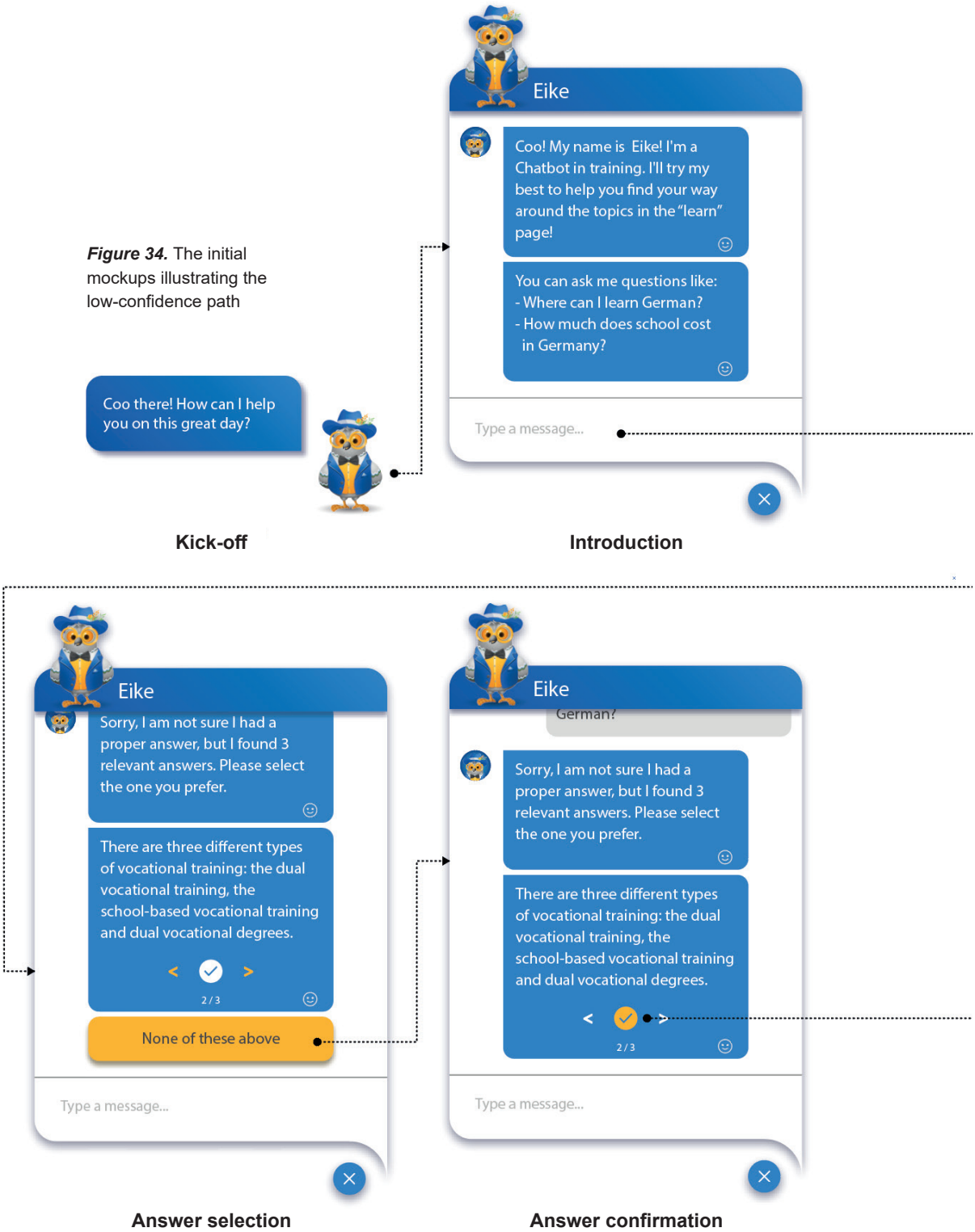
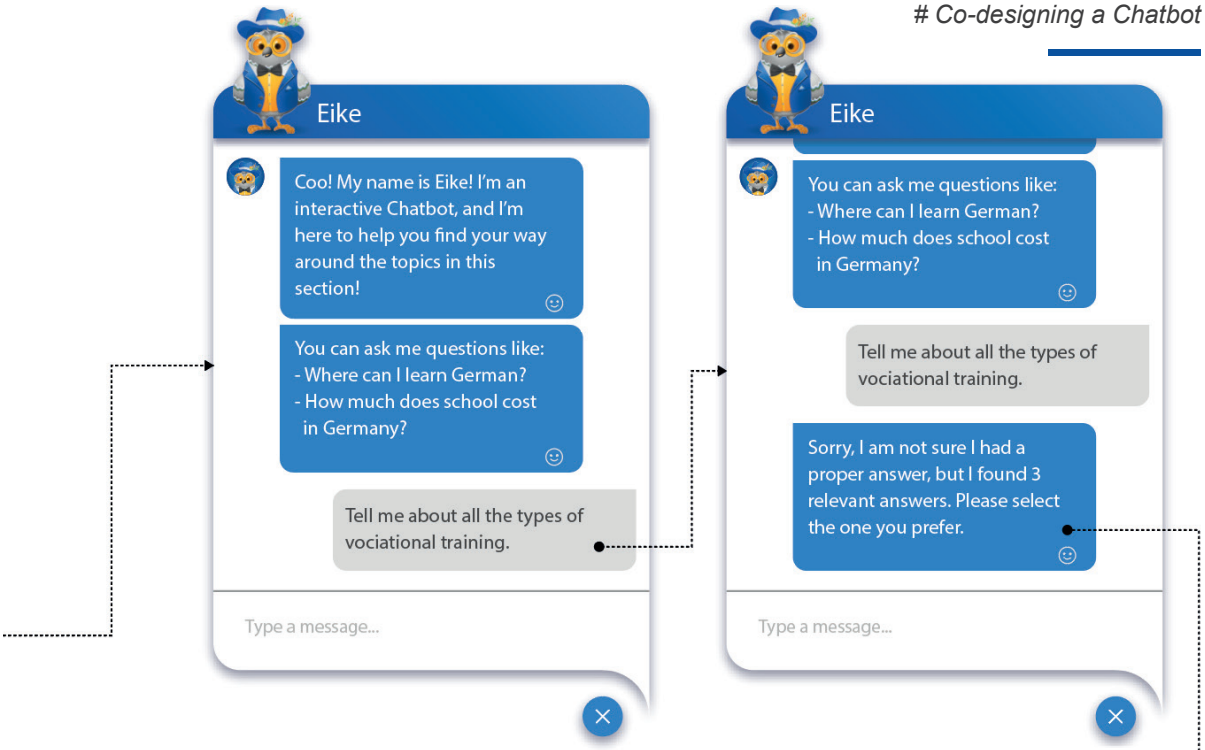
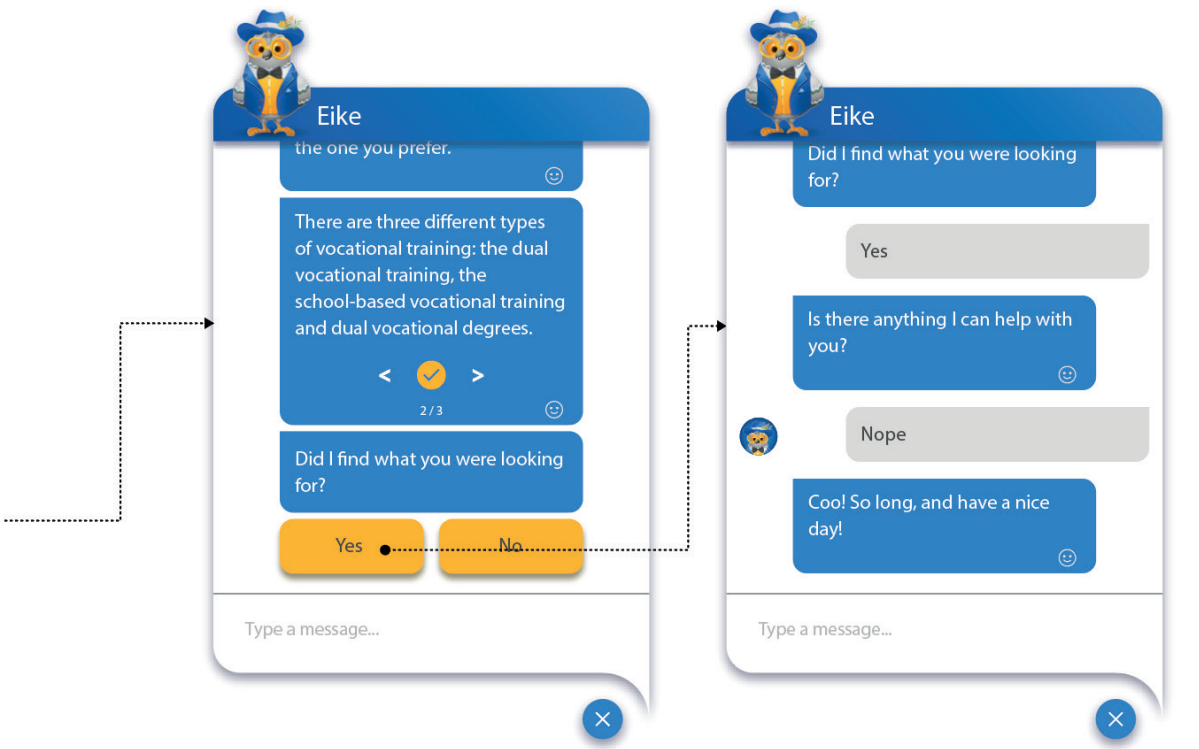


Figure 34. The initial mockups illustrating the low-confidence path



Query receiving **Low-confidence confirmation**



Answer selection **Closure check**

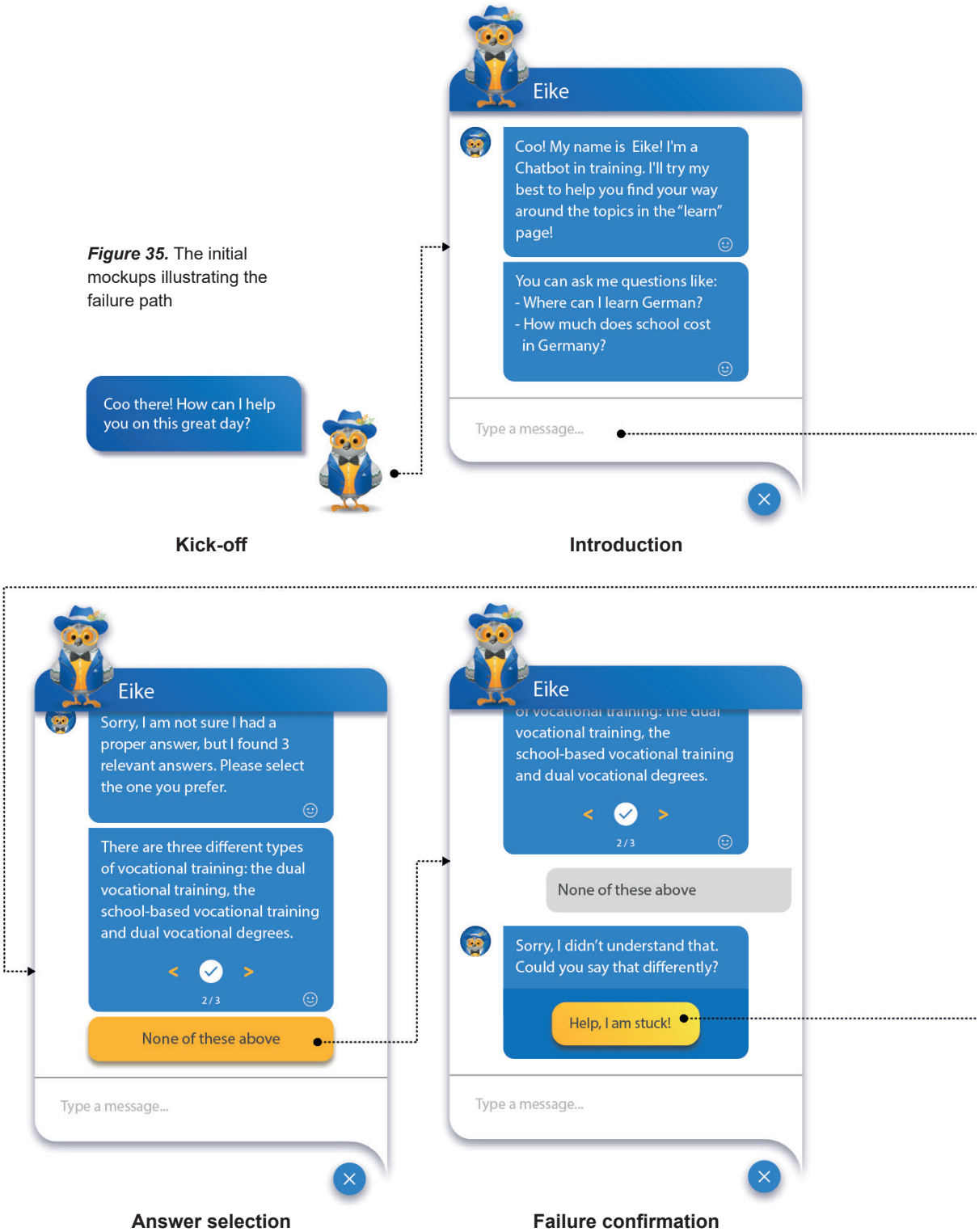
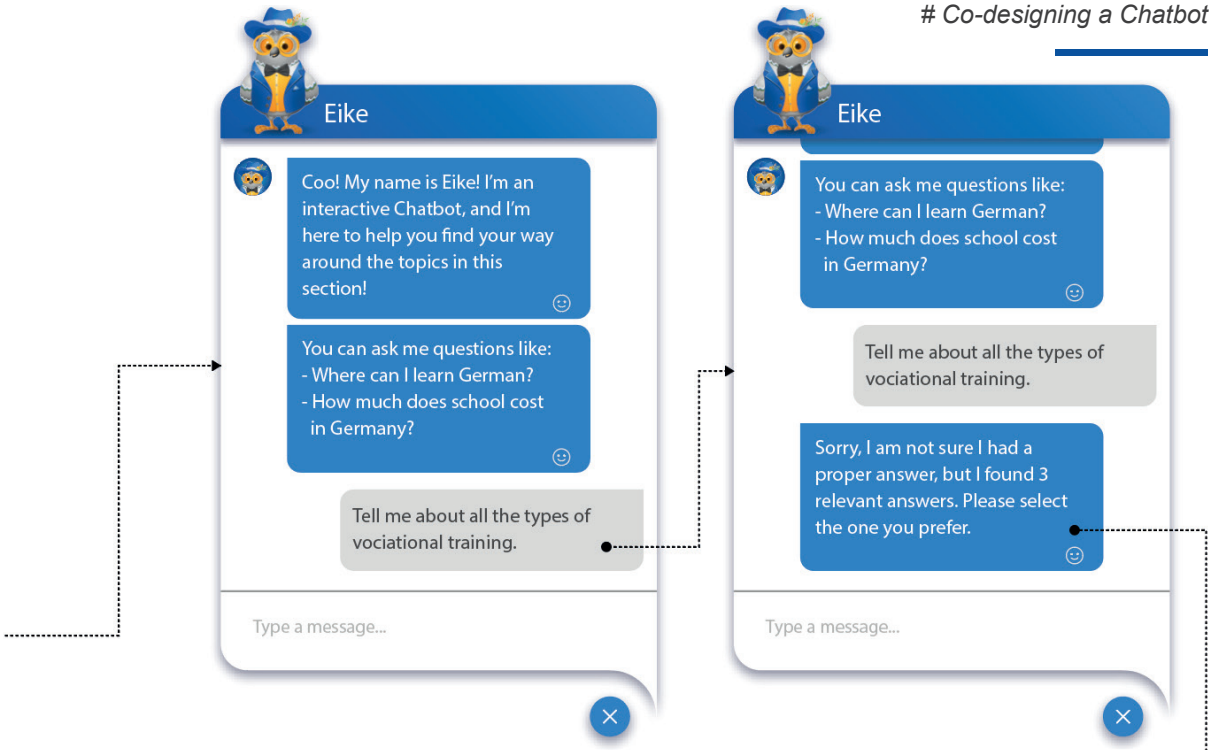


Figure 35. The initial mockups illustrating the failure path



Query receiving **Low-confidence confirmation**

Further support **Closure check**

Interaction

Proactive engagement. Once the user logs onto the 'learn' page of the HBG website, Eike greets the user and expresses his willingness to support (see Figure 36, top-left). As a means of engaging, this move demonstrates Eike's enthusiasm for soothing worries and providing refugees and migrants with his knowledge about living in Germany.

Chips as quick responses. To overcome technological shortcomings and avoid misunderstandings resulting in a back-and-forth dialogue with the user, Eike utilizes chips to help the dialogue progress, such 'Yes' and 'No' (see Figure 36, top-right). In addition, by using these, the user is able to take less time to navigate the chat interface than typing out a response.

Immersive conversational experience through typing indicators. Eike utilizes a tying indicator to create an impression where he is in the process of typing out a new response (see Figure 36, bottom-left). The tying indicator is beneficial for creating an impression that Eike is converse with the user naturally to let users immerse themselves in the conversation.

Answer selection through browsing carousels. When Eike has low confidence to parse the query and give the corresponding answer correctly, the relevant answer candidates are shown to the user through a browsing carousel (see Figure 36, bottom-right). A browsing carousel allowed users to select one of many items. To guide the user to determine a preferred answer to the undefined question, Eike clarifies the following information in the prompt:

1) How many items were in the carousel (e.g., "I found 3…")
2) Why these items were chosen (e.g., "I found 3 relevant answers.")
3) What action the user was encouraged to do (e.g., "Please select the one you prefer.")

Once the user selected the answer by clicking the check button, the answer is then accepted as Eike's responses. Except for given answer candidates, a "None of these above" chip is provided to let the user indicate he or she did not want any of the options.

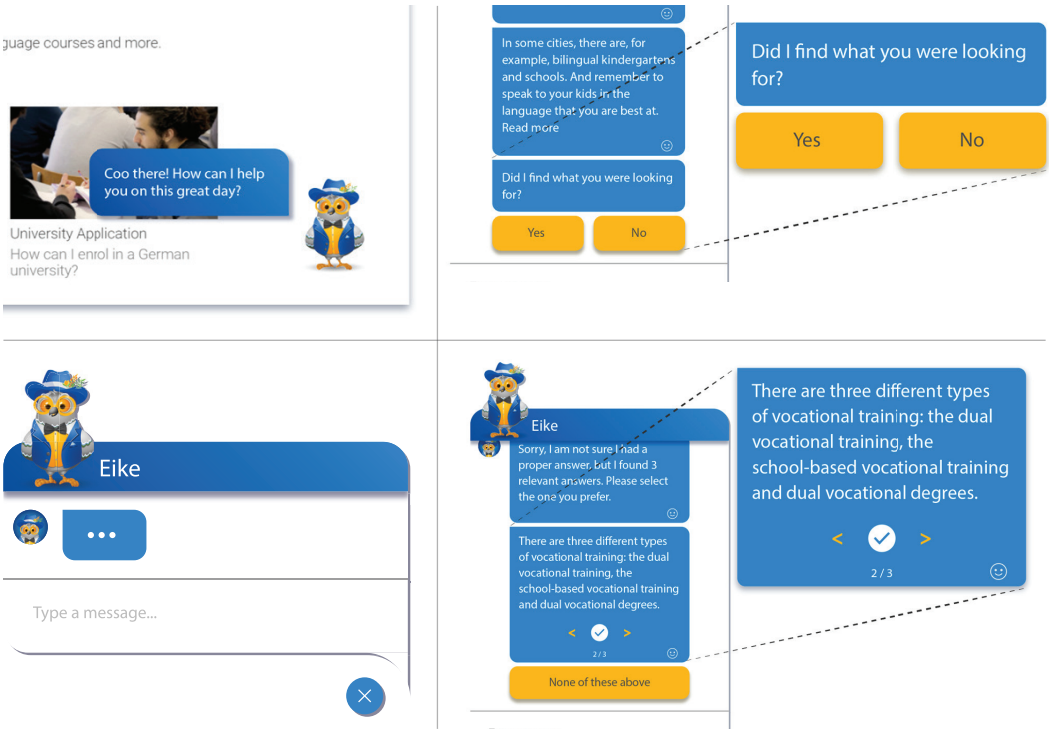


Figure 36. The interaction patterns of Eike the chatbot
Eike proactively greets the users when they log onto the 'learn' page of the HBG website (top-left);
The chips for confirming the validity of the given answer (top-left);
The typing indicator of Eike (bottom-left);
The answer candidates are presented in a browsing carousel (bottom-right).

Message rating in a visual fashion. Instead of using texts, Eike reminds the user of evaluating a particular message through emotive icons. A visualized drop-down five-degree rating chip occurs after the user clicked the rating icon, which is designed not to interrupt the conversation (see Figure 37).

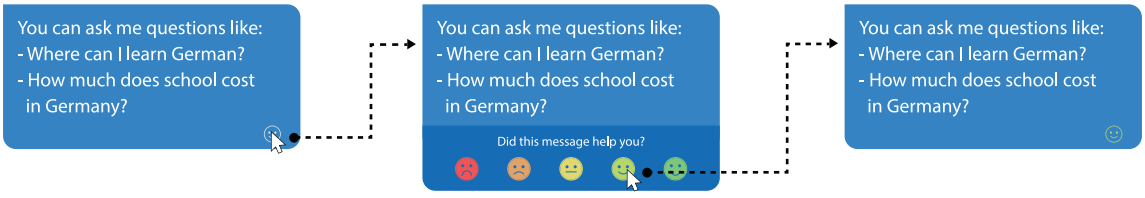


Figure 37. The process of rating an answer

Facal points for efficient navigation. Users can be easily distracted by a large amount of information and a plethora of interface elements. The chat interface is designed with considering focal points, which can ensure the users had a clear hint in a particular area where they are going to perform an action. For example, in the typing area of the chat window, the users only see a prompt, “Type a message…”, before they input a query. Once they start to type out a question, a ‘send’ icon then appears indicating the next step – sending the message) (see Figure 38).

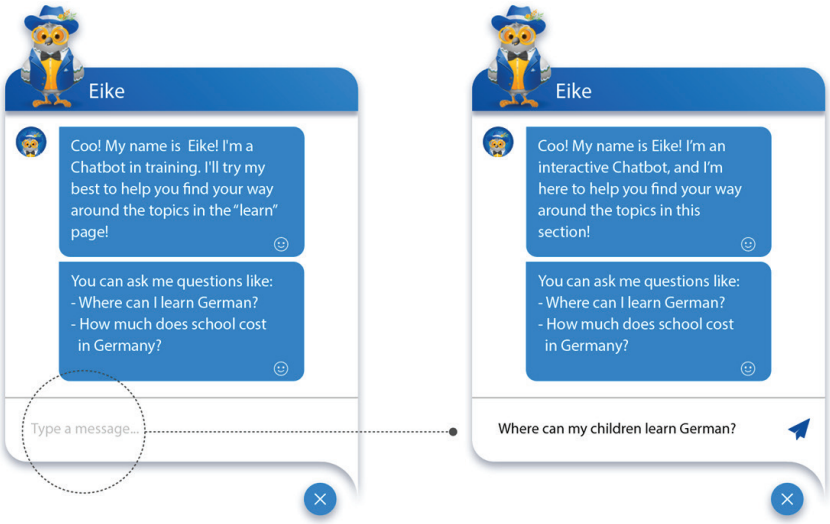


Figure 38. The hints guiding the user where to type and send a message

3.3 Prototype & Evaluate

3.3.1 Concept Prototyping

At the later stage of *make*, the defined concept is manifested in the form of a prototype, which functioned as a tool for testing the interaction quality and understandability of the chatbot. A prototype plays an important role in the explorative design process, as the solution is forced into a concrete manifestation allowing designers to identify new problems and generate insights that can contribute to the following iterations. An idea represented through tangible prototype (either built or visualized) can be easily verified, refined, and improved than the linguistical description (Eppinger & Ulrich, 1995).

In the ERICS project, the chatbot was prototyped by using ProtoPie®[16]. The prototype is clickable, and all essential chips on the chat window (e.g., the rating chips and the answer-selection chips) are interactive. However, due to the lack of technological capabilities in the prototype, users are only allowed to use pre-programmed queries instead of free-form questions. The prototype displayed the interaction and visual design in high fidelity to get accurate feedback from the participants.

[16] ProtoPie (https://www.protopie.io/) is a prototyping tool for smart devices. ProtoPie makes elaborate prototypes possible without any coding, and these prototypes can be tested on actual devices.

3.3.2 User experience evaluation

The goal of this study is to measure summative user experience and identify issues of the concept. User experience is evaluated by using the MAX (Cavalcante, Rivero, & Conte, 2015) and property checklists (Jordan, 1998). The MAX is a post-use method for assessing the general experience (emotion, ease of use, usefulness, and intention to use) through cards with an avatar and a board (see more information in Section 2.3.2 Methods for Validating Chatbots). A property checklist of the graphical user interface was integrated as a tool to unearth the problems further.

I decided to choose the MAX because it can allow the users to evaluate the hedonic and pragmatic aspects of the chatbot playfully. The Method for the Assessment of eXperience (MAX) is a proven UX evaluation method covering both the hedonic and pragmatic aspects. Utilizing the MAX in the context of evaluating chatbots, the category of *emotion* can capture how users feel about the graphic user interface and interaction with the chatbot in general. Based on the categories of *ease of use* and *usefulness*, users can reflect on the usability aspects of the chatbot and how much the chatbot contribute to successful question answering in conversation. In the category of *intention to use*, users can ponder on how much they would like to use or recommend the chatbot to their peers. The users are able to convey their opinions and feelings by picking the cards, which allows collecting the profound interpretation of UX behind the data. Besides, the property checklist allowed me to investigate the details of the design.

The evaluation session happened on two consecutive days. On the first day, I set up shop at International House Helsinki (see Figure 39), and seven different people (3 female, 4 male) of migrant background were invited to join the user experience evaluation. On the second day, ten refugees (seven of them were quite well-educated, and three were poor-educated) with varying in gender (6 female, 4 male) from Arffman Consulting involved in the chatbot evaluation. The evaluations were conducted individually. All sessions were recorded on video.

Figure 39. User experience evaluation at International House Helsinki

Evaluation processes

The user evaluation event consists of three sessions, lasting 30 minutes.

- First, the users experience the chatbot prototype.
- Next, they choose and place MAX cards on a board with prompts questions to report their feelings, opinions, and attitudes of utilizing the chatbot (see Figure 40).
- Finally, the users account for their actions where they selected a specific card and further reflect on a checklist of the product properties in a personal interview.

Figure 40. A refugee is evaluating the prototype by using the MAX

The MAX method

The MAX contains a pile of cards representing different attitudes with the intensity ranging from one to four and a board on which four questions are displayed (see Figure 41 & 42). The users are guided at the evaluation through the four prompts as following:

a) “How did you feel about it?” (emotion)
b) “Was it easy to use?” (ease of use)
c) “Was it useful?” (usefulness)
d) “Do you wish to use it?”(intention to use)

To answer the first question, “How did you feel about it?”, the users are able to choose positive responses such as “*interested*,” “*satisfied*,” “*happy*,” and “*excited*,” or negative responses such as “*bored*,” “*disappointed*,” “*confused*,” and “*sad*.” As for the second question, “Was it easy to use?”, the users responded with either “*It is easy to use*” and “*The use is intuitive*,” or “*I made mistakes*” and “*It was difficult to use*.” The third question, “Was it useful?”, can be answered through the cards saying “*It is useful for me*,” “*It would help me*,” “*It is not useful for me*,” or “*I lost my time using it*.” In regards to the last question, “Do you wish to use it?”, the users can give favorable responses (e.g. “*I liked to use it*” and “*I would use it frequently*”) if they have pleasure in using the chatbot. On the contrary, the cards such as “*I would not use it again*” and “*I would never indicate it*,” can be utilized to represent their attitudes when they undergo negative experience.

Property checklists

The properties of the chatbot incredibly influence user experience. In product development, Jordan claimed five important properties must be taken into consideration: performance, features, usability, aesthetics, and size (Jordan, 1998).

In the ERICS project, performance refers to Eike accomplishing its primary task (i.e., the successful question answering) in a human-like conversation. Besides, handling the low-confidence and failure paths of answering the user query is also part of the performance of the chatbot. Helpful features support

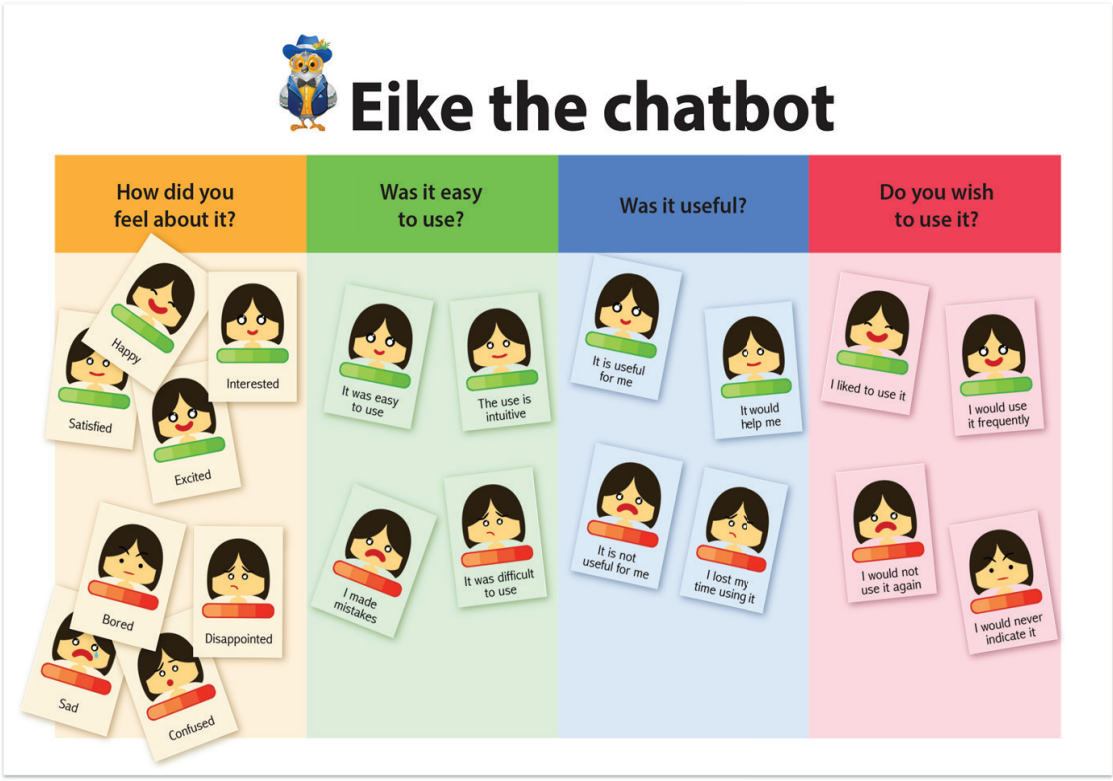


Figure 41. A The MAX for evaluating the chatbot prototype: a pile of cards representing different and a board on which four questions were displayed

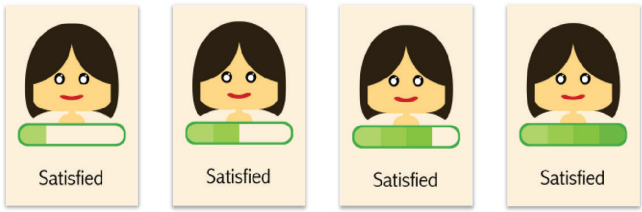


Figure 42. An example of a card representing an attitude with the intensity ranging from one to four

the user to operate the chatbot by doing what is supposed to do efficiently. For instance, the user is allowed to evaluate the given messages (e.g., the answer prompt) for a greater chatbot service. Usability is an extent to which the chatbot can be used by the user to achieve goals with effectiveness and efficiency. The user, for example, can understand each button or icon quickly. Visual design contributes to improving the pleasure of interacting with the chatbot. More specifically, color, layout, and style are important factors to the aesthetic appeal of the chatbot. These five properties of the chatbot were evaluated through a property checklist (see Appendix 3).

Findings

The proposed chatbot concept received positive feedback. Table 19 displays the result of the participants' experiences in interacting with Eike in terms of four aspects: emotion, ease of use, usefulness, and intention to use. 47.3% of the participants reported they were interested in using the chatbot service, and 47.4% stated they felt satisfied when conversing with Eike. In addition, 5.3% of the participants had a happy experience in the chatbot. As for the ease of use, 68.2% claimed it was easy to use the chatbot, and 31.8% considered the usage was intuitive. Half of the participants admitted Eike was useful to facilitate social integration. 66.6% made their attitudes clear that they liked to use the chatbot, and 33.4% reported they would want to reach out to Eike for help frequently in the future.

By asking the participants to reflect on the property checklist (see the result from Table 20), I identify the following key findings:

1) Only rating for the answer. From the perspective of chatbot development, it is ideal to ask the user to give feedback to all the system prompts. However, it is too idealistic and impracticable in real life. The participants report the feature of asking them to verify all the system prompts can give rise to negative emotions and perceptions. It is more likely to be a "heavy task," which they will ignore in most cases. However, it is highly possible for them to verify the given answer if the action is effortless or they are extremely satisfied or dissatisfied with it:

> *"I would not give feedback to every message, as there was too much clicking. I would only rate the answer to my questions or the whole conversation at the end. Again, it really depends on the time it cost and how useful the answer is."*

2) A visual hierarchy for presenting system prompts. According to the participants, when reading the received prompts such as confirmations, answers, and requests of rating, the graphical difference is not evident. As a result, it becomes less efficient to visually track or quickly identify a particular message that the users are interested in:

"It could be much easier to read if the surface of the answer prompt was different with others."

3) Visual difference between system prompts and user actions. Some participants think that utilizing the same visual elements to present the prompts (e.g., confirmations) and the user actions (e.g., the answer selection) can lead to misunderstanding to operate the chatbot. For example, one participant cannot recognize the further-support chip which is placed within the system prompt:

"The further-support button looks like an option for Eike rather than me since it is in the message sent by Eike."

4) Guided rating in an emotive fashion. Some of the participants complain the rating icon is not understandable. They would like Eike to communicate through more clear instructions both visually and literally. Although they are satisfied with the emoji-based rating chips, which allows them to express their emotions and feelings vividly, they still prefer the emojis representing different attitudes can be more "emotive":

"To be honest, I do not know what the smiley icon (the rating icon) means. I thought it meant that Eike was smiling or helping."

"I liked the rating in a visually emotive way. However, it was a little bit hard to differentiate the facial expressions on those emojis (i.e., the rating chips)."

5) Data privacy. Some of them need to revisit their previous conversation and review the answer that they required. They consider this is a convenient move saving them time from re-experiencing the procedure of question answering. According to some other participants, however, the public usage scenarios such as public libraries or refugee centers give rise to worries about the privacy of their information. Exposing their problems to the next user of the chatbot service on the same computer can psychologically produce anxiety:

"I want to erase my chat history after closing the chat window. Because I may use this service through a computer in the library, and I do not want people to see what I am looking for."

"Sometimes I might ask for the same information several times, so I would prefer the chatbot could allow me to save the chat history."

6) Enabling effortless answer selections. More than half of the participants complained about choosing an answer from the list of candidates is not "user-friendly." This request, to some extends, makes no difference between a chatbot and a search engine. Scanning several answer candidates to the undefined query is "time-consuming" and "undelightful." To make matters worse, Eike cannot ensure the users to have an answer they preferred after reviewing all the candidates:

"I did not see a difference between the chatbot and the web retrieve if Eike asked me to go through a list of possible answers and pick one. Also, it was very time-consuming to read all the candidates."

"I was here to get the answer rather than make the question-answer pair."

7) Not human supports at the beginning. Although all the participants are satisfied with the failure handling where the users can be given the contact information of human supports, from the perspective of chatbot development, providing that information too soon may not be beneficial for Eike. Because the user would be more likely to reach out to the human for help instead of paraphrasing their questions to the less intelligent chatbot at that moment. Consequently, Eike has fewer opportunities to be involved in the training:

"I think giving the contact information (the further-support action) should really be the last thing. Otherwise, everyone will press the help button instead of rewriting their question. This will definitely flood the customer service of the HBG website."

8) The user-centered system prompts. Some participants dissatisfied with a prompt because they ponder it does not take the user's feelings, attitudes, and emotions into account. For example, when dealing with the unsuccessful question answering, the further-support chip, "Help I'm stuck!", is regarded as offensive to the users, as it causes the users to feel like they are attributed to the mistakes instead of Eike admitted guilt for failure:

> *"I do not like the phrase 'Help I'm stuck!', as it makes me seem stupid."*

Table 19. The result of the MAX evaluation

"How did you feel about it?" (emotion)		
	%	Overall
"Interested" (Intensity: 2)	5.3	"Interested" – 47.3%
"Interested" (Intensity: 3)	21.0	
"Interested" (Intensity: 4)	21.0	
"Satisfied" (Intensity: 2)	10.6	"Satisfied" – 47.4%
"Satisfied" (Intensity: 3)	15.8	
"Satisfied" (Intensity: 4)	21.0	
"Happy" (Intensity: 2)	5.3	"Happy" – 5.3%
"Was it easy to use?" (ease of use)		
	%	Overall
"It was easy to use" (Intensity: 2)	4.5	"It was easy to use" – 68.2%
"It was easy to use" (Intensity: 3)	13.6	
"It was easy to use" (Intensity: 4)	50.1	
"The use is intuitive" (Intensity: 3)	13.6	"The use is intuitive" – 31.8%
"The use is intuitive" (Intensity: 4)	18.2	
"Was it useful?" (usefulness)		
	%	Overall
"It is useful for me" (Intensity: 2)	12.5	"It is useful for me" – 50%
"It is useful for me" (Intensity: 3)	18.75	
"It is useful for me" (Intensity: 4)	18.75	
"It would help me" (Intensity: 2)	12.5	"It would help me" – 50%
"It would help me" (Intensity: 3)	25.0	
"It would help me" (Intensity: 4)	12.5	
"Do you wish to use it?" (intention to use)		
	%	Overall
"I liked to use it" (Intensity: 2)	13.3	"I liked to use it" – 66.6%
"I liked to use it" (Intensity: 3)	40.0	
"I liked to use it" (Intensity: 4)	13.3	
"I would use it frequently" (Intensity: 2)	6.8	"I would use it frequently" – 33,4%
"I would use it frequently" (Intensity: 3)	13.3	
"I would use it frequently" (Intensity: 4)	13.3	

Table 20. The result of the property checklist

Essential properties of the chatbot	Yes	May be	No	Comments/concerns
Performance				
The conversation is natural.	100%	–	–	"Eike spoke in a very humanlike way, and the dialogue with him was quite natural."
System prompts are written in simple and understandable language.	80.0%	20.0%	–	"It was written in pretty basic and simple English." "For some less-educated refugees, some words used in the answer might be difficult to understand."
When in the low confidence of question answering, the topic-related answer selection is helpful.	46.7%	53.3%	–	"I did not see a difference between the chatbot and the web retrieve if Eike asked me to go through a list of possible answers and pick one. Also, it was very time-consuming to read all the candidates." "I was here to get the answer rather than make the question-answer pair."
The error handling is satisfactory.	73.3%	26.7%	–	"It was good to have contact information that I could reach out to a human for help if Eike could not help me." "I think giving the contact information should really be the last thing. Otherwise, everyone will press the help button instead of rewriting their question. This definitely floods the customer service of the HBG website."
Features				
The rating of messages is useful.	46.7%	26.7%	26.6%	"I wouldn't give feedback to every message, as there was too much clicking." "I would only rate the answer to my questions or the whole conversation at the end. Again, it really depended on the time it cost and how useful the answer was."
The request for verifying the answer is acceptable.	100%	–	–	"I was more willing to verify the answer if Eike proactively asked me to do during the conversation."
Usability				
The following interface components are legible.				
System avatar	100%	–	–	"It was noticeable."
Rating icons	60.0%	40.0%	–	"I did not notice it at first."
Rating chips	100%	–	–	"They were pretty clear to see."
Messages	73.3%	26.7%	–	"Eike quickly sent couples of messages in a row, and it was hard to read when the messages are moving all the time." "It could be much easier to read if the surface of the answer was different from other prompts."
Further-support buttons	100%	–	–	"It was clear to see."
Send buttons	100%	–	–	"It was legible."
Close buttons	100%	–	–	"I could find it easily."

Table 20 (continued).

Essential properties of the chatbot	Yes	May be	No	Comments/concerns
The following interface components are understandable.				
System avatar	100%	–	–	"I knew I needed to click the avatar to start the conversation."
Rating icons	6.7%	20%	73.3%	"To be honest, I don't know what this smiley icon (the rating icon) means. I thought it meant that Eike was smiling or helping." "There could be a more clear instruction to the rating."
Rating chips	86.7%	13.3%	–	"The difference between the chips with different facial expressions was not obvious. I needed to look carefully to distinguish the different degrees behind the icons."
Messages	100%	–	–	"It was good to have a confirmation prompt before the answer, which reminded me the answer was coming."
Further-support buttons	73.3%	20%	6.7%	"I do not like the phrase 'Help I'm stuck!', as it makes me seem stupid." "The further-support button looked like an option for Eike rather than me since it is in the message sent by Eike."
Send buttons	100%	–	–	"Typical design, and it was easy to understand."
Close buttons	100%	–	–	"It is a classic close button in accord with my mental model."
Aesthetics				
The visual design is harmonious (adapted to the HBG website).	100%	–	–	"The chat interface is very elegant. It is colorful and well-designed."
Size				
The size of the chat window is appropriate	100%	–	–	"I liked the size and layout of the chat window on the website, as it was just big enough to show the messages, while avoided badly interrupting the view of the website."

3.3.3 Refining the solution – Final designs

This improvement of the chatbot solution took place after the analysis of the feedback received in the evaluation session. I first refined interface components based on the critical findings mentioned in section 3.3.2 Refining the Solution – Final Designs. After that, the high-level conversation flow was redefined, and the final mockups for the dialogue samples were rebuilt.

Visual hierarchy of different system prompts

To enable users to efficiently and quickly obtain information required, hierarchically, I presented information by applying color difference. The color indicates which element is interactive and the level of prominence. To create a contrast between the interface elements, the dark variant of the primary color, sapphire, is used on the answer prompt differentiating with the confirmation prompt in the primary color (see Figure 43). The secondary color, yellow, is utilized to draw attention on the call-to-action such as the request of rating and the rating chips.

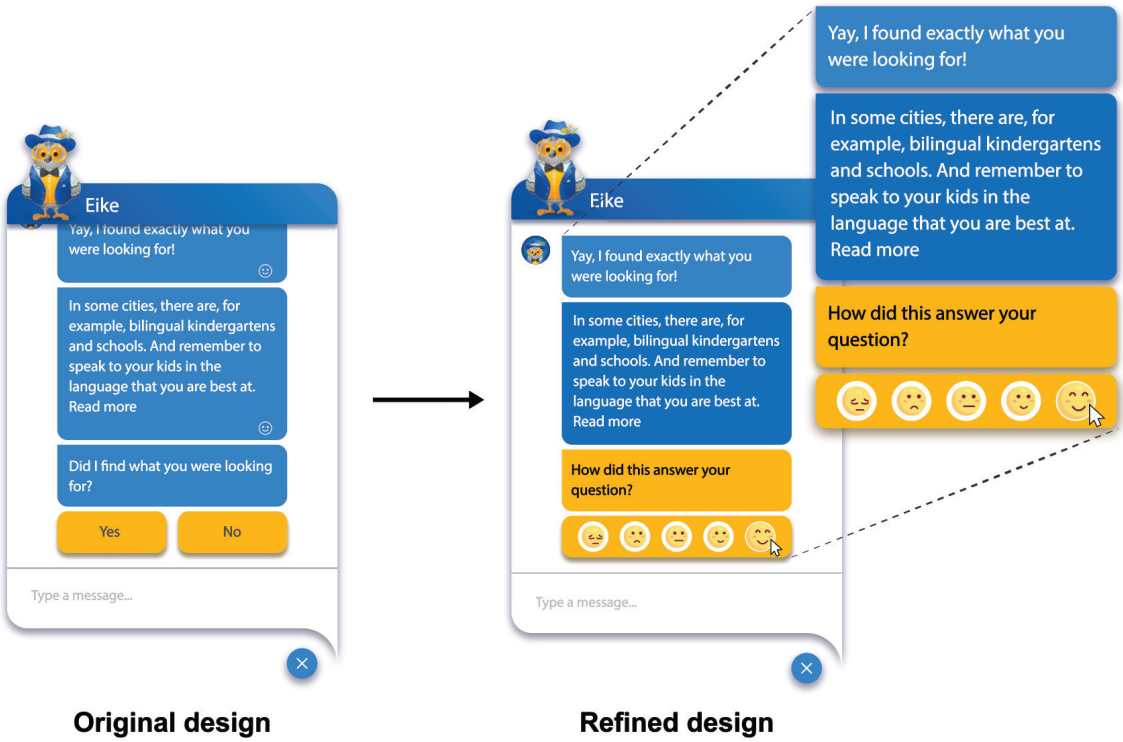


Figure 43. Color differences amongst regular prompts, answer prompts, and the request of rating answers

Visual Likert scales rating the given answer

Given the negative feedback towards surveying the performance of every prompt, it is removed from the chatbot design. However, verifying the question-answer pairs is acceptable to the users. As the quality of language is difficult to be quantified and gauged, using dichotomous scales (i.e., yes and no) cannot provide insight into the degree of response. To be specific, dichotomous scales can lead to inaccurate data of the respondent's satisfaction and feelings. Therefore, I decide to employ Smiley Face Likert scales as a rating scale for quantitative questions in evaluations (see Figure 44). A scale of faces – emojis –conveys emotions and attitudes in a simple and instantly recognizable format. In addition, it makes survey-taking experience less tedious and overcomes problems of survey fatigue.

Figure 44. Smiley Face Likert scales for rating the answer

Question selections when undefined user queries appear

Scanning and selecting an answer from a list of candidates is troublesome, which may cause the users to abandon the chatbot. To avoid that, I suggest to ask the users to confirm which topic-related question from the top three candidates was in line with their intention. These three queries correspond to the three answer candidate, respectively (see Figure 45). By doing this, the users are allowed to save more time and energy when using the chatbot service. Furthermore, it functions as a cover-up of failing to provide an appropriate answer to the users' query due to the lack of adequate intelligence. It can create an impression that Eike can intelligently understand and analyze their questions and then accordingly clarify the users' intention by proposing possible questions from the QA system for confirmation, rather than directly displaying keyword-relevant information such as a search engine. Figure 46 illustrates the graphical interface of showing the question candidates generated by the chatbot.

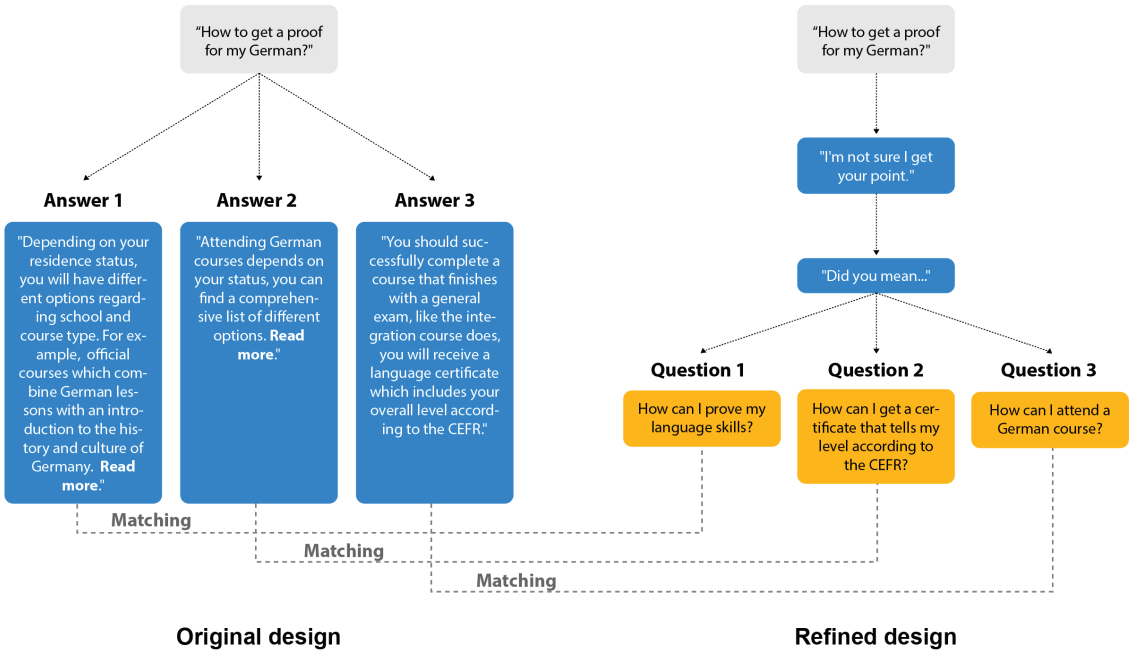


Figure 45. The logic and relation between the question selection and the answer selection in the low-confidence path

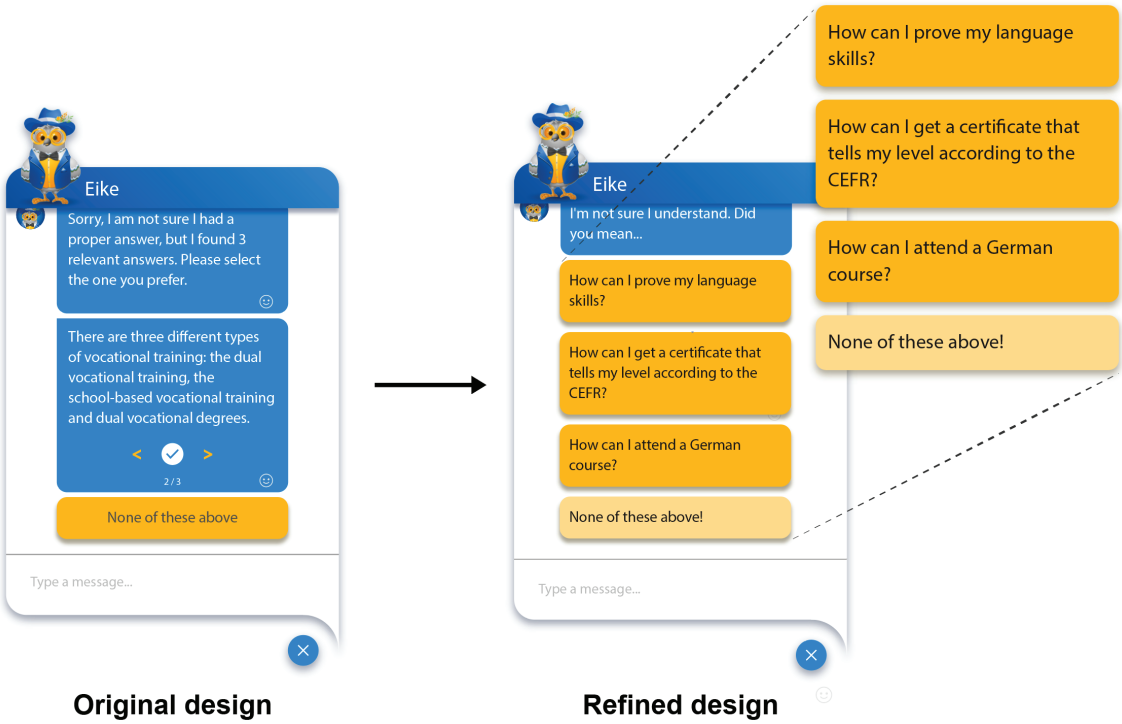


Figure 46. The graphical interface of showing the question candidates to the undefined query

Data privacy

To enable the users to revisit their chat history but also to protect their data privacy, Eike is designed to ask the confirmation of preserving the chat history for the users' next use (see Figure 47). Once the users click 'yes,' Eike will then have a double-checking by informing the risk of saving information on a public computer. This move ensures the users are not doing that by mistakes. After that, Eike says goodbye to the user to provide them with delightful exit experiences.

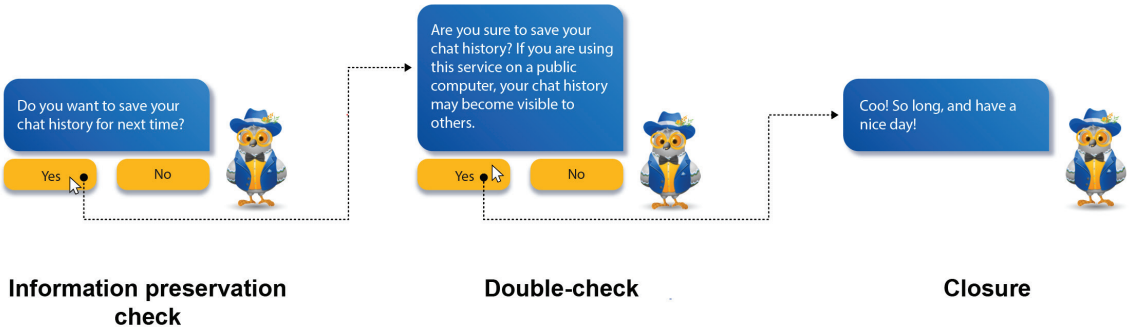


Figure 47. The process of information preservation checking

“Eike can’t help me!”

The further-support chip is separated from the system prompt to avoid the misunderstanding that it is an action for Eike rather than for the users (see Figure 48). Moreover, the prompt changed from “Help, I am stuck!” to “Eike can’t help me!”, which not only indicates the nature of a call-to-action button for the users but also admits the failure. However, the further-support chip will not be exposed to the users too soon. The users are required to paraphrase their query if their questions cannot be parsed by Eike correctly and they are not able to choose a preferred question which is accord with their intention. If Eike still fails to understand their rephrased question and provide them with a question representing their purposes, the users can access to the contact information of human supports by pressing the further-support chip, “Eike can’t help me!”.

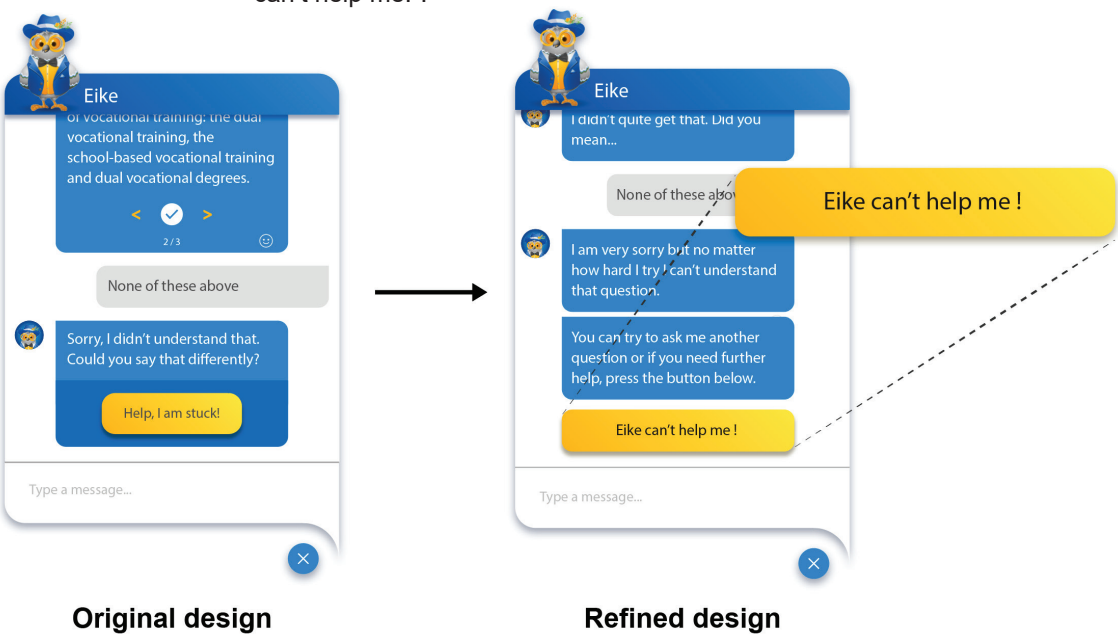


Figure 48. The refined further-support chip, “Eike can’t help me!”

The refined high-level conversation flow

I restructured the high-level conversation flow based on the improvement of interaction (see Figure 49). In the high confidence path, Eike responds with greetings, self-introduction, and suggestions of queries when the users initiate conversation. After the users receive a preferred answer to their questions, Eike asks them to evaluate the quality of the answer using visual Likert scales. Once the question answering completes, Eike seeks confirmations to

that whether the users want to continue another round of question answering. If the users confirm the dialogue is finished through utterances or the action of proactively closing the chat window, Eike will check whether they would like to save the chat history or not and then say goodbye to them. However, if Eike has troubles in parsing the user query, then the low confidence path occurs. In that case, Eike provides the users with the top three question candidates, which possibly match their intention. The users will be required to paraphrase their query if they cannot find a preferred question. When Eike comes to the second defeat of providing a question in line with the users' purpose, then the further-support chip is presented to them, by which the users obtain the contact information of human assistance.

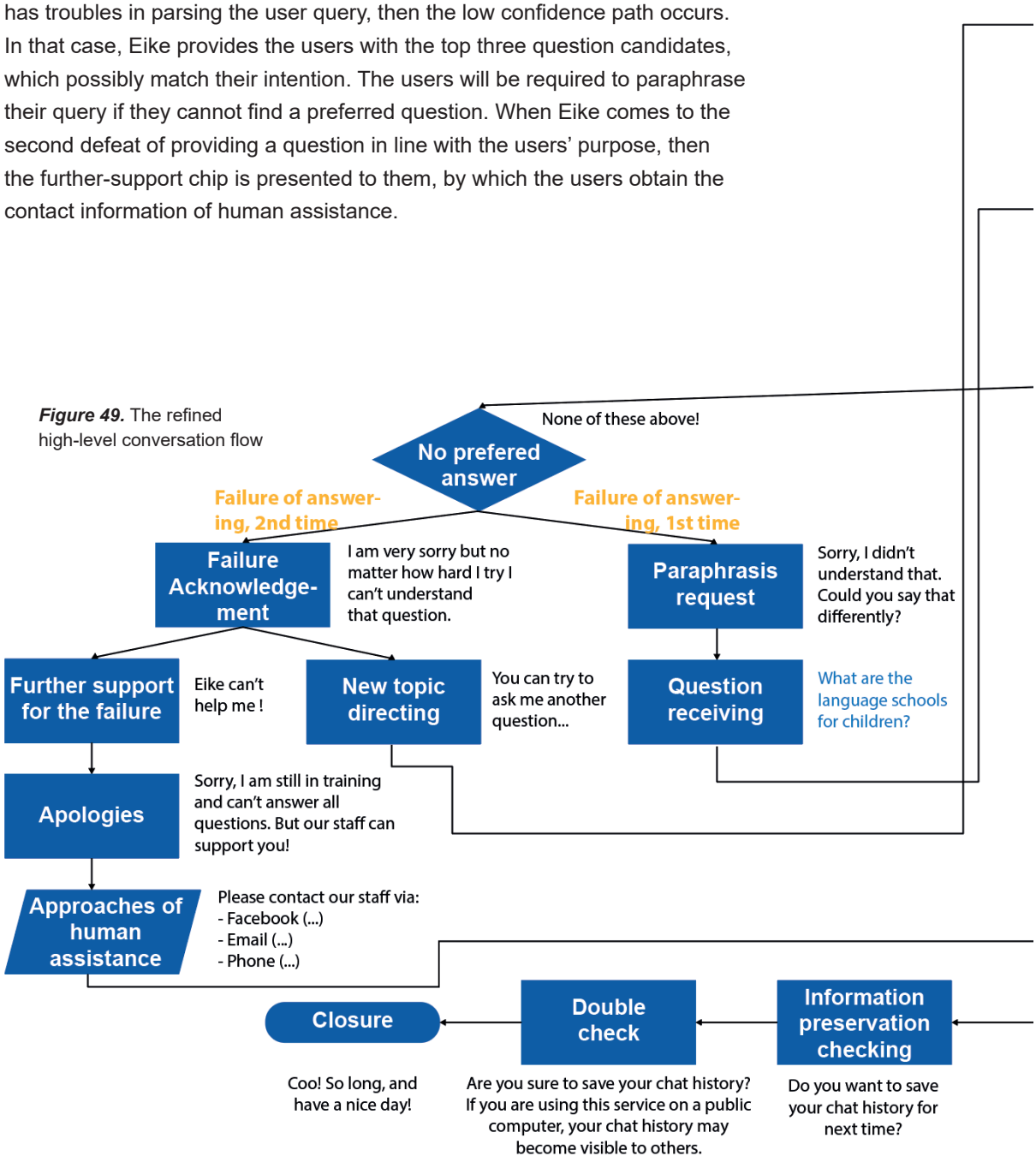
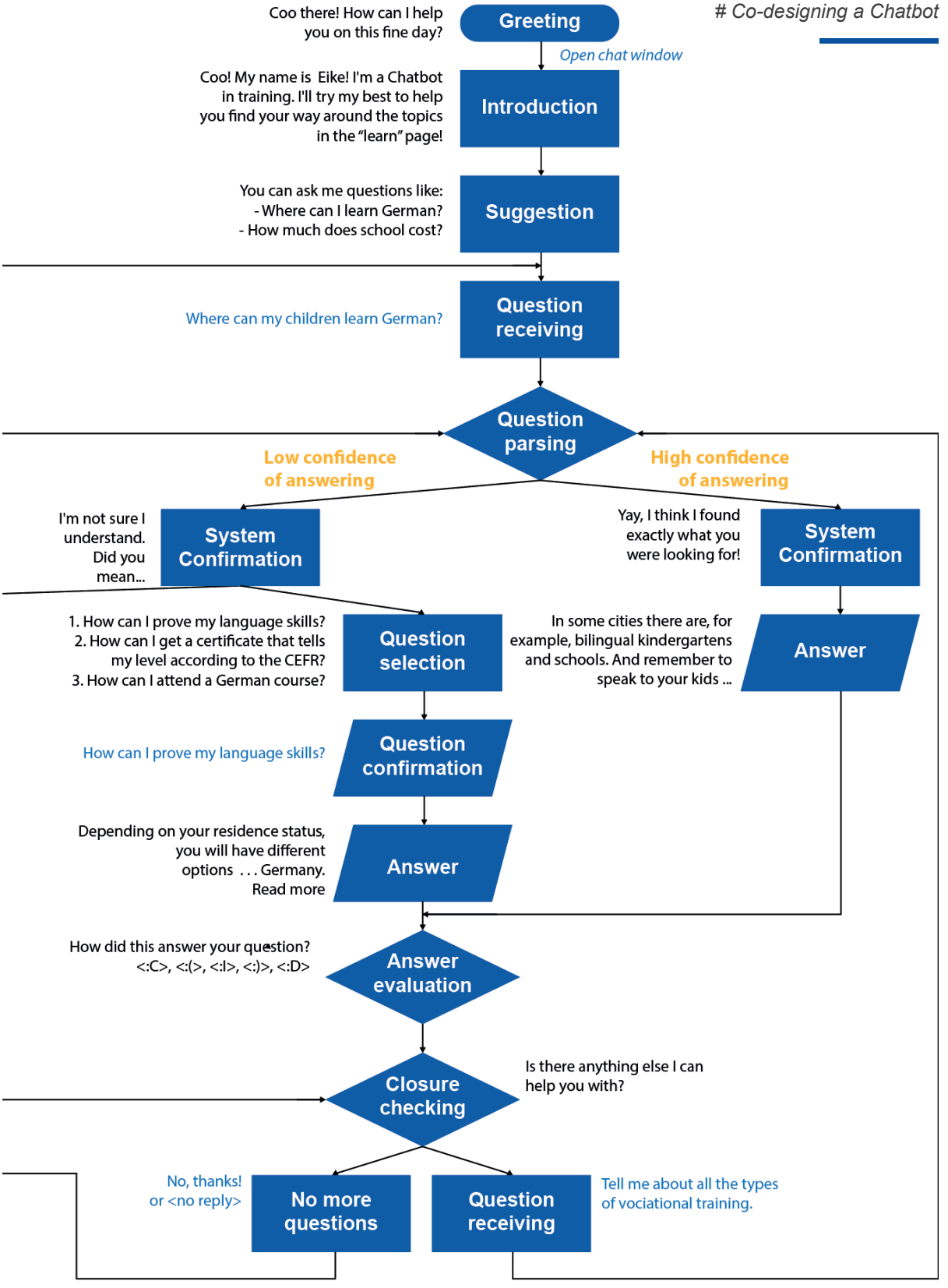


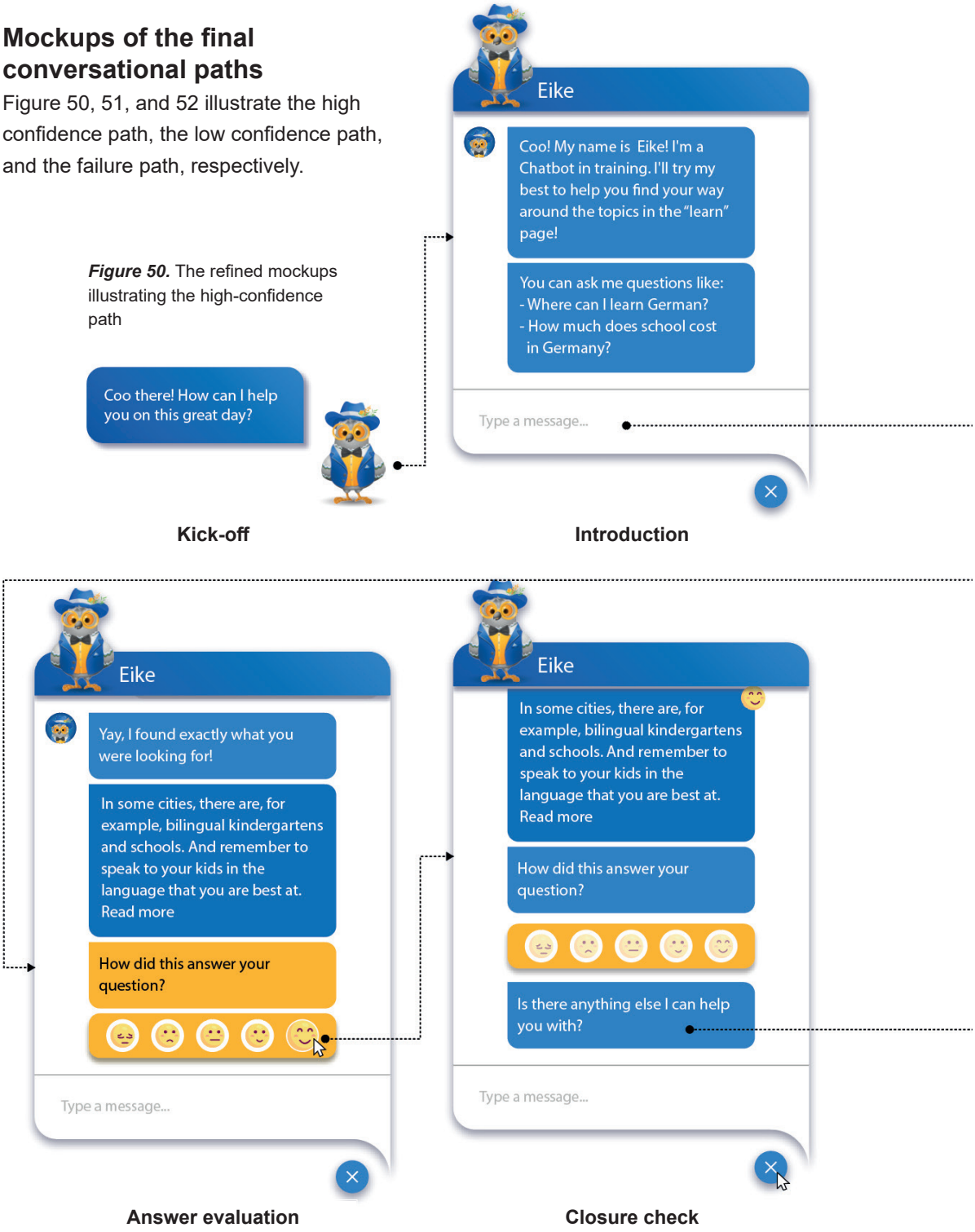
Figure 49. The refined high-level conversation flow

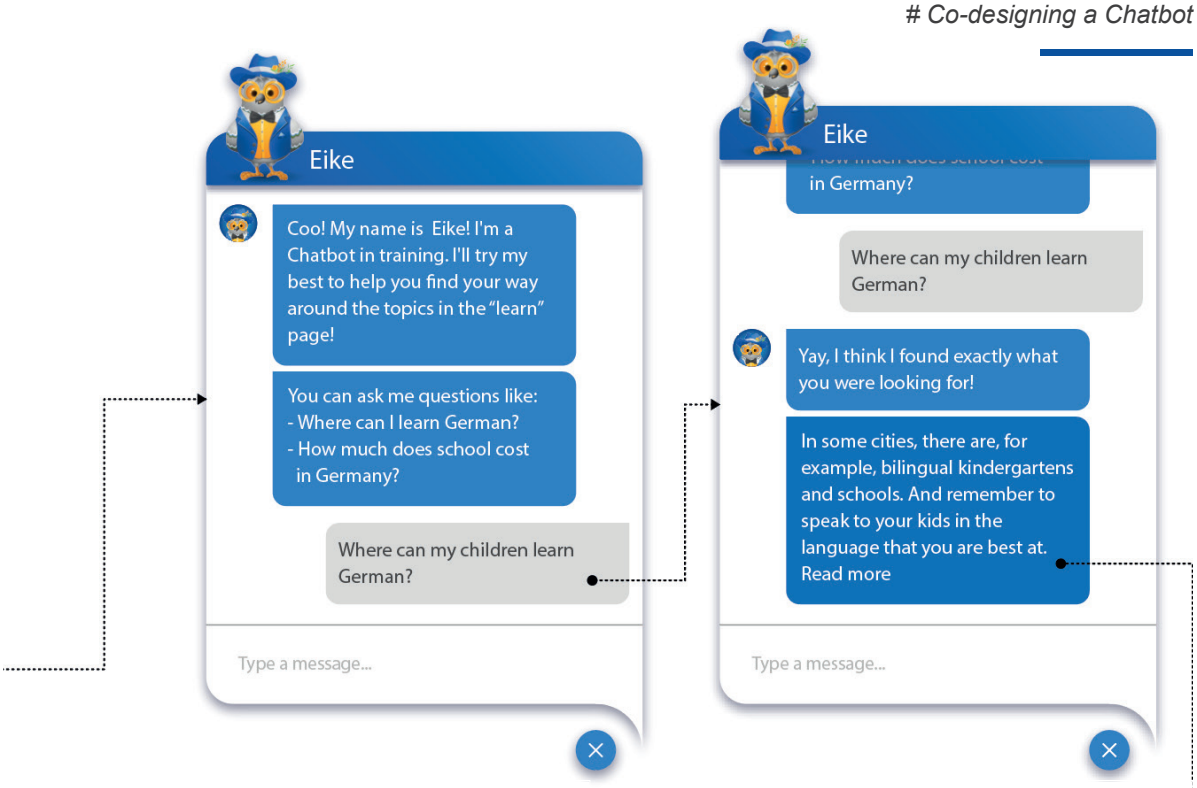

Coo there! How can I help you on this fine day?
Greeting
Open chat window
Coo! My name is Eike! I'm a Chatbot in training. I'll try my best to help you find your way around the topics in the "learn" page!
Introduction
You can ask me questions like:
- Where can I learn German?
- How much does school cost?
Suggestion
Where can my children learn German?
Question receiving
Question parsing
Low confidence of answering
High confidence of answering
I'm not sure I understand. Did you mean...
System Confirmation
Yay, I think I found exactly what you were looking for!
System Confirmation
1. How can I prove my language skills?
2. How can I get a certificate that tells my level according to the CEFR?
3. How can I attend a German course?
Question selection
In some cities there are, for example, bilingual kindergartens and schools. And remember to speak to your kids ...
Answer
How can I prove my language skills?
Question confirmation
Depending on your residence status, you will have different options . . . Germany. Read more
Answer
How did this answer your question? <:C>, <:(>, <:I>, <:)>, <:D>
Answer evaluation
Is there anything else I can help you with?
Closure checking
No, thanks! or <no reply>
No more questions
Question receiving
Tell me about all the types of vociational training.

Mockups of the final conversational paths

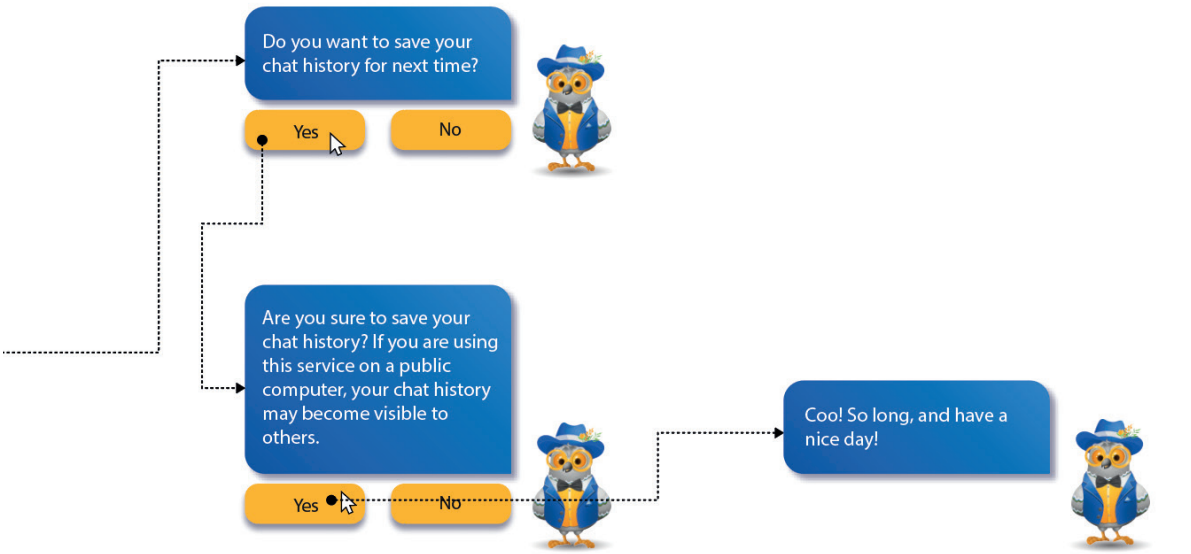
Figure 50, 51, and 52 illustrate the high confidence path, the low confidence path, and the failure path, respectively.

Figure 50. The refined mockups illustrating the high-confidence path

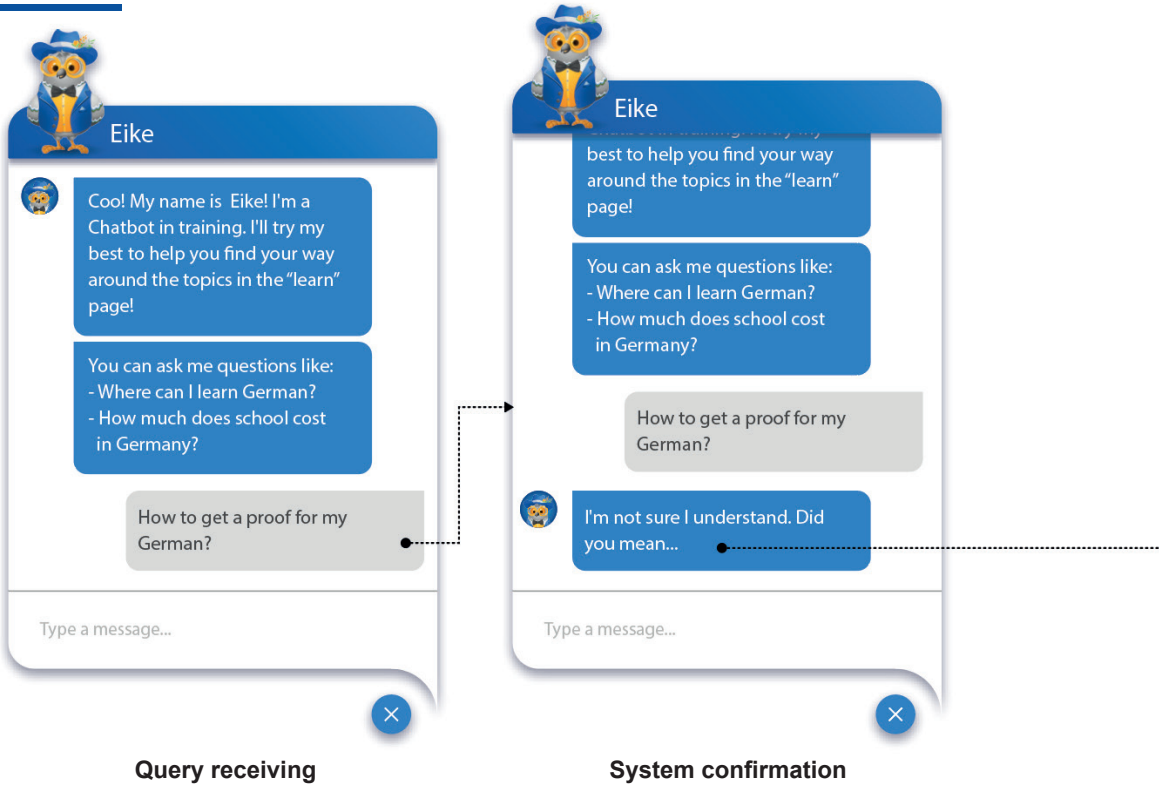




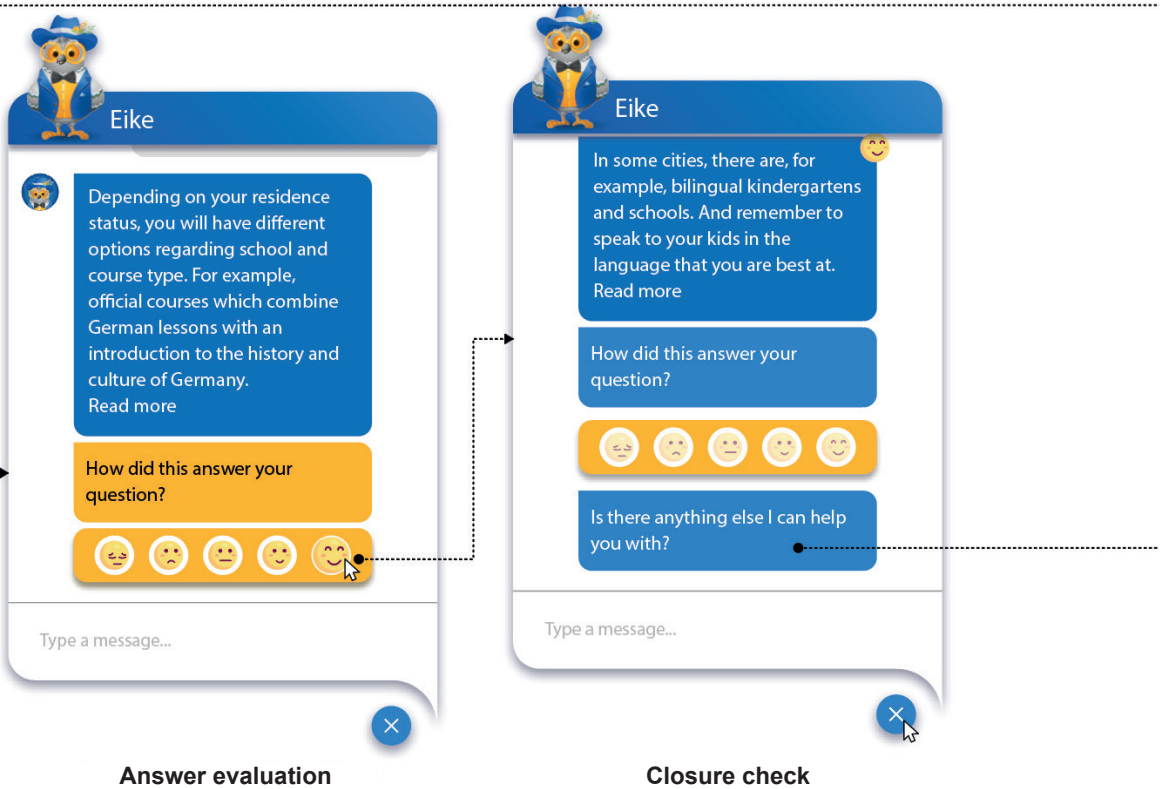
Query receiving **Confirmation & answer**



Information preservation check **Closure**



Query receiving **System confirmation**



Answer evaluation **Closure check**

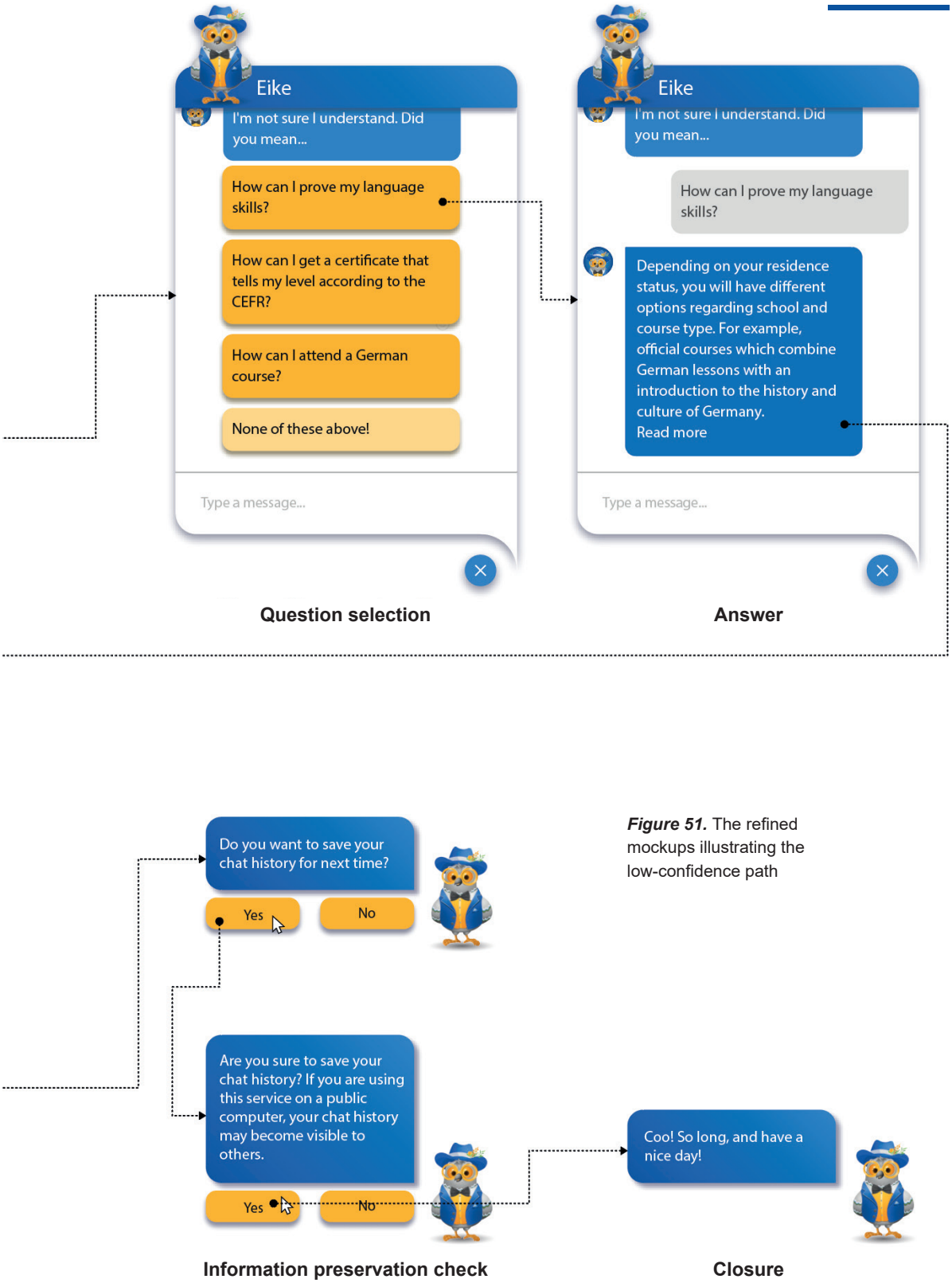


Figure 51. The refined mockups illustrating the low-confidence path

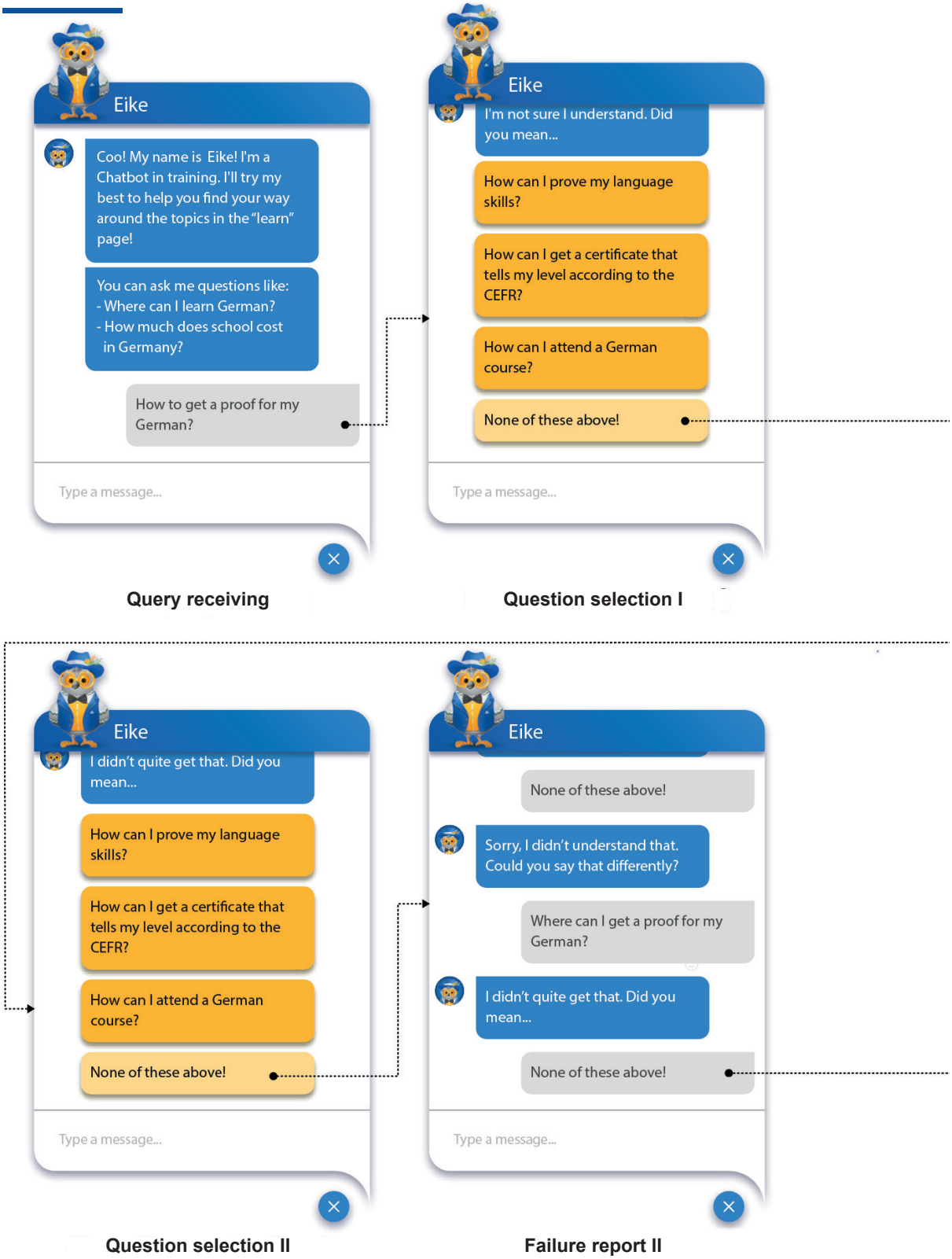
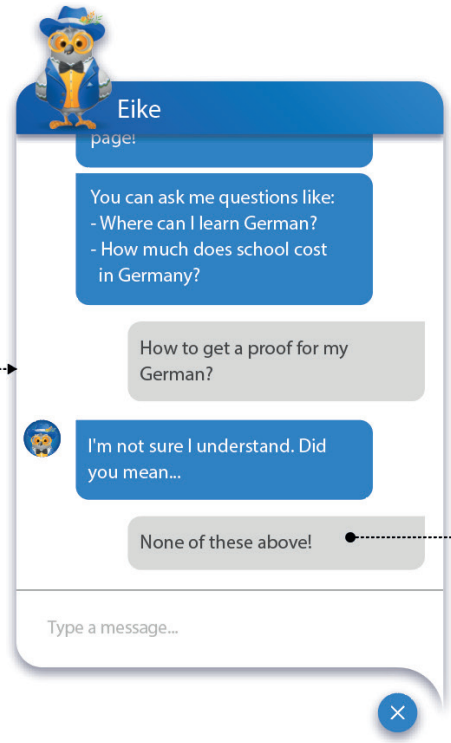
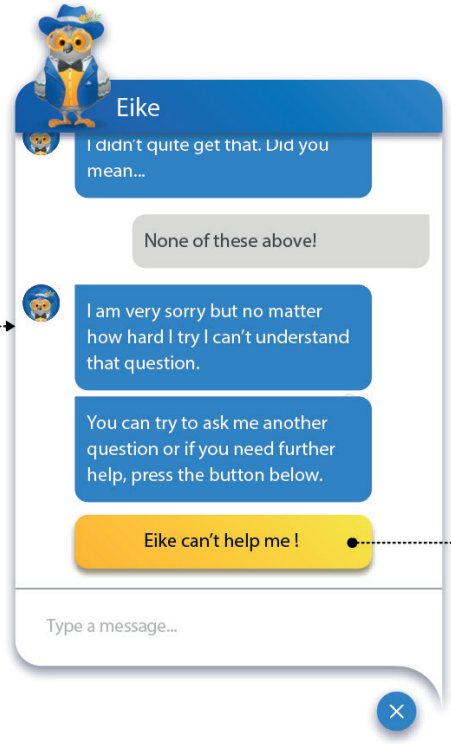


Figure 52. The refined mockups illustrating the failure path

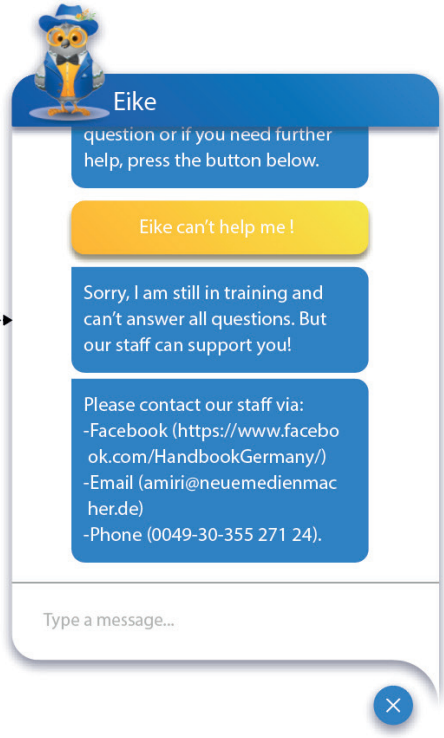


Failure report I

Paraphrasis request



Further support request



Further support

3.4 Deliver

3.4.1 Communication between Designers and Developers

The actual building of the chatbot occurs in the phase of implementation where I communicate the designs to the front-end developer writing program code to bring the software into existence. The collaboration frequently takes place in the form of discussing design specifications, such as impromptu or scheduled face-to-face meetings. The design specification is represented through accessible artifacts determining in much greater detail what is created. The final deliverable is a comprehensive design document – interface proxies – handing off to the developer. Interface proxies stand in for the chatbot and support reflection on the chat interface (e.g., visual elements and interaction patterns). Interface proxies contain a collection of clear and polished pictures with a description of details and annotation explaining the behaviors. After delivering those documents, I am in more of a support role, clarifying designs to the front-end developer as he implements Eike the chatbot.

3.4.2 Design Specifications

The chatbot is planned to be built through the two computer technologies, Hyper Text Markup Language 5 (HTML5)[17] and Cascading Style Sheets (CSS)[18]. HTML5 gives content structure and meaning by defining that content as headings, paragraphs, or images, whereas CSS is a presentation language created to style the appearance of material including layouts, fonts, and colors. As a rule, HTML5 always represents content, and CSS regularly renders the appearance of that content. In order to successfully facilitate the implementation, the final designed chatbot is required to be in line with the specification of HTML5 and CSS.

To explicitly communicate the designs of the chatbot, I create interface proxies to represent the chatbot properties: layout, shape, color, typography, and iconography (see Figure 53). A tool called *CSSmatic* is utilized to ensure the defined interface elements (e.g., borders, shadows, textures, and gradients) to adapt to the standards of CSS. Interaction is a dynamic procedure, which cannot be readily presented through static pictures. Consequently, except for the interface proxy, a flash animation about the process of interacting with Eike (e.g., how the state of interactive components changes) is produced to the help the developer understand the behaviors.

[17] Hyper Text Markup Language gives content structure and meaning by defining that content as, for example, headings, paragraphs, or images.

[18] Cascading Style Sheets is a presentation language created to style the appearance of content.

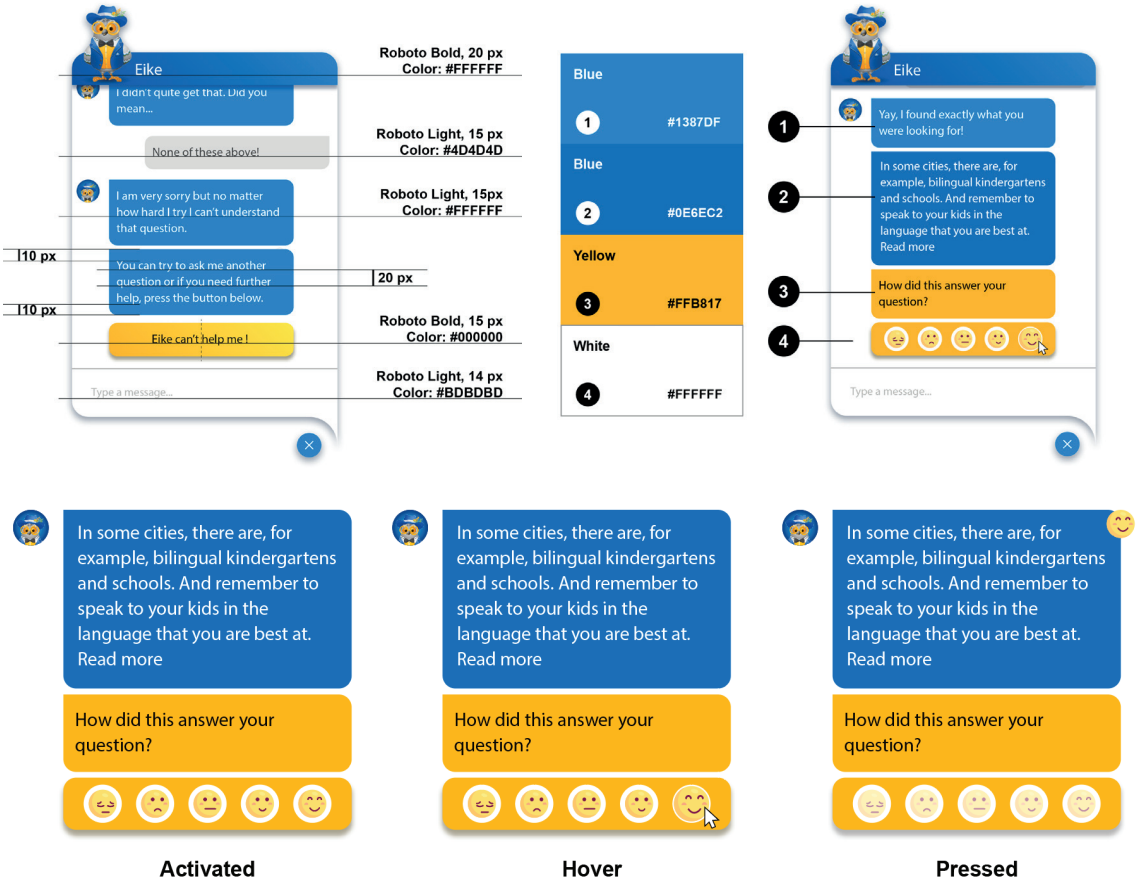


Figure 53. Examples of the interface proxies of the chatbot design
Typography (top-left); **Color scheme** (top-left); **Interaction behaviors of answer rating** (bottom).

4.
Discussion & Conclusion

4.1 Co-design Tools in Practice

With the support of this *research through design*, I found that probes (i.e., cultural probes and empathy probes) and generative toolkits to be engaging and supportive for collecting information from users and experts (i.e., refugees, migrants, and public service agents) and generating concepts by co-design participants (i.e., migrants and developers).

In a playful way, probes are effective in enhancing the informant's engagement and motivation as a proactive and capable informant. One informant explicitly reported later on that the task of reflecting on the conversation between him and refugees or migrants through the cultural probe was very playful. He preferred performing the task like that than filling out a traditional questionnaire.

Probes support refugees and migrants in expressing their thoughts, feelings, and opinions. In the session of exploring how the users feel about the pre-design chatbot prototype through empathy probes, sometimes the refugees and migrants having troubles in fluently using written English to convey their thoughts were able to use stickers (e.g., emojis) to represent their emotions and attitudes.

Probes are beneficial to inspire and evoke more qualitative feedback from the informants. For example, in the interview session of empathy probes, occasionally the informants have little to say towards a given topic. The illustrated cards produced by them then smoothly evoked their memories of using the pre-design chatbot prototype and aided them to provide valuable insights.

Furthermore, probes also provide the designer with an effective way to empathize with the informants in a short period. After going through the probing artifacts created by the refugees and migrants, I obtained an understanding of their experiences of using the pre-design chatbot prototype rapidly. However, the visual data without specific annotations, such as images, is difficult to be managed, analyzed, and defined. As a result, the interpretation of the ambiguous visual information by designers may be a

bit biased. In light of this, empathy probes can help avoid the designers' subjective judgments through a follow-up personal interview.

In the context of chatbot design, a generative toolkit is a right fit for concept development, as the co-design participants (e.g., users and developers) can build the content of the conversational user interface productively rather than merely generate inspiring artifacts for designers. Unlike other software development calling for specific subject domain knowledge, the critical expertise required for designing chatbots – knowledge about dialogues – can be easily accessed and learned by the users because they have practiced conversation all the time in their everyday life. Plus, the chatbot is going to be built with adapting to users' conversational behaviors. In the ERICS project, the generative toolkit provides easy-to-understand step-by-step guidance. The co-design participants without design backgrounds can simply produce concepts by leveraging their own expectations towards of the chatbot, experiences of using chatbots or seeking information in real life, and the provided tools. Although the provided tools can give the participants 'abilities' to design, they may, to some extent, limit the participants' imagination. For example, when building the chatbot avatar, all the participants directly chose an image to represent their ideal embodiment of avatars instead of drawing their own characters. As a result, the possibility of the visual representation of the chatbot avatar was still restricted within the pictures provided by the designer.

4.2 Co-design Participants in the ERICS Project

The central idea of co-design is to involve potential end users in the whole co-design process (Sanders, 2006). In the ERICS project, the users (i.e., refugees and migrants) are mainly involved in the first three phases of the co-design process: *pre-design*, *generative*, and *evaluative*. They facilitate the designer to explore and define the problem, ideate the design solution, and evaluate the prototype. Instead of playing a passive role as an object surveyed by the designer, they, as proactive 'informants' and 'designers' are empowered to express their opinions, generate ideas, and make design decisions with the support of co-design tools in different forms of collaboration such as reflection sessions on probes and co-design workshops. Despite that the designer is taking the lead in the early interviews, they still have enough room to proactively reflect on and explore their experiences of information seeking without the designer's intervention.

The end users of the chatbot are defined as refugees and migrants in Germany. Due to the geographical gap and the lack of resources for accessing them, inviting the comparable users (i.e., refugees and migrants) in Finland is considered as an alternative solution. Martínez-Solimán (2016) claims the refugees and migrants in Finland and Germany have the same problems and demands of social integration generally. Although the 'comparable' users cannot speak for all the 'actual' users, their experiences of migration and desires of obtaining information on everyday life have significant reference values when defining the problem and designing the chatbot solution aligning with the context of German refugees and migrants.

Nevertheless, the concept generation may have been affected by sampling bias of participants, as only three migrants representing users participated in the co-design workshop. No refugees took part in the session of ideation because of the unavailable contact and access to the refugees at that time, although they played a critical role in verifying the design proposals, and their thoughts, opinions, and suggestions directly influenced the final chatbot solution. All the participants (including users [migrants], the designer [the author] and the developers) have a higher educational background (postgraduates). But previous surveys on refugees' education, in Germany,

have concluded that their overall qualification structure is very heterogeneous ranging from school diplomas and vocational training to university degrees. Only twelve percent of refugees have higher educational degrees (Federal Office for Migration and Refugees, 2017a).

On the other hand, thanks to the co-design participants' higher educational background, they have better context availability to access German culture and an understanding of refugees and migrants' issues. It turns out that the refugees involved in the evaluation session have little knowledge about Germany culture and, culturally, they cannot comment on the proposed concept. In light of this, I suggest ensuring the comparable users (co-design participants) having experience or knowledge of the design context and user culture when the real users cannot be reached.

Furthermore, I discover involving different users or stakeholders in different co-design sessions can contribute to reflecting, evoking, and generating new ideas. For example, in the phases of say and do, I investigated users' expectation towards the chatbot by applying questionnaires and participant observation with different refugees and migrants. After that, the findings were fed into the co-design workshops, which provoked lively discussion among the participants and inspired them to generate more ideas. The concept became holistic when all the people's voices were taken into consideration. Besides, keeping the consistency of participants in a session can efficiently promote the success of that co-design event. Due to the complexity of the chatbot, the design goals cannot be easily achieved just through one single co-design workshop. Therefore, several co-design workshops are organized. Through the consecutive co-design workshops where the participants remain unchanged, it shows that the participants can quickly get into the position where they know what and how to do without familiarizing themselves with the background and the context of chatbots again. Moreover, they possess a mutual understanding and know how to collaborate with each other.

Revisiting the UCD Competency Model (Nieminen, 2015), I spot that co-design participants (i.e., users and developers) can obtain 'design

competencies' by leveraging co-design tools integrating a designerly way of problem-solving (e.g., conceptualization, visualization, and validation) and subject domain knowledge. In a combination of the users' strengths (i.e., subject domain experience, context availability, and user cultures) and the developers' technological knowledge about the product, the design proposals generated by co-design participants can be very mature, directly forming a foundation for the final concept. For example, an intricate design process of chatbot avatar and its knowledge can be deconstructed, modularized, and represented in a generative toolkit. By using the step-by-step instructional kit, the co-design participants can produce design concepts like a trained designer. In the co-design workshops, the users bring ideas based on their experience and demands, and the developers provide suggestions making sure the design concepts fall into the scope of technological feasibility. Through this collaboration, the resulting design concepts tend to be holistic and mature rather than being merely inspiring for the designers.

4.3 Answering to Research Questions

This section answers the research question in this thesis:

> ***How can a co-design approach contribute to designing a chatbot supporting social integration within the context of refugees and migrants?***

Defining the problem to shape design questions. Due to the complexity and heterogeneity of refugees and migrants, exploring and defining their problems and demands can be difficult. The co-design approach allows the users and other stakeholders to take part in the design process and add their voice to the goals of chatbot design. By employing co-design methods and tools to listen to what the refugees, migrants, and related experts say about social integration, the user needs are revealed. At the meantime, the user behavior is discovered by observing their interaction with the chatbot. All these sessions play an essential role in helping to identify the problem. Based on the problems, design questions are then formed to explicitly indicate design direction, which significantly contributes to the promotion of the co-design process.

Developing mature concepts to lay a foundation for the final solution. In the ERICS project, the co-design approach helps to yield well-developed ideas. Co-design happens at different stages with people from diverse countries, cultures, and backgrounds. One of the co-design tools applied in the ERICS project is the generative toolkit (Sanders & Stappers, 2014), by which the users (migrants) work with the designer and the developers to produce design concepts in co-design workshops. Different competencies brought by different co-design actors are beneficial to form holistic and mature initial concepts directly shaping a foundation of the final design solution. The initial concepts are reasoned and useful because they integrate and leverage the users' strength (subject domain experience [i.e., user experience], context availability, and user culture), the developers' knowledge of available technologies, and continues validation and improvement.

Refining solutions through co-design tools. In the ERICS project, generative

toolkits are extensively used by co-design participants to build design artifacts expressing their experience, ideas, and proposed chatbot concepts for the chatbot. In a tangible and visualized way, those resulting artifacts allow the participants to reflect and yield feedback and further refinement. For example, in the first co-design workshop, the participants utilize toolkits to create a visualized storyboard to share the ideas of the chatbot avatar, which effectively prompts discussion and reflection. They can review the concrete design ideas and then share their opinions and suggestions — the whole process of reflection functions as a quick iteration of design concepts. Later on, the refined artifacts demonstrating the possible candidates of chatbot avatars work as prototypes to be evaluated by more refugees.

4.4 Limitations & Future Research

In the ERICS project, the development team consists of the frontend developer who collaborates with the designer in the co-design process in Finland and the backend developers in Germany who build the database and are in charge of integration and implementation. Implementing the chatbot falls outside of the scope of my responsibilities. After the design is delivered, the deployment of the chatbot changed from answering questions under the “Learn” section on the HBG website to responding to queries regarding all the topics in the site.

The interaction between the designer and the frontend developer mainly occurs in the later phases of co-design where they discuss the availability of the designed interface components and design specifications based on the created artifacts such as interface proxies. Although the frontend developer closely works with the design team and takes part in the co-design activities, there is still some moment of detachment. For example, as the developer is not involved in the user research, sometimes he raises doubts about a particular defined user requirement or function, and then an explanation is required to resolve the puzzles. Unfortunately, the backend developers are entirely excluded in the co-design process because of the geographic gap. As a result, this noncooperation slightly leads to the inconsistency between the final product and the ideal design designs in terms of visual design and conversation flows. According to the frontend developer, this gap between the launched chatbot and the proposed solution took place because there was not a shared version control where the frontend and the backend could cooperate and build the chatbot as a whole. Instead, the backend developers copied and pasted the codes created by the frontend programmer without jointly walking through and examining the codes. As for the removed prompts, the backend developers, as the final decision maker in the ERICS project, considered them unnecessary. This one-sided understanding is attributed to the backend developers’ disengagement in the co-design process.

Based on the reflection on the ERICS project, I propose an assumption of a co-design process for future research, in which users, designers, and developers jointly create a user-centered shippable solution of software

products (e.g., chatbots) (see Figure 54). This co-design process incorporates four phases: *pre-design*, *generative*, *evaluative* and *post-design*. In the stage of *pre-design*, by listening to what users say and observing how users interact with the product in their places, designers jointly explore user experience and define the problem with them. After that, designers support developers to understand the studied user experience and the issues identified by presenting research findings. In the second phase, *generative*, all the co-design participants (e.g., users, designers, and developers) meet together in workshop-like activities to generate the possible ideal solutions to the defining problem with taking business requirements and technological capabilities into consideration. After the early concept is formed, with developers' supports, designers build prototypes in line with technical feasibility. Subsequently, users are invited to experience and evaluate the prototype and then collaboratively refine the design solution with designers. Finally, designers facilitate implementation by providing design specifications. Although the process is illustrated linearly, it is encouraged to be iterative in practice. During the whole process, designers function a bridge connecting users and the development team. The users are encouraged to be involved in the development as late as possible, whereas the developers are expected to participate in the research and co-design as early as possible.

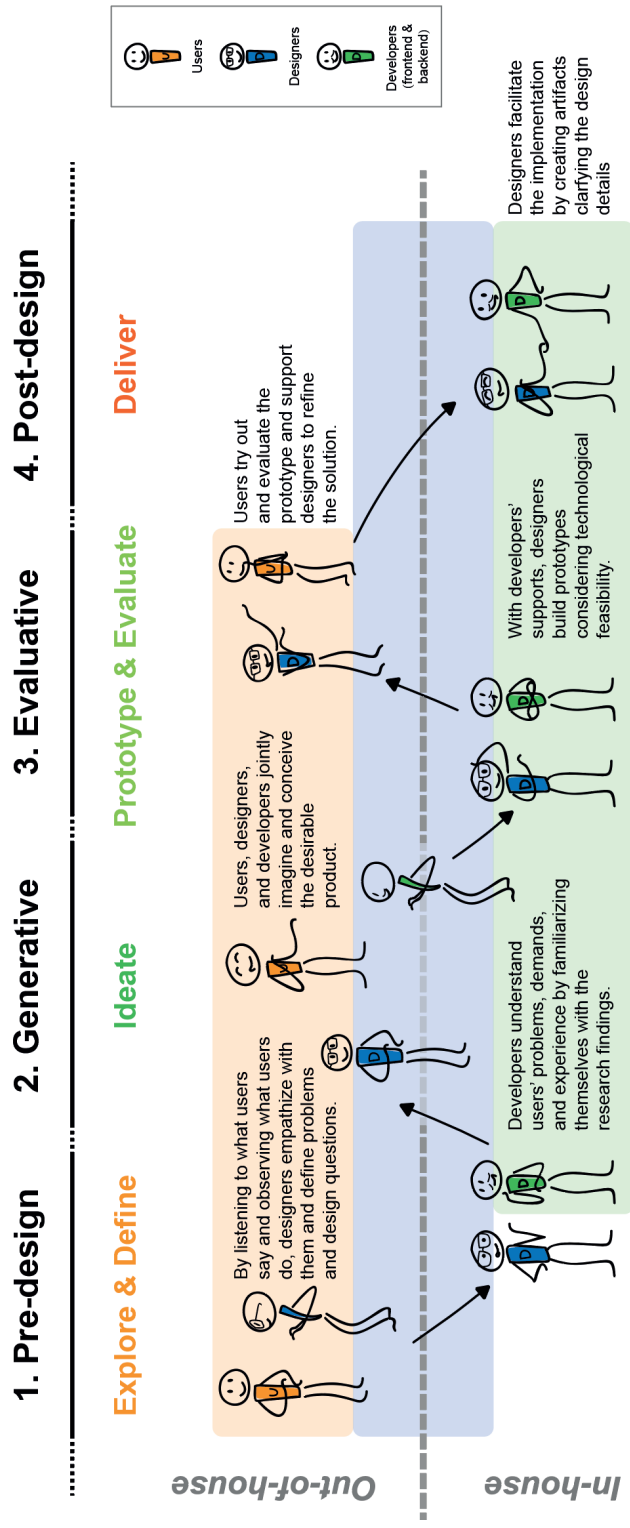


Figure 54. Co-design processes in chatbot development where users, designers, and developers jointly create a user-centered shippable design solution

4.5 Implications

This thesis provides an empirical case of co-design practice in the context of refugees and migrants, which contributes to a body of research from a variety of perspectives incorporating co-design, user interface design, and conversation design. In addition to adding the theoretical and empirical knowledge and perspectives connecting co-design approach with chatbot development, this thesis can benefit companies and research institutes that require references for designing a chatbot through co-design approach. More specifically, within the complex context of refugees and migrants, this thesis investigates the co-design methods and tools are beneficial for defining problems as well as generating and refining mature design concepts contributing to the final solution in the field of chatbot design. Based on *say, do, and make* (Sanders, 2003), *framework of co-design processes* (Sanders & Stappers, 2014), *joint inquiry and imagination* (Steen, 2013) and the empirical case study, this thesis proposes a co-design process in software (chatbot) development, in which users, designers, and developers can build strong engagement and shared understanding so as to ensure the success of product development.

4.6 Conclusion

For my *research through design* process, I investigated how a co-design approach can contribute to designing a chatbot which supports social integration within the context of refugees and migrants. In order to fulfill this goal, in-depth document studies, user questionnaires, and cultural-probe-aided expert interviews were conducted to grasp an understanding of user experience, needs, and expectations. Meanwhile, participant observation and empathy probes were applied to unveil user behaviors. After that, the problems were defined, which led to the generation of design questions guiding the co-design. In order to solve the design questions, two co-design workshops were implemented using generative toolkits. The resulting design concepts were later on evaluated by the users. After reflecting on the whole co-design process, I concluded that the co-design approach was beneficial for defining problems in the complex context of refugees and migrants, developing mature concepts laying a foundation for the final solution, and refining the solution efficiently. Although the previous research indicated the efficacy of utilization of co-design when designing services for refugees and migrant, my thesis fills the gap between the co-design approach and designing chatbots for refugees and migrants. However, the disengagement of developers (especially the backend developers) in the co-design process led to the inconsistency between the final product and the proposed design concept. Therefore, future researchers should consider investigating the impact of collaboration and communication among users, designers, and developers throughout the whole design process of software (chatbot) development.

5.
References & Appendices

References

Angga, P. A., Fachri, W. E., Elevanita, A., & Agushinta, R. D. (2015, October). Design of chatbot with 3D avatar, voice interface, and facial expression. In Science in Information Technology (ICSITech), 2015 International Conference on (pp. 326-330). IEEE.

Arce, O. (2004). Participatory design, challenges and experieneces using design in development. Design without borders–Experiences from incorporating industrial design into projects for development and humanitarian aid, 45-50.

Archer, B. (1995). The nature of research. Co-Design Journal, 2(11), 6-13.

Assembly, U. N. (1951). General. Universal declaration of human rights." Resolution adopted by the General Assembly, Treaty Series, vol. 189, p. 152.

BAMF, 2017a, IAB-BAMF-SOEP-Befragung von Geflüchteten 2016: Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen, Forschungsbericht 30

Bevan, N. (2009, July). Extending quality in use to provide a framework for usability measurement. In International Conference on Human Centered Design (pp. 13-22). Springer, Berlin, Heidelberg.

Brandt, E. & Agger Eriksen, M. (2010) Co-design events. In Halse, J., Brandt, E., Clark, B., Binder, T. (eds) Rehearsing the future. The Danish Design School Press. Denmark, 19.

Buxton, W. (2005). Innovation vs. invention. Rotman magazine, 2, 52-53.

Cahn, J. (2017). CHATBOT: Architecture, design, and development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment (Vol. 17). Sage publications.

Casey, P. (2018, February). 6 Chatbot UX Design 'Must-haves' for 2018. Retrieved from https://chatbotsmagazine.com/6-must-haves-for-chatbot-ux-design-cc2a21b944e

Cavalcante, E., Rivero, L., & Conte, T. (2015). Evaluating the feasibility of max: a method using cards and a board for assessing the post-use UX. International Journal of Software Engineering and Knowledge Engineering, 25(09n10), 1759-1764.

Conversational components. (n.d.). June 2, 2018, Retrieved from https://designguidelines.withgoogle.com/conversation/conversational-components/overview.html

Conversation design process. (n.d.). June 2, 2018, Retrieved from https://designguidelines.withgoogle.com/conversation/conversation-design-process/how-do-i-get-started.html

Cook, M. (2016). Personnel selection: Adding value through people-A changing picture. John Wiley & Sons.

Creat a persona. (n.d.). June 2, 2018, Retrieved from https://designguidelines.withgoogle.com/conversation/conversation-design-process/create-a-persona.html

Daiute, C. (2010). Critical narrating by adolescents growing up in war: Case study across the former Yugoslavia. In Narrative development in adolescence (pp. 207-230). Springer, Boston, MA.

Di Gaetano, S., & Diliberto, P. (2018). Chatbots and Conversational Interfaces: Three Domains of Use.

Donetto, S., Pierri, P., Tsianakas, V., & Robert, G. (2015). Experience-based co-design and healthcare improvement: realizing participatory design in the public sector. The Design Journal, 18(2), 227-248.

Eppinger, S. D., & Ulrich, K. T. (1995). Product design and development.

Fisher, K. E., Yefimova, K., & Bishop, A. P. (2016, May). Adapting design thinking and cultural probes to the experiences of immigrant youth: Uncovering the roles of visual media and music in ICT wayfaring. In Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems (pp. 859-871). ACM.

Gaver, B., Dunne, T., & Pacenti, E. (1999). Design: cultural probes. interactions, 6(1), 21-29.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse education today, 24(2), 105-112.

Hartson, R., & Pyla, P. S. (2012). The UX Book: Process and guidelines for ensuring a quality user experience. Elsevier.

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products–Facets of user experience. Interacting with computers, 22(5), 353-362.

Holmes, S. M., & Castañeda, H. (2016). Representing the “European refugee crisis” in Germany and beyond: Deservingness and difference, life and death. American Ethnologist, 43(1), 12-24.

Huang, T. H. K., Lasecki, W. S., Azaria, A., & Bigham, J. P. (2016, September). “Is There Anything Else I Can Help You With?” Challenges in Deploying an On-Demand Crowd-Powered Conversational Agent. In Fourth AAAI Conference on Human Computation and Crowdsourcing.

Huang, J., Zhou, M., & Yang, D. (2007, January). Extracting Chatbot Knowledge from Online Discussion Forums. In IJCAI (Vol. 7, pp. 423-428).

Hussain, S., Sanders, E. B. N., & Steinert, M. (2012). Participatory design with marginalized people in developing countries: Challenges and opportunities experienced in a field study in Cambodia. International Journal of Design, 6(2).

ISO, W. (2009). 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO). Switzerland.

ISO, (1998). 9241-11: Ergonomic requirements for office work with visual display terminals (UDTs) – Part 11: Guidance on usability. Germany.

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018, June). Evaluating and Informing the Design of Chatbots. In Proceedings of the 2018 on Designing Inter active Systems Conference 2018 (pp. 895-906). ACM.

Jan, F. M., (2017, October). Tinka – the clever chatbot of T-Mobile Austria. June 12, 2018, Retrieved from https://www.b2b-europe.telekom.com/blog/2017/10/25/tinka-the-clever-chatbot-of-t-mobile-austria

Jordan, P. W. (1998). Human factors for pleasure in product use. Applied ergonomics, 29(1), 25-33.

Kaasinen, E., Koskela-Huotari, K., Ikonen, V., & Niemeléi, M. (2013). Three approaches to co-creating services with users. Advances in the Human Side of Service Engineering, 286.

Kelman, H. C., & Hovland, C. I. (1953). " Reinstatement" of the communicator in delayed measurement of opinion change. The Journal of Abnormal and Social Psychology, 48(3), 327.

Kirmayer, L. J., Narasiah, L., Munoz, M., Rashid, M., Ryder, A. G., Guzder, J., ... & Pottie, K. (2011). Common mental health problems in immigrants and refugees: general approach in primary care. Cmaj, 183(12), E959-E967.

Kleinsmann, M (2006) Understanding collaborative design, Ph.D. thesis, Delft University of Technology, Delft, 30.

Kleinsmann, M., & Valkenburg, R. (2008). Barriers and enablers for creating shared understanding in co-design projects. Design studies, 29(4), 369-386.

Kocaballi, A. B, Laranjo, L, & Coiera, E. (2018). Measuring User Experience in Conversational Interfaces: A Comparison of Six Questionnaires.

Kuligowska, K. (2015). Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents.

Language. (n.d.). June 12, 2018, Retrieved from https://designguidelines.withgoogle.com/conversation/conversational-components/overview.html#overview-examples

Learn about conversation. (n.d.). June 12, 2018, Retrieved from https://designguidelines.withgoogle.com/conversation/conversation-design/learn-about-conversation.html#

Lucero, A., Lashina, T., & Diederiks, E. (2004, November). From imagination to experience: the role of feasibility studies in gathering requirements for ambient intelligent products. In European Symposium on Ambient Intelligence (pp. 92-99). Springer, Berlin, Heidelberg.

Lucero, A. (2009). Co-designing interactive spaces for and with designers: supporting mood-board making. Vaajakoski, Finland. Gummerus Printing.

Lucero, A. (2015, September). Using affinity diagrams to evaluate interactive prototypes. In IFIP Conference on Human-Computer Interaction (pp. 231-248). Springer, Cham.

Lucero, A., Aliakseyeu, D., & Martens, J. B. (2007, October). Augmenting mood boards: flexible and intuitive interaction in the context of the design studio. In Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07) (pp. 147-154). IEEE.

Mammarella, N., Di Domenico, A., Palumbo, R., & Fairfield, B. (2016). When green is positive and red is negative: aging and the influence of color on emotional memories. Psychology and aging, 31(8), 914.

Martínez-Solimán M. (2016, September 18). Migrants and refugees: A global problem or a local solution?. June 17, 2018, Retrieved from http://www.undp.org/content/undp/en/home/blog/2016/9/18/Migrants-and-refugees-A-global-problem-or-a-local-solution-.html

Mattelmäki, T. (2005). Applying probes–from inspirational notes to collaborative insights. CoDesign, 1(2), 83-102.

Mattelmäki, T., & Battarbee, K. (2002, January). Empathy probes. In PDC (pp. 266-271).

Mattelmäki, T., & Sleeswijk Visser, F. (2011, October). Lost in Co-X: Interpretations of Co-design and Co-creation. In Proceedings of the 4th Conference on Design Research (pp. 1-12).

McTear, M., Callejas, Z., & Griol, D. (2016). Dialog Management. In The Conversational Interface (pp. 209-233). Springer, Cham.

Migrant/Migration. (n.d.). June 17, 2018, Retrieved from http://www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/migrant/

Miles, R. S., Greensmith, J., Schnadelbach, H., & Garibaldi, J. M. (2013, September). Towards a method of identifying the causes of poor user experience on websites. In Computational Intelligence (UKCI), 2013 13th UK Workshop on (pp. 258-265). IEEE.

Moreno, A. M., Seffah, A., Capilla, R., & Sanchez-Segura, M. I. (2013). HCI practices for building usable software. Computer, 46(4), 100-102.

Musante, K., & DeWalt, B. R. (2010). Participant observation: A guide for fieldworkers. Rowman Altamira, (pp. 1-16).

Nandan, S. (2005). An exploration of the brand identity–brand image linkage: A communications perspective. Journal of brand management, 12(4), 264-278.

Nan, X., Anghelcev, G., Myers, J. R., Sar, S., & Faber, R. (2006). What if a web site can talk? Exploring the persuasive effects of web-based anthropomorphic agents. Journalism & Mass Communication Quarterly, 83(3), 615-631.

Nieminen, M. P. (2015). User-Centered Design Competencies-Construction of a Competency Model.

Pelachaud, C. (2000). Some considerations about embodied agents. In Int. Conf. on Autonomous Agents, Barcelona.

Pumariega, A. J., Rothe, E., & Pumariega, J. B. (2005). Mental health of immigrants and refugees. Community mental health journal, 41(5), 581-597.

Quarteroni, S., & Manandhar, S. (2007). A chatbot-based interactive question answering system. Decalog 2007, 83.

Reeves, B., & Nass, C. I. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge university press.

Russo, T. C. (2001). Operationalizing mediated presence: Initial steps toward a measure of the construct. In 4th International Workshop on Presence, Phil adelphia, PA.

Sanders, E. B. N. (2000). Generative tools for co-designing. In Collaborative design (pp. 3-12). Springer, London.

Sanders, E. B. N. (2003). From user-centered to participatory design approaches. In Design and the social sciences (pp. 18-25). CRC Press.

Sanders, E. B. N. (2006). Design serving people. Cumulus Working Papers, 28-33.

Sanders, E. B. N., Brandt, E., & Binder, T. (2010, November). A framework for organizing the tools and techniques of participatory design. In Proceedings of the 11th biennial participatory design conference (pp. 195-198). ACM.

Sanders, E.B. N. & Dandavate, U. (1999) Design for experiencing: New tools. Proceedings of the First International Conference on Design and Emotion, Delft, The Netherlands. 87-92.

Sanders, E. B. N. & Stappers, P. J. (2008). Co-creation and the new landscapes of design. Co-design, 4(1), 5-18.

Sanders, E. B. N., & Stappers, P. J. (2014). Probes, toolkits and prototypes three approaches to making in codesigning. CoDesign, 10(1), 5-14.

Sanders, E. B. N., Singh, S., & Braun, E. (2018). Co-Designing with Communities.

Sansonnet, J. P., Leray, D., & Martin, J. C. (2006, August). Architecture of a framework for generic assisting conversational agents. In International Workshop on Intelligent Virtual Agents (pp. 145-156). Springer, Berlin, Heidelberg.

Seeger, A. M., Pfeiffer, J., & Heinzl, A. (2017). When do we need a human? Anthropomorphic design and trustworthiness of conversational agents.

Shawar, B. A., & Atwell, E. S. (2005). Using corpora in machine-learning chatbot systems. International journal of corpus linguistics, 10(4), 489-516.

Stappers, P. J. (2007). Doing design as a part of doing research. Design researchnow, 81-91.

Steen, M. (2013). Co-design as a process of joint inquiry and imagination. Design Issues, 29(2), 16-28.

Steen, M., Manschot, M., & De Koning, N. (2011). Benefits of co-design in service design projects. International Journal of Design, 5(2).

T-Mobile Austria wins consumer experience award, Tinka gains best chatbot accolade among DAX firms. (2018, November). December 11, 2018, Retrieved from https://www.telecompaper.com/news/t-mobile-austria-wins-consumer-experience-award-tinka-gains-best-chatbot-accolade-among-dax-firms--1270716

The color system. (n.d.). July 20, 2018, Retrieved from https://material.io/design/color/the-color-system.html#tools-for-picking-colors

Tony, Russell-Rose. (2017, January 24). A framework for chatbot evaluation. Retrieved from https://isquared.wordpress.com/2017/01/24/a-framework-for-chatbot-evaluation/

Tribe, R. (2002). Mental health of refugees and asylum-seekers. Advances in psychiatric treatment, 8(4), 240-247.

Vaajakallio, K., Lee, J. J., & Mattelmäki, T. (2009, June). It has to be a group work!: co-design with children. In Proceedings of the 8th international conference on interaction design and children (pp. 246-249). ACM.

Vaajakallio, K., & Mattelmäki, T. (2007, August). Collaborative design exploration: envisioning future practices with make tools. In Proceedings of the 2007 conference on Designing pleasurable products and interfaces (pp. 223-238). ACM.

Vaajakallio, K., & Mattelmäki, T. (2014). Design games in codesign: as a tool, a mindset and a structure. CoDesign, 10(1), 63-77.

Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (pp. 521-530). ACM.

Walker, M. A., Litman, D. J., Kamm, C. A., & Abella, A. (1997, July). PARADISE: A framework for evaluating spoken dialogue agents. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics (pp. 271-280). Association for Computational Linguistics.

What is conversation design. (n.d.). June 12, 2018, Retrieved from https://designguidelines.withgoogle.com/conversation/conversation-design/what-is-conversation-design.html#

What is wireframing. (n.d.). July 5, 2018, Retrieved from https://www.experienceux.co.uk/faqs/what-is-wireframing/

Xu, Y., Maitland, C., & Tomaszewski, B. (2015, May). Promoting participatory community building in refugee camps with mapping technology. In Proceedings of the Seventh International Conference on Information and Communication Technologies and Development (p. 67). ACM.

Appendices

Appendix 1: The most relevant conversational components by *Actions on Google* (1/3)

Conversational component	Guidelines	Examples	Usage
Acknowledgments	Acknowledgments reassure the user that they have been heard and that your persona is keeping track of the conversation. It also helps the dialog feel fluid and natural.	“Okay, ” “Sure, ” “Alright, ” and “Thanks”	Use them to acknowledge acceptance, confirmation, refusal, disconfirmation, correction, and before changing the subject.
Apologies	It’s okay to use “sorry” when it serves a transitional social or phatic function and is not a full-fledged, heartfelt apology.	“Sorry”	“Sorry” is most helpful in no match prompts to make it clear to the user that your persona couldn't understand or interpret their response in context.
Commands	Indicating actions the user can take	“Create a bouquet of yellow daisies and white tulips.”	After a No-Match error, it’s okay to offer suggestions of things the user could say when they could benefit from more support.
Confirmations	Confirmations give users feedback on how their input was understood. This not only empowers users to correct mistakes immediately, but it also reassures them in a socially and conversationally appropriate way by establishing common ground. Furthermore, confirmations help carry the thread of the conversation forward by maintaining context.	“Got it. The men’s running shoes in royal blue and neon green. In what size?”	Acknowledge that an action has been completed (unless it is self-evident) or double-check with the user before performing an act that would be difficult to undo, for example, deleting user data, completing a transaction, etc.
Discourse markers	When communicating, one marks how upcoming words or phrases related to previous discourse (i.e., spoken or written language used in a social context). These discourse markers ease comprehension by providing a preview of what’s coming up next; they are essential for making a conversation sound natural and fluid, as opposed to robotic and stilted.	“By the way, ... ” and “For example, …”	- Use discourse markers like “and” or “also” to show how an utterance adds to the previous one. - Use discourse markers like “now” to change topics. Use discourse markers like “by the way” to introduce additional, often tangential information that’s highly relevant.

Appendix 1: The most relevant conversational components by *Actions on Google* (2/3)

Conversational component	Guidelines	Examples	Usage
Endings	Users abandon tasks for all kinds of reasons. Maybe they got interrupted. Maybe they lost interest. Or maybe the system persona misunderstood them and consequently took them down the wrong path. Regardless of the reason, one should let the user decide when the conversation should end.	"Anything else I can help you with right now?"	- Say goodbye. - Once the user has indicated the conversation is over, assume that you've lost their attention.
Errors	- A No-Match error occurs when the Action can't understand or interpret the user's response in context. - No Inputs, The Action hasn't detected response from the user.	"Sorry, for how many?"	- Cooperation. Assume the user is cooperative, and what they're saying is relevant and valid. - Transparency. Be honest and transparent when explaining why something doesn't work. - Context-specific. Good error handling is context-specific.
Greetings	The chatbot has to introduce itself and make an excellent first impression by showing value. The goal is to make the user feel confident and in control as quickly as possible, so it's essential to help users discover what they can do with the chatbot without making it feel like a tutorial.	"Welcome"	There are three main goals you want to accomplish with your greeting: 1) Welcome the user, 2) Set expectations, and 3) Let the user take control
Questions	One of the most effective ways to get the user to continue the conversation (e.g., make a choice) is to ask a question. When the call to action isn't clear, the user won't know when, or how, to respond.	"What kind of flowers would you like in your bouquet?"	The way of how a question is phrased sets the user's expectations for what they can say. This phrasing can range from open-ended, or wide-focus questions, to close-ended, or narrow-focus questions.
Suggestions	The chatbot can provide suggestions to help the user answer a question. Suggestions can also be used as hints to help the user discover new features.	"I can tell you more about I/O. For example, you might like to know about keynotes, codelabs, or app reviews. I can also help you find sessions or office hours. So, what do you want to know?"	- Use chips to suggest answers. - Suggestions are most helpful for wide-focus questions, although all questions can benefit from the quick-tap response they enable.

Appendix 1: The most relevant conversational components by *Actions on Google* (3/3)

Conversational component	Guidelines	Examples	Usage
Chips	Chips help users 1) refine topics, 2) discover related topics, next steps, and pivots, and 3) take action. Sometimes it's faster and more comfortable for users to tap a chip than it is to say or type their response. When users touch a chip, that text becomes part of the conversation as the user's response.	"By the way, ... " and "For example, …"	- Refine topics. - Discover related topics, next steps, and pivots. - Take action.

Appendix 2: Chatbot Design Survey for the ERICS Project 2018

We are a research group at Aalto University, and we're developing the ERICS (European Refugee Information and Communication Service) project. We intend to develop a high-quality Chatbot service providing valuable information about daily life and work to refugees and newcomers.

This survey will take about 10 minutes of your time. The purpose of this survey is to gather insights into migration and corresponding chatbot services. Your anonymous answers will not be given to third parties and are used solely for research purposes.

Among the participants, we will raffle two movie tickets. Please fill out your email at the end of this survey to participate in the raffle.

Feel free to contact us if you have any questions.
zhifa.chen@aalto.fi

Section 1:

1. How old are you?

2. What is your gender?

□Male □Female □Other

3. What is your home country?

4. How long have you been in Finland?

□ 0 - 3 months □ 3 - 6 months □ 6 - 12 months □ 1 - 2 years
□ Over 2 years

5. How did you find information about living/working in Finland?

6. What kind of problems do you face when trying to get information about life and work?

7. What helps you get that information about life and work smoothly?

8. What makes you feel safe and comfortable when you are in a new country?

Section 2:

Note: A chatbot is an AI-based conversational service that you interact with via a chat interface.

9. What personality traits would you expect from the ERICS chatbot? (at least 3 adjectives)

10. Which avatar would best portray the ERICS chatbot as an information provider?

□ A human □ An animal □ An alien □ An artificial intelligence □ An item □ Other

11. Elaborate why this avatar?

12. Which gender would you like the ERICS chatbot to be?

□ Don't care □ Male □ Female □ Androgynous

13. How mature would you like the avatar to be?

□ Don't care □ Baby □ Kid □ Teen □ Adult □ Middle age □ Senior

Appendix 2: The property checklist for the prototype

Essential properties of the chatbot	Yes	May be	No	Comments/concerns
Performance				
The conversation is natural.	[ ]	[ ]	[ ]	
System prompts are written in simple and understandable language.	[ ]	[ ]	[ ]	
When in the low confidence of question answering, the topic-related answer selection is helpful.	[ ]	[ ]	[ ]	
The error handling is satisfactory.	[ ]	[ ]	[ ]	
Features				
The rating of messages is useful.	[ ]	[ ]	[ ]	
The request for verifying the answer is acceptable.	[ ]	[ ]	[ ]	
Usability				
The following interface components are legible.				
System avatar	[ ]	[ ]	[ ]	
Rating icons	[ ]	[ ]	[ ]	
Rating chips	[ ]	[ ]	[ ]	
Messages	[ ]	[ ]	[ ]	
Further-support buttons	[ ]	[ ]	[ ]	
Send buttons	[ ]	[ ]	[ ]	
Close buttons	[ ]	[ ]	[ ]	
The following interface components are understandable.				
System avatar	[ ]	[ ]	[ ]	
Rating icons	[ ]	[ ]	[ ]	
Rating chips	[ ]	[ ]	[ ]	
Messages	[ ]	[ ]	[ ]	
Further-support buttons	[ ]	[ ]	[ ]	
Send buttons	[ ]	[ ]	[ ]	
Close buttons	[ ]	[ ]	[ ]	
Aesthetics				
The visual design is harmonious (adapted to the HBG website).	[ ]	[ ]	[ ]	
Size				
The size of the chat window is appropriate	[ ]	[ ]	[ ]	

Aalto University
School of Arts, Design
and Architecture