Development and evaluation of the chatbot solution for Wikipedia editors

Wioletta Gluza

Izabela Turaj

171 515 characters, 26 107 words, 71 pages

Chatbot preview: https://heywiki.herokuapp.com
Chatbot codebase: https://github.com/w-gluza/hey-wiki

Summary

Wikipedia strives to improve the diversity and inclusion of the community and make knowledge available to everyone. The goals set for 2030 to dedicate effort to knowledge gaps, knowledge entity, and foundations resemble the pursuit of being globally representative and more accessible. This paper explores how technology can support this initiative by improving the editor experience on Wikipedia.

The problem statement describes the development and evaluation process of the chatbot solution that can potentially impact editor activation and retention on Wikipedia. The research questions addressing the problem focus on evaluating the usability and user experience of developed solutions and seeking possible future improvements to the conversational interface.

The research draws from the related literature about chatbot user experience, human-chatbot interactions, usability, and chatbot platforms and frameworks. The chatbot is designed using the atomic design approach. The development is powered by IBM Watson Assistant and results in a full-stack application coded in React and Node.js and deployed to Heroku. The chatbot evaluation is gathered through a think-aloud test completed by eleven participants who rate the usability of the chatbot in post-task Single Ease Question and post-test Chatbot Usability Questionnaire. The qualitative data collected in the study is consolidated in the thematic analysis, which uncovers two themes of positive chatbot user experiences and negative chatbot user experiences. Each theme describes a different perspective on the usability of the chatbot, functional, and social human-chatbot interactions. The sub-theme of usability in thematic analysis describes chatbot's understanding of utterances, intent recognition, features such as typing, predefined options, multi-path functionality, suggested topics, or interactive resources.

The negative user experience reflecting on usability concerns poor error handling or misunderstanding user input. The thematic analysis described the negative social characteristics of user experience that should be improved as user confusion when interacting with the chatbot and highlighted the chatbot's personality issues. The chatbot seems neutral, not human enough, or too robotic. The functional characteristics that need improvements include the necessity of rephrasing or using keywords to adjust the language to the chatbot. The chatbot provided questionable information or inefficient synthesis of the essential information.

The final solution is assessed as helpful, informative, and friendly; however, further work is necessary to improve the chatbot's personality. The future research direction could be chatbot assessment in the organic setting.

Table of contents

Development and evaluation of the chatbot solution for Wikipedia editors

Wioletta Gluza[1], Izabela Turaj[2]

[1] Information Studies, Aalborg University in Copenhagen
wgluza20@student.aau.dk

[2] Information Studies, Aalborg University in Copenhagen
ituraj20@student.aau.dk

Supervisor: Florian Meier
fmeier@hum.aau.dk

Abstract. Wikipedia strives to improve the diversity and inclusion of the community and make knowledge available to everyone. The goals set for 2030 to dedicate effort to knowledge gaps, knowledge entities, and foundations resemble the pursuit of being globally representative and more accessible. This paper explores how technology can support this initiative by improving the editor experience on Wikipedia. The study describes the development and evaluation process of the chatbot solution that can potentially impact editor activation and retention on Wikipedia. The chatbot is designed using the atomic design approach. The development is powered by IBM Watson Assistant and results in a full-stack application coded in React and Node.js and deployed to Heroku. The chatbot evaluation is gathered through a think-aloud test completed by eleven participants who rate the usability of the chatbot in post-task Single Ease Question and post-test Chatbot Usability Questionnaire. The qualitative data collected in the study is consolidated in the thematic analysis, which uncovered two themes of positive chatbot user experiences and negative chatbot user experiences. The final solution is assessed as usable, informative, and friendly; however, further work is necessary to improve the chatbot's personality.

Keywords: Wikipedia, editor experience, human-chatbot interactions

1 Introduction

Wikipedia is a digital encyclopedia with the objective of making knowledge available to everyone. The project has gathered millions of people worldwide for the past 20 years. Despite being a comprehensive source of knowledge written in over 300 languages, Wikipedia still comes across challenges with inclusion and diversity. Therefore, the resolution to celebrate its twentieth anniversary grounded the initiatives to improve the overall health and diversity of the organization (Wikimedia Foundation, 2021a).

The whitepapers about knowledge gaps, knowledge entity, and foundations were published to draw the direction of future initiatives to increase the diversity of the Wikipedia community. The most prominent reflections from the publications highlight the impact of Wikipedia editors and encourage initiatives supporting the contributors. The proposed activities include diversifying the community, networking, improving onboarding and retention of novice editors, guiding community members in preventing and fighting Wikipedia policy violations (Wikimedia Foundation, 2019).

Several research papers emerged to identify the factors affecting editor experience on Wikipedia. Bryant, Forte, and Bruckman (2005) study the gaps in editor demographics, and Hargittai and Shaw (2014) focus on gender bias. Halafaker et al. (2012) explore the reasons for resistance towards new editors, DeDeo (2016) described different types of conflicts between them, and Halafaker, Kittur, and Riedl (2011) mention the recurrent content rejection as one of the explanations for poor editor retention. The researchers suggest potential solutions to improve editor experience by encouraging socialization and positive interaction (Halfaker et al., 2013), moving the current focus from more experienced editors to the new contributors by providing quality feedback and personal guidance (Halfaker, Kittur, & Riedl, 2011), or teaching underrepresented groups how to function in complex technoscientific infrastructure (Ford, & Wajcman, 2017).

The Growth Team from Wikipedia provides a more technical perspective that highlights the role of information systems. Software solutions can support content creation and building a strong community by limiting editor retention and engaging new contributors through better onboarding. According to the Growth Team, the core issue of current editor experience is the lack of cultural and technical skills. The proposed projects provide technological solutions for in-context help, human-to-human help, and task recommendations (Contributors to Wikimedia Projects, 2021).

The live chat was an initiative by the Growth Team from Wikipedia to help novice users navigate through the information available on help pages and reach out asynchronously to more experienced community members. The project did not bring satisfactory outcomes since the user testing uncovered the necessity of instant reply to the query rather than extended time to receive answers from the fellow editor in the chat (Wikipedia Contributors, 2021). However, the data collected in the experiment shows positive signs. About 20% of the editors interacted with the panel, and 50% used the contents. The activities initiated by the Growth Team show the need for such a tool. Further research is encouraged to determine a successful solution for the

activation and retention of inexperienced editors (Wikipedia Contributors, 2021). A conversational interface could support Wikipedia editors in their struggles and potentially solve the issue of editor retention and activation of inexperienced editors on Wikipedia.

1.1 Problem statement

This research documents the development and evaluation of the chatbot solution to improve editor experience on Wikipedia and support Wikipedia's pursuit of enhancing editor retention and onboarding. The development and evaluation of a chatbot solution for Wikipedia editors are considered in regard to the following research questions:

RQ 1: How usable is the chatbot solution developed for Wikipedia editors?

RQ 2: What user experience aspects of the chatbot need to be improved?

The definition of user experience used in this paper was formulated by the international standard on ergonomics of human-system interaction to identify the characteristics of chatbot user experience for Wikipedia. Three aspects influence user experience: the system, the user, and the context of use (International Organization for Standardization [ISO], 2019). Peter Morville (2004) draws from this definition and formulates the facets of user experience that describe the experience as usable, useful, findable, credible, accessible, desirable, and valuable (Morville, 2004). This research focuses on evaluating the usability and user experience of the chatbot solution for Wikipedia editors.

The International Standards Organization (ISO) characterizes usability as the extent to which a service or a product can be used in context by individuals to accomplish specific objectives with efficiency, effectiveness, and satisfaction (International Organization for Standardization [ISO], 2019). Although there are multiple usability evaluation methods for software solutions (Murillo et al., 2018), a consolidated conversational interfaces method is not yet established. Therefore, the research direction is set by the Chatbot Usability Questionnaire (CUQ), which introduces specific usability attributes for chatbot evaluation (Holmes et al., 2019).

In the following chapters, we provide the case description to introduce the work that has been done so far before the development and evaluation of the chatbot. Next, we summarize the related work in the fields of chatbot user experience, human-chatbot interaction, usability, and chatbot platforms and frameworks. The chapter about methodology includes the description of the process from design, through development, to the deployment of the chatbot, and consequently the evaluation of the solution in the remote unmoderated usability test, including think-aloud evaluation, post-task, and post-test evaluation of the chatbot. The collected outcomes are summarized in the results section and discussed in conclusion about the usability of the application and user experience aspects that need further improvements.

1.2 Case description

This paper caps the series of studies conducted in the previous semesters about improving the editor experience on Wikipedia. The earlier work explores the background, motivations, and frustrations of Wikipedia editors (Gluza, & Turaj, 2021a), the impact carried by technological solutions to improve editor onboarding and retention (Gluza, & Turaj, 2021b), and chatbots as a potential software solution that can help accomplish the community's goals (Gluza, & Turaj, 2021c). The subsequent sections outline the most prominent findings from previous work and describe each study shortly to provide context and show their impact on chatbot development.

The essential papers supporting the chatbot development process so far include reflections on social (Chaves, & Gerosa, 2020) and functional characteristics of user experience with chatbots (Brandtzaeg, & Følstad, 2017), technical aspects of design and development (Böhm et al., 2020), and perspective on creating conversational agent especially for Wikipedia editors (Skorupska et al., 2020). User intent recognition was in the center of attention during the process so far, meaning identifying people's intention behind their message to a chatbot (Bohm et al., 2020). An adequate practice of generating the intent and entity pairs from the online resources provides the knowledge foundation for the chatbot (Kobayashi et al., 2015). Therefore, the chatbot development started from the analysis of the frequently asked questions pages about contributing to Wikipedia (Wikipedia contributors, 2021a; Wikipedia contributors, 2021b) and Wikipedia help desk archives that provide authentic challenges described by inexperienced editors (Wikipedia contributors, 2021h).

The contribution taxonomy created by the Wikimedia Foundation Product team is a helpful resource for the consecutive iterations of chatbot development (Contributors to Wikimedia, 2021b). The contribution taxonomy includes the index of all the editor activities and the visualization structure. The master inventory consists of the catalog of workflows, assessment criteria, scoring system, and estimations. The activities are illustrated in a format that outlines the 88 workflows and over 500 steps recorded in the inventory (Wikimedia Product, 2019). Based on this database, we received access to the diverse collection of user intents that can be beneficial in training the chatbot. The inventory accessible online consists of the information about the focus of each activity user decision, task, and the further division into steps. Potential utterances can be formulated based on the decisions, tasks, and workflows described in the inventory (Wikimedia Product, 2018).

The first study we conducted gives an overview of the knowledge gaps by exploring the editor experience. The mixed-method concept of a questionnaire and participatory observation focuses on Polish Wikipedia editors' background, motivations, and experiences. It explores the factors that inhibit the editors when contributing to Wikipedia and questions how the portal's usability impacts the editing behavior (Gluza, & Turaj, 2021a). The aspects inspiring editors to contribution include knowledge sharing, fixing issues, and closing content gaps. Communication problems and the unintuitive user interface are the most occurring frustration sources.

The system usability of the editor portal needs further evaluation and improvements to support the editors in their tasks (Gluza, & Turaj, 2021a).

The objective of improving editor experience on Wikipedia continues in the next paper by exploring how technology can support this process. The human-centered design process is applied to support the creation of the conversational interface that considers functional and social aspects of interactions. The learnings from the study show that the chatbot could support the need for instant responses, improved productivity, and provide a knowledge base to improve editor activation (Gluza, & Turaj, 2021b). The chatbot should demonstrate proactivity, damage control capabilities, manners, and emotional intelligence. The conversational elements, such as utterances, intents, and entities (Abdellatif et al., 2020), refer to the contribution workflows expressed as creation, curation, and moderation (Wikimedia Product, 2018). The chatbot interface is based on the atomic design approach and provides editors with accessible and usable interactive elements. As a result of the research, a chatbot prototype has been developed, and further improvement needs were established (Gluza, & Turaj, 2021b).

The most recent paper follows the CRISP-DM process for data mining to gather and analyze the data for intent classification and further training the chatbot solution. Three machine learning approaches are applied to find the best-performing model in classifying user intents scraped from the Wikipedia archives (Wikipedia contributors, 2021h) into categories (Wikipedia contributors, 2021e). Multinomial Naive Bayes, Support Vector Machine, and Random Forest algorithms are evaluated in three iterations with CountVectorizer and TfidfVectorizer. The findings suggest Multinomial Naive Bayes with CountVectorizer as the best performing model, while Support Vector Machine with TfidfVectorizer performs well in natural language processing context for machine learning problems (Gluza, & Turaj, 2021c).

The research proceeds to develop and evaluate the chatbot solution for Wikipedia based on the knowledge acquired so far. We determine which user experience aspects should be improved depending on the evaluation. The next chapters cover the related research and present the methodology chosen to address the research questions. Finally, we analyze the results and evaluate the chatbot solution to improve the editor experience on Wikipedia.

2 Related work

In the related work chapter, we introduce the studies that explore the factors impacting the user experience of the chatbot. The related studies describe the usability and the overall user experience of conversational interfaces.

Følstad et al. (2021) consolidate the findings in the field of conversational agents after years of research and propose future directions that cover six interest areas: users and implications, user experience and design, frameworks and platforms, chatbots for collaboration, democratizing chatbots, ethics, and privacy (Følstad et al., 2021). The direction of users and implications looks for an understanding of user behavior and the social value of the chatbots. The strategies for improving, modeling, and

evaluating user experience are described in the direction of chatbot user experience and design. Context understanding and interpretation capabilities are mentioned in the frameworks and platforms direction and consider how they are sustained by new chatbot design and testing methods. The exploration and modeling of human-chatbot interaction are described in the research direction about chatbots for collaboration. The implementation of technology for the common good and promoting inclusive information systems are described in chatbot democratization. The ethics and privacy research subject describes the potential challenges and risks of conversational agents (Følstad et al., 2021).

Table 1: Topics and directions for future research on chatbots (Følstad et al., 2021).

Topics	Proposed future research directions
1. Users and implications	(a) Emerging chatbot user groups and behaviors (b) Social implications of chatbots
2. Chatbot user experience and design	(a) Design for improving chatbot user experience (b) Modeling and evaluating chatbot user experience
3. Chatbot frameworks and platforms	(a) Interpretation capabilities and context understanding (b) Emerging techniques for chatbot design and testing
4. Chatbots for collaboration	(a) Modeling human-chatbot collaboration (b) Empirical investigations of human-chatbot collaboration
5. Democratizing chatbots—chatbots for all	(a) Chatbots for social good (b) Inclusive design with and for diverse user groups
6. Ethics and privacy in chatbots	(a) Understanding chatbot ethics and privacy (b) Ethics by design

This research draws from the considerations about chatbot frameworks and platforms for the development process and delves into the topic of chatbot user experience and design for evaluation purposes.

2.1 Chatbot user experience

State of the art regarding chatbot frameworks and platforms suggests that technological advances are the key enablers of the current interest in conversational interfaces. The fundamental advances include implementing supervised machine learning for information retrieval and classification and identifying user sentiment and intent prediction, which is crucial for task-oriented interactions (Følstad et al., 2021). Moreover, the prediction of appropriate responses to the user input is possible due to models constructed from large conversational datasets (Adiwardana et al. 2020). Some of the new architectural approaches to natural language processing and high-quality open-source libraries enabled creators and lowered the entry barriers to conversational design (Wolf et al., 2019). Nowadays, it is possible to build conversational models that exhibit high coherence and generalization (Roller et al., 2020). Opening large-scale generative models such as GPT-3 by OpenAI (Brockman, 2021) or BERT by Google (Delvin, & Chang, 2018) has the potential to accelerate chatbot development even further (Følstad et al., 2021). However, there is an ethical concern regarding the emergence and growing impact of large models concerning misinterpretation and misconduct (Zou, & Schiebinger, 2018), carbon footprint (Strubell, Ganesh, & McCallum, 2019), and research opportunity costs (Hao, 2021).

Regardless of substantial advances in chatbot frameworks and platforms, several challenges endure. There is a lack of needed technological background to support critical aspects of chatbot applications, such as understanding user input, modeling according to context, chatbot development and deployment to production, integrating chatbot solutions in different technological environments (Følstad et al., 2021).

While machine learning approaches have empowered both intent prediction and natural language understanding, user input comprehension remains difficult. Human-chatbot interaction is inclined to conversational breakdowns mainly due to the misinterpretation of everyday situations (Porcheron et al., 2018). Modeling and adapting to the interlocutor and conversational context is increasingly important. Chatbots are more often deployed for use in sensitive scenarios such as health care, and it becomes crucial for them to adapt to social and emotional literacy (Jovanovic, Baez, & Casati, 2021). Another challenge is supporting chatbot development and testing by ensuring easier deployment to production. Finally, as chatbots are becoming a part of broader software ecosystems, they should be integrated accordingly to provide a seamless experience for both users and developers (Baez, Daniel, & Casati, 2020).

Følstad et al., 2021 formulate future research directions for chatbot frameworks and platforms as interpretation capabilities and context understanding and emerging techniques for chatbot design development and deployment. Understanding advances in quality training data are necessary to develop interpretation capabilities and context (Følstad et al., 2021). Furthermore, new techniques are needed to include domain specialists in content development, dialog management, and natural language processing. The critical aspects of future research in this field remain user understanding and context recognition. Future research in the area of chatbot design,

development, and deployment can support creators and enable conversational agents' access to information and services (Følstad et al., 2021).

Another broad topic for future research directions is chatbot user experience and design. It relates to the users' perception and response to the conversational agents. Conversational interfaces, interaction mechanisms, and content should be designed to address the user experience and behavior. User-centered evaluations of chatbots are necessary to collect insights and measure chatbots' impact (Følstad et al., 2021).

As of now, chatbot user experience has been the main topic in recent research efforts. The related studies have identified the factors affecting user experience both positively and negatively (Zarouali et al., 2018) and addressed particular aspects such as perceived social support (Lee, & Choi, 2017), trust (Zierau et al., 2020), motivation (Brandtzaeg, & Følstad, 2017), and how these features are impacted by chatbot design (Go, & Sundar, 2019). There is an expanding base of knowledge about human-chatbot interactions concerning conversational design (Ashktorab et al., 2019), personalization (Laban, & Araujo, 2020), or the application of social cues (Feine et al., 2019). The expectations for conversational dialogs are rising to demonstrate advanced social behaviors characterizing human-to-human interactions. Chaves and Gerosa (2020) move away from studies on user motivation and explore possibilities of chatbot design and development that minimize user frustration and dissatisfaction by responding to social needs. Fifty-six research papers have been analyzed to identify the benefits, obstacles, and future directions for human-chatbot interaction (Chaves, & Gerosa, 2020). The research pursues the development of a chatbot solution that brings positive human-chatbot interaction and provides an engaging user experience.

The social characteristics identified in the research by Chaves and Gerosa (2020) cover four primary categories, including conversational intelligence, social intelligence, and personification. The categories are further divided into eleven features distinctive for different chatbot domains. Conversational intelligence help chatbots in managing interactions with people, social constructs are covered by social intelligence, and chatbot's personality is described by personification (Chaves, & Gerosa, 2020).

There is also a body of knowledge on measures and methods for chatbot assessment. User-centered evaluation in chatbot research is conducted through questionnaires (Vanichvasin, 2021), experiments (Van den Broeck et al., 2019), observations with interviews (Luger, & Sellen,2016), heuristic evaluations (Höhn, & Bongard-Blanchy, 2021), or mixed-methods studies (Lin, & Chang, 2020).

Some of the research challenges in the field of chatbot user experience and design formulated by Følstad et al. (2021) include a lack of knowledge on how to use the findings from the research in chatbot design to engage and amuse users consistently. Users encounter multiple issues while using conversational interfaces, where chatbots fail to help them achieve their goals or fulfill their tasks. Apart from the pragmatic problems such as difficulty in user understanding, there are also hedonic issues such as failing to engage the users over time. Despite the number of research initiatives to face these challenges in the fields of healthcare (Cameron et al., 2019), customer service (Adam, Wessel, & Benlian, 2020), and education (Hobert, & Berens, 2020), strengthening of chatbot user experience remains a crucial research challenge (Følstad

et al., 2021). There is also a need to assess and measure chatbot user experience from a more holistic approach to resolve whether conversational agents are actually beneficial. Some of the challenging aspects include usefulness, instance, efficiency, and process support. Even though there are several studies on the overall chatbot user experience, there is a lack of definitions and metrics for key aspects of chatbot evaluations. Moreover, commonly applied approaches to assessment would be beneficial in contrast to a broad range of perspectives (Følstad et al., 2021).

Therefore, the proposed future research directions include improving chatbot user experience and modeling and evaluating the experience that answers the challenges mentioned above. The research should translate theoretical findings and apply conclusions of practical impact on design. In the future, the researchers should move away from exploring and assessing the effect of chatbot design elements and pursue how this knowledge may improve chatbot user experience in industrial applications. Even though theoretical knowledge is still needed, there is a lack of design-oriented objectives and guidelines that can directly inform human-chatbot interaction and conversational design (Følstad et al., 2021). As for modeling and evaluating chatbot user experience, there is a need for convergence of models and measurements for evaluation. In the emerging field of research, interest diversity in definitions is natural; however, there is a need for agreement and consistency in applying terminology of user experience constructs and standardized measurements for these constructs. There is a benefit in including standard metrics across the studies to enable cross-study comparison for tracking progress over time. Well-established evaluation strategies from fields such as human-computer interaction may be beneficial for standardization (Følstad et al., 2021).

2.2 Human-chatbot interactions

The primary domain of the chatbot can be described as information search defined by Chaves and Gerosa (2020) and based on the previous research analysis mentioned in case description (Kobayashi et al. 2015; Skorupska et al., 2020; Brandtzaeg, & Følstad, 2017; Følstad et al., 2021; Wikipedia Contributors, 2021d). The information search functionality is crucial for inexperienced editors to contribute to Wikipedia as soon as possible. The social characteristics of the human-chatbot interaction belonging to the information search domain include proactivity, damage control, manners, and emotional intelligence (Chaves, & Gerosa, 2020).

The application can manifest proactivity in human-chatbot interactions by starting the conversation, suggesting related topics, providing additional information, or asking follow-up questions. Discussions facilitated by proactive chatbots are more natural (Chaves, & Gerosa, 2020). The value of interaction can be increased by proactively providing useful and often unexpected information. The users are inspired by introducing new ideas and dynamic conversational flow. Without realizing, the users participate in onboarding, gain knowledge, or learn about the chatbot's capabilities (Chaves, & Gerosa, 2020). Some of the benefits of proactivity for chatbots include quick recovery from failure described by Chaves and Gerosa (2020) as damage control. The potential improvements for Wikipedia editors using a proactive

chatbot can be visible, especially in task-oriented interactions. Proactive chatbots can potentially motivate and engage users to finish the task by notifying them (Chaves, & Gerosa, 2020).

On the other hand, timing and relevance, privacy issues, or the user's perception of surveillance can be the challenging aspects of proactivity (Chaves, & Gerosa, 2020). The main user motivation to use chatbots is productivity (Brandtzaeg, & Følstad, 2017); therefore, quick responses and informative messages are especially applicable to Wikipedia editing. Moreover, the subjects provided by the chatbot should be relevant, the initiation of the new topics selected carefully, and responses should be timely (Reboot, 2017).

Efficient proactivity can be achieved by following two strategies of leveraging conversational context or suggesting a random discussion (Chaves, & Gerosa, 2020). The conversation topics should be relevant; therefore, the chatbot should leverage the conversational context to improve editor retention and onboarding on Wikipedia. The chatbot can do it by proactively offering editors a set of valuable topics. Although it is one of the most common approaches (Chaves, & Gerosa, 2017), emotionally aware agents should benefit from user intent recognition. Progressive chatbots should support novice editors differently than experienced contributors and offer not only information but also encouragement (Chaves, & Gerosa, 2017).

The next social characteristic described as damage control depends on the social intelligence of the chatbot. Conversational agents need to manage failure and conflict. People are more demanding towards the software than other humans; therefore, chatbots should be equipped with tools to handle difficult situations (Hill et al., 2015).

Some of the benefits of damage control affecting human-chatbot interactions include the capability to react to violence or disturbance, user testing, or acknowledgments of false steps due to system limitations or unsatisfactory information. The researchers propose strategies to control damage depending on the desired human-chatbot relationship, including emotional reaction, authoritative response, or change of the topic with no answer to the user's utterance. More advanced systems can leverage consciousness. However, no matter if the partner is a human or a machine, none of the interactions are conflict-free. Since there are no apparent solutions to avoid crisis, the related research suggests politeness to manage challenging conversations (Chaves, & Gerosa, 2017).

Human habits and polite behavior can be manifested in chatbots through a social characteristic called manners. Examples of chatbot activities expressing manners include greetings, apologies, or conclusions (Morrisey, & Kirakowski, 2013). Using manners in a chatbot design can improve unsuccessful interaction and reduce frustration. Understanding the context of use and engaging in small talk prevent invasive behavior and are common strategies for equipping chatbots with manners (Chaves, & Gerosa, 2020).

The chatbots do not have a natural ability to show emotions, but the social characteristic called emotional intelligence can help them manifest social cues when helping with information search. Conversational agents should identify and influence people's emotions to improve overall human-chatbot interactions. Enriched relationships, increased credibility, and engagement are among the perks of

transferring emotional intelligence to the system (Chaves, & Gerosa, 2020). Increased attention could support the process of transitioning novice Wikipedia editors into experienced contributors and retaining them (Bryant, Forte, & Bruckman, 2005). In conversational design, user engagement is affected by the chatbot's capability of sustaining extended conversations, receiving feedback, and providing user interactive interface components (Chaves, & Gerosa, 2020).

User motivation impacts the way chatbots are used. The researchers recognize productivity as the most prominent motivational factor in human-chatbot interactions; however, curiosity, entertainment, and social purposes can be observed as well (Brandtzaeg & Følstad, 2017). Purposeful human-chatbot interactions can be designed based on these findings by setting clear objectives for user experience. We argue that conversational interface can be valuable for contributors on Wikipedia who are often motivated by sharing knowledge and fixing existing issues which contributes to productivity (Reboot, 2017). The researchers further explain that productivity is about speed, ease, and convenience when using conversational agents.

Moreover, users value guidance provided by the chatbot and open access to information. The chatbot is perceived as beneficial based on its usefulness. The users are more likely to interact with the conversational interface and emphatically evaluate their experience when the productivity goal is fulfilled through instant and informative support (Brandtzaeg, & Følstad, 2017).

The past initiative to introduce the help center for Wikipedia contributors was unsuccessful mainly because it did not provide timely responses (Wikipedia Contributors, 2021d). The issue could be solved by creating a chatbot providing suggestions and answers in real-time. The related literature supports this view with the questionnaire. The most prominent perks of using the chatbot were presented as 24-hour support, instant replies, and quick responses to trivial questions (Modrzejewski, & Rokita, 2018).

Følstad and Brandtzaeg (2020) researched the positive and the negative user experience with chatbots. The chatbots used commonly by users need to be helpful and engaging. The researchers identified and reflected on the chatbots' satisfactory and frustrating experiences. The findings guide chatbot creators by analyzing both positive and negative perspectives to pragmatic and hedonic attributes of user experience (Følstad and Brandtzaeg, 2020).

The positive, pragmatic experiences refer to usefulness, helpfulness, support in task completion, or the advantages of acquiring general knowledge (Følstad and Brandtzaeg, 2020). The findings show that the chatbot for Wikipedia editors should be informative and helpful, help complete specific editorial tasks. Examples of such activities could be providing general information about Wikipedia policies and guidelines or supporting the novice editor in adding a reference to an article.

The positive, hedonic experiences presented in the study suggest that the chatbot should be entertaining, provide an impression of innovation or motivational content (Følstad and Brandtzaeg, 2020). Wikipedia editors could benefit from these hedonic experiences as currently, the chatbot solution supporting them does not exist. Many struggle with inspiration to edit due to the frequent reverts of their work (Halfaker, Kittur, & Riedl, 2011).

However, negative pragmatic experiences occur and usually derive from the poor usability of the chatbot. The most prominent negative factor of pragmatic experiences is the limited ability to understand user input, tiring repetitiveness, or unsuccessful support attempts (Følstad, & Brandtzaeg, 2020). The related literature raised the issue encountered by the potential solution for Wikipedia editors suggesting recovery techniques to ensure a positive user experience with the chatbot regardless of its failure. Apologies or requests to rephrase are proposed methods of recovery followed by switching the subject, task, or answer type (Skorupska et al., 2020).

Although the pragmatic negative experience dominated the responses in the study, the hedonic attributes usually take the form of uncanny or rude answers, unwanted or random information, monotonous content (Følstad, & Brandtzaeg, 2020). When designing the chatbot for Wikipedia editors, profound content selection for chatbot answers is necessary to support different unsuccessful scenarios. The conversational designers should consider average response time, options to abandon the task, rephrasing or correcting answers, accepting multiple input types, and analyze user satisfaction overtime to prepare the chatbot for humanlike conversations (Skorupska et al., 2020). The above-mentioned indicators were used in the conversational flows to test the chatbot's abilities.

The conclusions from Følstad and Brandtzaeg (2020) highlight the importance of usefulness in task-oriented chatbots. A positive user experience can be achieved by solving the problems for the users to fulfill the hedonic attributes. Moreover, monitoring the insights about user behavior with the chatbot is crucial for the chatbot creation process as it helps understand user needs and motivations. The customization of the chatbot and user group segmentation is expected to bring significant benefits, yet implementing such an advanced solution is challenging.

2.3 Usability

Usability is the fundamental concept of human-computer interaction (Nielsen, 2012). Nielsen (2012) describes usability as the quality aspect defining the interface's ease of use. Five quality components representing usability include learnability, efficiency, memorability, error tolerance, and satisfaction. Finalizing a primary task when the person uses the interface for the first time is described as learnability. Efficiency represents how quickly a user can finish the task once they have learned the interface. Memorability measures how easily people can reestablish fluency while using the design after some time. Error tolerance evaluates the number and severity of errors made by users. The last usability aspect is satisfaction from using the interface (Nielsen, 2012). The chatbot usability questionnaire used in this research draws from the usability principles.

Borsci et al. (2021), in the comparative study and literature review of four research papers on chatbot usability, develop a quality checklist that supports conversational designers in evaluating their products. Borsci et al. (2021) formulate seventeen attributes to be used as a tool for controlling the quality of chatbot functionalities during the design phase for short-term and long-term interactions. In the case of short-term interactions, the researchers describe fifteen attributes. The first attribute is

the ease to start the conversation and the ability to respond appropriately regardless of the situation. Access to the chatbot should be easy, including the accessibility of the functions and location in the interface (Borsci et al., 2021). Expectation setting, another quality attribute is the ability of the chatbot to create fair expectations for the end-users regarding its capabilities (Bostrom, & Yudkowski, 2014). Flexibility and communication effort describe how chatbots manage and adapt to different conversational styles of the users (Meira, & Canuto, 2015). Another feature is maintaining a themed conversation once introduced and preserving the context to understand users' utterances (Radziwill, & Benton, 2017).

Regarding context, it is also essential for a chatbot to sustain the reference to the system or environment (Staven, 2017). Protecting users' privacy and security is a quality feature that ensures trust and allows a chatbot to make appropriate decisions to support the user (Eeuven, 2017). Chatbots need to recognize and facilitate users' goals and intents. The relevance of the information provided by the conversational agent allows users to achieve their goals (Wilson et al., 2017). Chatbots should respond with quality support without adding too much information to fulfill the maxim of quantity (Borsci et al., 2021).

Moreover, chatbots should respond appropriately in situations when they are not equipped to handle the conversation or request (Morrisey, & Kirakowski, 2013; Vetter, 2002). Understandability and politeness allow chatbots to understand user input and carry correct statements and responses with acceptable manners and certainty (Morrisey, & Kirakowski, 2013). Perceived conversational credibility and speed of answer provide chatbot with the capability of providing valuable and informative responses in a timely manner (Borsci et al., 2021). Finally, the chatbot's attribute of meeting neurodiverse needs (Radziwill, & Benton, 2017) creates an accessible environment for the users with diverse needs regardless of their well-being, health condition, or age (Borsci et al. 2021). The researchers identify two quality attributes of interaction enjoyment and personality for long-term interactions. Interaction enjoyment means that the conversation with the chatbot is enjoyable and engaging (Ramos, 2017). The chatbot's personality is expressed in self-introduction, greeting, providing information, and empathetic behavior (Borsci et al., 2021).

Kvale et al. (2020) formulate the characteristics of successful and unsuccessful chatbot dialogs to help designers in chatbot training. The crucial findings for successful conversational flow include the chatbot's linguistic elements reflecting on the user's vocabulary, recognizing user input regardless of the misspelling or other mistakes, and efficient integration with a base information system that enhances the user journey. Some of the common issues with the dialogs included failing in understanding user input or identifying wrong intent, gaps in the chatbot's knowledge base, or poor integration with the information system providing the chatbot. Intent recognition issues or missing content typically motivated the suggestions for improvements in the chatbot user experience. While failures in identifying existing intents are relatively easy to recognize as chatbots respond with a fallback message, the wrong interpretation of the utterance is more problematic and may require manual analysis (Kvale et al., 2020). The findings from the research can help evaluate the chatbot and formulate possible improvements for a better editor experience.

Cameron et al. (2018) advocate that the chatbot development lifecycle differs from traditional software development processes. In the case of the conventional user interface, the focus is on the prototyping user interface, whereas for the conversational design, the primary goal is modeling the dialog between the user and the chatbot (Cameron et al., 2018). The best practices for user experience practitioners are still applicable; however, there is a need for more specific guidelines that can support the design and development of chatbots (Følstad et al., 2021). Moore et al. (2017) formulate three basic principles of conversation design. The principles include recipient design, minimization, and repair. The recipient design means that the conversation is tailored to match the user's understanding level (Moore et al., 2017). The minimization corresponds to the maxim of quantity (Borsci et al., 2021) and suggests keeping interaction short and straightforward (Moore et al., 2017). The repair principle is about recovering from failures and guiding a user through conversation to provide necessary support by repeating or rephrasing (Moore et al., 2017). Gibson et al. (2016) see the need to modify evaluation techniques to accommodate different characteristics of the technologies when using testing methods such as thinking aloud, task completion, or usability questionnaires (Gibson et al., 2016). Therefore, there is a need to modify the traditional evaluation methods when testing chatbots (Holmes et al., 2019).

The user experience experts suggest that certain aspects of chatbots should be considered to validate the conversational interfaces. These factors include interaction style, conversation flow, language, and privacy (Budiu, 2018). Therefore, Holmes et al. (2019) argue that chatbot user experience design and usability evaluation require nontraditional methods and multiple metrics to provide a more comprehensive view of usability (Holmes et al., 2019). In their study of usability testing of the healthcare chatbot, they use think-aloud test, post-task test formulated as Single Ease Question (SEQ), and three post-test metrics, including two established methods of System Usability Score (SUS) and User Experience Questionnaire (UEQ). Additionally, one measure called Chatbot Usability Questionnaire (CUQ) was developed primarily to evaluate the usability of the chatbot (Holmes et al., 2019).

In this research, we first design, develop and deploy a chatbot solution for Wikipedia editors considering the challenges and future directions established for chatbot frameworks and platforms (Følstad et al., 2021). Then, we apply the proposed methodology by Holmes et al. (2019) in the qualitative manner conducting the concurrent think-aloud test with ten users who answered the post-task question using a standard questionnaire of SEQ enhanced with a question about the reason of the score to derive helpful qualitative feedback (Laubheimer, 2018). In the post-test questionnaire, we carried out the usability test introduced by Holmes et al. (2019) to follow the directions drawn by Følstad et al. (2021) regarding chatbot user experience and design. Finally, based on the findings from chatbot usability evaluation, we establish the necessary advancements for the chatbot to improve the editor experience on Wikipedia truly.

2.4 Chatbot platforms and frameworks

Several platforms and frameworks have emerged to ease chatbot design. Large tech companies like Google, IBM, Microsoft, or Amazon have implemented chatbot development applications, and a few smaller alternatives exist, such as the open-source platform Rasa (Böhm et al., 2020). These solutions provide versatile functionalities for natural language processing, the formation of the conversation flow, the capability to integrate the chatbot into information systems, and support deployments and testing processes (Perez-Soler et al., 2021). Selecting the most suitable development technique for a particular project is challenging, and making a wrong decision can cause non-compliance with technical requirements. Perez-Soler et al. (2021) describe the process of chatbot development in four steps and highlight that it is not necessarily linear and requires iterations. The phases introduced by research include intent identification, dialog construction, action recognition, and chatbot deployment to a specific channel (Perez-Soler et al., 2021).

The first step is the identification of the user's intents. The intents express that the user's intentions from the message were recognized and include, i.e., requests, needs, or motivations. The program can effectively analyze the content of the conversation and identify the requirements based on the intents (Wochinger, 2019a). While conventional applications generally offer functionality through graphical user interfaces, chatbots present it via conversations. In order to match the intent corresponding to the user input, the natural language understanding libraries can be used as they allow analyzing the phrase structure and grant facilities for tokenizing utterances. However, more straightforward libraries also allow training the chatbot with example phrases characterizing each intent that are more cost-effective for narrow domain projects. The frameworks can also support extracting parameters from user input (Perez-Soler et al., 2021). The categories generated from user intents and applied in further analysis are called entities. The entities identify the data necessary for the application of user commands. Entities can be used to extract values such as number, geolocation, dates, time. The user's needs are exposed in intents, and entities extract details from the dialog (Wochinger, 2019b). Most chatbot development platforms serve predefined entities such as numbers and dates; however, custom entities can be extracted from the available dataset. Additionally, chatbots can formulate fallback intents to ensure that the communication with the user is smooth even if the chatbot does not recognize the utterance (Perez-Soler et al., 2021).

Apart from intents recognition, the chatbot needs to have a defined dialog structure to accomplish the task. The conversational agents need to store the dialog state in context to handle sequential information from the user and guide the conversation naturally. The information needs to be carried through the different stages of the conversation (Perez-Soler et al., 2021).

Moreover, user intents trigger actions that should be identified. Therefore, an integrated chatbot should interact with the user and be able to apply changes in the information system. The responses to the entities that a chatbot should execute are called actions. Single intent can initialize multiple actions. The possible formats of the actions include text strings, links, or graphical elements (Böhm et al., 2020).

The last step in the process is chatbot deployment to a channel. Resources adopted by chatbots to interact with users and accomplish their requests are called channels

(Singh et al., 2019). Conversational software development frameworks integrate various channels such as custom websites, phones, live agents, and popular social media networks. Moreover, each channel may support different interaction prospects that can be leveraged to achieve an effective chatbot solution. For example, chatbots can prompt the user through the channel to interact with the information system by providing a passage that can potentially reduce the number of errors. However, various channels may sustain distinct interaction mechanisms that bring both opportunity and risk (Perez-Soler et al., 2021).

With all the capabilities of current chatbot development platforms, choosing the right tool is challenging and requires a focus on different technical factors. The tools range from low-level natural language processing services helping in intents recognition and training phrases to comprehensive software solutions covering most phases of the chatbot creation process (Perez-Soler et al., 2021).

Perez-Soler et al. (2021) compare the leading available software for chatbot development as well as proposals from smaller companies specializing in conversational agents to create a selection tool for the most suitable solution. The researchers establish technical and managerial features for a total of fourteen different chatbot development tools. Technical features include input processing, dialog, deployment, system integration, development and testing, execution, and security. Input processing category supports evaluation of the following features: regular expressions, phrase patching with natural language processing, text processing for parameters extraction, number of languages, sentiment analysis, and speech recognition. In the dialog category, the research mentions volatile or persistent storage of phrase parameters, support for intents, support for predefined and user-defined entities, dialog structure, utterances to re-engage users, and specification of chatbot answers. The evaluation factors for the deployment include integration with social networks and websites and interaction support for specific social media networks. System integration properties for chatbot development tools assessment include a call to services from chatbot and chatbot usage via API. Development and testing capabilities consist of prebuilt components such as chatbot templates, intents, small talks and services, native or code-based version control, chat console for testing, debug mechanisms, and validation support. Execution features require hosted deployment, support for analytics, and user message persistence. The final category is security, and it consists of the security of cloud services (Perez-Soler et al., 2021).

Managerial features focus on organization, development experience, and operational factors. The organization category considers the pricing model of the chatbot development tool that can be free, pay-as-you-go, quota, or advanced for specific features. The organizational factor is also developer expertise required that is assessed as low or high. The development category includes code hosting, group work, internalization and localization support, and an open-source project. Operational features for evaluation are adding new channels and no vendor lock-in (Perez-Soler et al., 2021).

The comparison also includes the distinction of whether the software is a library, service, framework, or platform. Libraries and services support only specific steps in the chatbot lifecycle, usually related to natural language processing. Frameworks

provide sets of classes that require programming and custom code. Whereas most platforms are cloud-based, require little to no code, provide a ready-to-use interface, and frequently support the deployment of the chatbot to the channel (Perez-Soler et al., 2021).

Overall, the presented tools involve a wide spectrum of opportunities to ease chatbot development in different scenarios. However, designing, developing, and evaluating chatbots still brings some challenges. Most platforms provide general guidelines for the creation process, but quality metrics and design patterns are missing. The tools rely on training phrases to generate the intents which work for closed domains while supporting less constrained dialogs requires more sophisticated natural language processing mechanisms (Arora et al., 2019). Existing tools neglect the support for evaluating chatbots in an automated and systematic manner by incorporating a basic debugging system and console for manual testing (Perez-Soler et al., 2021).

Ultimately, the success of the conversational agent depends on its user experience and usability. Development tools should provide heuristics and guidelines for improved chatbot usability to complement the technical features such as natural language processing that enables more natural conversations. Parameters save users from the necessity of rephrasing a sentence multiple times for the same information or sentiment analysis that contributes to the chatbot's personality (Perez-Soler et al., 2021).

3 Chatbot development method

Since there is limited research on the specific chatbot development method (Folstad et al., 2021), we use the generic software development life cycle approach (Dawson, & Dawson, 2013). According to Dawson and Dawson (2013), all software processes can be described by the generic software development life cycle approach as they consist of the phases of requirements gathering, design, system development, deployment, and evaluation (Dawson, & Dawson, 2013). We have covered requirements development in the previous research and described some features in the case description and related research chapters. This part of the study involves the subsequent stages of the software development process, including design, development, and deployment. In the following sections, we describe the chatbot creation phases from the design of conversational flow and interface, through full-stack application development, to deployment of the solution.

3.1 Design

Perez-Soler et al. (2021) argue that chatbots require graphical user interface design and conversational design (Perez-Soler et al., 2021). In the following paragraphs, we describe the process of creating dialogs between users and a chatbot concerning the related research findings for human-chatbot interaction (Chaves, & Gerosa, 2020; Brandtzaeg, & Følstad, 2017). Furthermore, the atomic design approach to graphical

user interface design is applied to follow the findings from research on chatbot design elements (Jain et al., 2018, Janssen et al., 2020), a specific solution for Wikipedia editors (Skorupska et al., 2020), and universal user interface design practices.

Currently, conversations with chatbots are perceived as impersonal and unnatural. Enhancing more empathetic responses is possible by implementing an engaging communication style, which positively impacts people's perceptions of a chatbot (Kelleher, 2009). This approach to communication contributes to the persuasiveness of online communication between organizations and users, ad it is especially relevant for chatbot design and development. Liebrecht and van Hooijdonk (2020) describe different linguistic elements that can positively impact conversation design and improve chatbot perceived personality as more human. The researchers identify three main linguistic categories: message personalization, informal speech, and invitational rhetoric. The message personalization elements include greeting, addressing the user and organization, signature. Informal speech is characterized by abbreviations, non-verbal cues such as emojis or injections. Invitational rhetoric focuses on conversational features such as acknowledging, apologizing, expressing empathy, stimulating dialogs, joking, and well-wishing (Liebrecht, & van Hooijdonk, 2020).

Following the findings of related research (Liebrecht, & van Hooijdonk, 2020; Jain et al., 2018; Cameron et al. 2019) and chatbot design practitioners (Liebl, 2019), we established conversational flow for the chatbot, including components such as greeting, asking, informing, checking, error handling, apologizing, suggesting, and concluding the conversation. In the table below, we present the example elements of this conversational flow.

Table 2: Examples of conversation elements.

Conversation element	Example
Greeting	Hello 👋, I'm Wiki chatbot. I'm excited that you are a part of Wikipedia 🚀! How can I help you?
Asking	Alright! What exactly is the issue?
Informing	You can request a username change by filling in this form.
Checking	What kind of resources do you prefer? Article, video, or template?
Error handling	Could you rephrase your question?
Apologizing	I'm sorry, I don't understand. 😔
Suggesting	I can recommend to you a helpful article about changing the username.

Concluding	It was great to hear from you! Bye! 👋
Asking	Happy to help! Is there anything else I can do?

Since most of the studies focus on conversation design rather than graphical user interface in this design challenge, we apply the universal design guidelines for interface design and implement the findings from the study by Jain et al. (2018). Researchers in their study on evaluating and informing the design of chatbots provide design implications for chatbot user experience designers that consider combining text-based interface with interactive elements, enabling efficient input from users, providing a persistent view of chatbot capabilities and context, and providing effective chatbot discovery.

Combining textual interface with interactive elements such as buttons or media makes chatbot more engaging. However, redirecting to external resources needs to be handled with care since users express dissatisfaction when the chatbot opens a resource in a new browser tab. A possible solution is implementing the interactive element inline to show up within the frame of the conversational interface. When implementing additional features such as an extendable menu, the system should indicate to the user the functionality; otherwise, the item is neglected. Finally, due to the increase in horizontal scrolling functionality, the users are aware of it, and in the carousels, they expect more than five articles when browsing the chatbot's suggestions (Jain et al., 2018).

Enabling efficient input from users includes providing auto-suggestion buttons for efficient interaction. Users find this functionality time-saving; therefore, reducing the interaction cost. If the chatbot provides users with an option to select a parameter, it should also edit their selection. This is especially important when the edit affects a search query. Moreover, click functionality should be enabled on the entire element when it is interactive and not only on the button or link.

Providing a persistent view of chatbot abilities and context means avoiding a mismatch of expectations. The interface should provide a clear view of the chatbot's competence. A description of the chatbot should be included in the greeting or introduction but also persist throughout the conversation. Consequently, users can identify with the chatbot's contextual state and its presumptions to maintain the same state of mind for both dialog participants. Finally, providing effective chatbot discovery allows inexperienced participants to leverage the software.

During the design process, we incorporated knowledge gathered from related research about user experience (Følstad et al., 2021), human-chatbot interaction (Chaves, & Gerosa, 2020; Brandtzaeg, & Følstad, 2017), and usability (Borsci et al., 2021), as well as specific research on the graphical user interface (Jain et al., 2018).

The atomic design approach to design was followed in the study allowing to break down the complex concept of the chatbot into absorbable elements (Frost, 2016). The process consists of five stages: atoms, molecules, organisms, templates, and pages. The most elementary components are atoms, as they cannot be separated without

losing their function (Frost, 2016). Buttons, inputs, or avatars are examples of atoms in this project.

The combination of atoms creates molecules. They are elementary but can be further diminished (Frost, 2016). Examples of molecules can be a bot avatar connected with a message bubble, a text input field, or a send button.

The organisms are created from atoms and molecules. The organisms are aggregations of molecules that form a specific area that repeats in different interface areas. The characteristic feature of organisms in atomic design is that they are reusable. Due to this application, sections with similar content and structure can have a consistent presentation no matter where they appear in the interface (Frost, 2016). The header of a chatbot card illustrates the use of organisms as it consists of text, an avatar, and a close button icon.

Previously prepared organisms create templates that are assembled into specific layouts. Templates ensure consistency of the interface across different devices (Frost, 2016). The chatbot card is a template that includes a header, chat session, and footer.

The higher-level components with unique structures and layouts are called pages. The crucial aspect of the atomic design allows editing all elements after constructing a page and checking established assumptions (Frost, 2016). The home page of the chatbot is an example of a page in atomic design.

3.2 Development

In the following sections, we describe the development of the chatbot, including consideration about the chatbot development platform and the creation of a full-stack application. The app back-end uses Node with Express framework that integrates third-party API provided by IBM. The app's front-end is built upon React and Redux with React-Thunk for handling side effects. Finally, CSS preprocessor SASS is used for stylesheets writing.

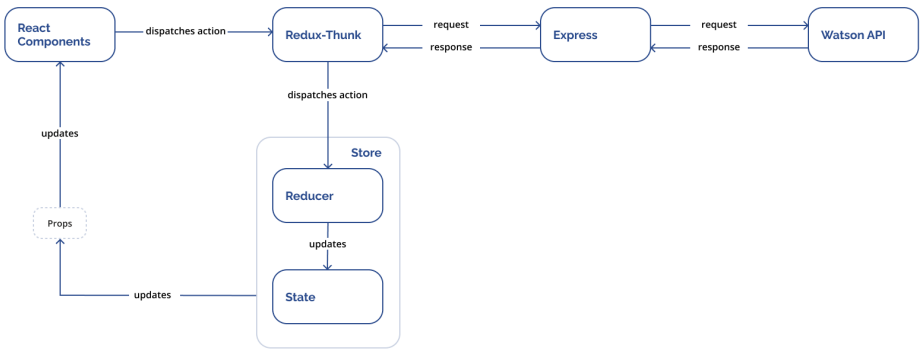

Figure 1: Fullstack app overview.

3.2.1 IBM Watson

At the beginning of the development process, we focused on selecting the most appropriate chatbot platform to satisfy all user requirements for Wikipedia editors and support developers with an efficient, well-documented, and cost-effective solution. The natural approach was the consideration of the most developed open-source platform called Rasa (Rasa, 2021), as it corresponds to the values of Wikipedia in making knowledge available to everyone (Wikimedia Foundation, 2021). In the comparative research of different platforms and frameworks by Perez-Soler et al. (2021), Rasa places seventh place in regard to technical and managerial features. However, the development experience for this platform is quite demanding (Perez-Soler et al., 2021), and with the lack of previous knowledge in the field, we decided to leverage one of the platforms with lower entry points for developers. The two platforms with the most technical and managerial features assessed in the mentioned study are Google Dialog (Google Cloud, 2021) and IBM Watson (IBM Watson, 2021g). As both platforms offer almost the same set of features, we selected IBM Watson based on the chatbot testing interface, cost-effectiveness, and availability of learning resources.

Chatbots can be customized by adding the skills that improve the experience for the users. IBM Watson provides different types of skills, including conversational skills and search skills. Conversational skills convey custom responses to common questions, whereas search skill explores provided datasets and returns response passages (IBM, 2021a).

Conversational skills focus on understanding and addressing requests or questions from the user based on the information provided by the chatbot constructor. The chatbot developer provides the information about the tasks or subjects and how users ask about them, and the Watson Assistant gradually builds a machine learning model tailored to understand the same or similar tasks. There are two types of conversational skills, including actions skills and dialog skills. Actions skills offer a straightforward interface for building conversational flow from scratch. The process of training the data happens automatically behind the scenes. Dialog skill provides a set of services that can define training data and the conversation flow. The conversation takes the form of a dialog tree. The graphical dialog editor is used to construct a script for the chatbot when interacting with the users. The dialog focuses on common customer goals based on the training you provide to respond with valuable responses (IBM, 2021b).

The action skill is the optimal approach for the chatbot focused on content as it offers the following benefits. Creating a conversational flow is simplified and does not require a deep understanding of machine learning. The actions setup provides a better overview of user satisfaction as they progress through a task that can be tracked visibly. Identifying obstacles or blockers is more straightforward as each action has a clear beginning and end. Moreover, the data does not need to be stored in any form of a database. By default, the data gathered during the conversation is not available, but certain types of information can be stored for the duration of the session, such as user names. Multiple people can work on the same skill simultaneously in self-contained,

separate actions in the actions skill setup. Only the order of the steps within one action matters, and all actions can be easily changed for optimal flow (IBM, 2021b).

On the other hand, the dialog skill is more appropriate for chatbots that require greater control of the logic of conversational flow. The dialog skill editor leverages the underlying artifacts of intents and entities used to apply the machine learning model. The dialog page corresponds to the if-else login known to developers but can confuse nontechnical chatbot creators. (IBM, 2021b).

Additionally, the functionality of search skills currently only available for premium users leverages information collected from existing knowledge bases or datasets to address more sophisticated user inquiries. Each query uses the IBM Watson Discovery service to explore a data source and return an appropriate answer. When the chatbot does not have an accurate solution, it redirects the user query to a search skill to find a response from the provided source of knowledge (IBM, 2021c).

As developing a chatbot is an iterative solution and the goal is to achieve a minimum viable product, we started by implementing the chatbot based on the actions skill. The long-term project objective was to leverage the search skill functionality and train the chatbot based on the Wikipedia help desk archives dataset (Wikipedia contributors, 2021).

Actions serve as the questions or tasks that chatbot can help the users solve. Each action has a clear beginning and an end. The action is initiated when the chatbot recognizes a goal based on the words in user input. The action ends when all the steps required for a satisfactory user experience are completed. The body of action consists of steps that extract information for the chatbot to accomplish the user's goal. IBM Watson provides built-in actions called system actions that automatically apply to the chatbot (IBM, 2021d). In the following paragraphs, we introduce some of the terminology used for IBM Watson Assistant development.

Steps are single interactions between a chatbot and a user. Each step defines one or more conditions that determine the processing of the step at a run time, the response to the user when the step is processed, rules for the user reply type, and the next interaction. Each step in action is numbered and processed from first to last. During action building and testing, the order of the steps can be adjusted for the best conversational flow (IBM, 2021d).

Step conditions determine how the chatbot knows whether to include a step in the current conversation with the user, such as asking a yes or no question before assuming the user's goal. When defining the user response for a step, it is necessary to identify the type of data that the chatbot expects to receive in return. So if the chatbot asks for a number, the customer response type would be a number as well. The chatbot is not able to complete a task on the user's behalf without the necessary information. The information that the user shares can be saved as a variable. The chatbot can reference the variable to refer to the user's message. Variables help personalize the conversation between the chatbot and the user. The lifespan of a variable is as long as the duration of a single action. However, Watson Assistant also provides predefined, built-in variables that can be used in any action. The variables set by the assistant include values like current time and date.

Moreover, there are session variables that are available across all actions. A session is defined as an instance of a conversation between the chatbot and the user. As the name suggests, this kind of variable is available for the duration of the single session. Dialog skills or client applications that call action can set the value of the session variable to be used by the action when it is triggered. Expressions are tools for processing and formatting data collected from the user. The expression can be used to do a simple calculation of tip recommendations based on the subtotal of the bill (IBM, 2021d). The following paragraphs explain how the conversations can be built with actions by adding an action skill to the chatbot and adding different actions.

Building a conversation flow is fundamentally about deciding which steps or user interactions are required to complete the action. An action skill contains actions that resemble the chatbot's tasks to help the users solve. Each action includes a number of steps that represent individual exchanges with the user. After the list of steps is identified, the focus shifts from the architecture of the conversation to the quality of content that can turn each interaction with the chatbot into a positive experience for the users (IBM, 2021e).

3.2.2 Back-end

The process begins by creating a back-end that will intercommunicate with Watson's assistant API core. Hence, Node.js was used, the open-source, cross-platform software that allows developers to create software in JavaScript running on the server-side (Node, 2021). Once the initial app installation was completed, the Node package manager was utilized to add the dependencies (NPM, 2021). In that matter, the Express framework that allows building applications based on Node.js was added. Express enables intermediary layers to respond to HTTP requests; it is characterized by considerable flexibility and provides a routing mechanism (Express, 2021). IBM Watson software developer kit was also used, which is the official SDK provided by IBM Cloud that servers apps built with Node and effortlessly integrates Watson Assistant V2 (IBM, 2022c). Nodemon development dependency was additionally introduced to the project to look after hot reloading and refreshing of the app (Nodemon, 2021). Lastly, Concurrently, a development library that allows running commands simultaneously was installed (Concurrently, 2021).

After installing the dependencies, the server.js file was created to gather and manage all of them. On top of the server.js role is to import the routes and prepend them with the session token from the Watson file assembled later on. However, first, to call the Watson API and obtain the session token, there are few elements needed, precisely the keys such as Watson API key, Assistant URL, and Assistant ID, which can be found inside the Watson assistant graphical user interface within assistant settings under the API details (IBM, 2022a). They need to be duplicated to an environment file that will make them available through the entire application; however, it will prevent us from accidentally deploying them to the GitHub repository, where they would become open to the public.

Figure 2: API details stored in IBM Watson GUI.

Following this, it is time to assemble back-end routes. When the communication occurs with Watson, it is required to accompany all messages with a session token. Otherwise, Watson Assistant will return error 500 and not react to user messages accordingly, making the chatbot purposeless. It is possible to use the *"createSession"* function from the Watson assistant package installed before this stage (IBM, 2022c). Inside the function, Watson Assistant ID is being passed as a parameter while dispatching a request to the API in order to receive a session token. A *"try-catch"* statement handles successful requests and returns the token id to the app's front-end. If an exception occurs, a catch block is executed where an error is thrown inside the console.

Subsequently, another route to take charge of incoming messages is being created. The *"message"* Watson API was used to carry a payload as the parameter, which consists of three things: Watson Assistant ID, a session token received from preceding API call, and input object that specifies the message type in that case, text and user message itself (IBM, 2022c). The Watson assistant will return a response message if those three parameters are assigned correctly (IBM, 2022a). That was the final step in developing a back-end for the chatbot application. Until this stage, it is possible solely to test it using Postman software. Using it makes it possible to test requests and ensure that endpoints were implemented correctly. Postman is one of the tools utilized for communication with the external API before the front-end is being developed (Postman, 2021). Nonetheless, for this application to be practical, it needs to execute the next step and implement the application's client.

```
// Import dependencies
const express = require("express");
const router = express.Router();
const AssistantV2 = require("ibm-watson/assistant/v2");
const { IamAuthenticator } = require("ibm-watson/auth");

// Create instance of assistant

// Authenticate
const assistant = new AssistantV2({
  version: "2021-06-14",
  authenticator: new IamAuthenticator({
    apikey: process.env.WATSON_ASSISTANT_APIKEY,
  }),
  serviceUrl: process.env.WATSON_ASSISTANT_URL,
});

// Route to session token
// GET /api/watson/session
router.get("/session", async (request, response) => {
  try {
    const session = await assistant.createSession({
      assistantId: process.env.WATSON_ASSISTANT_ID,
    });
    response.json(session["result"]);
  } catch (error) {
    response.send("There was an error processing your request");
    console.log(error);
  }
});

// Handle messages
// POST api/watson/message
router.post("/message", async (request, response) => {

  // Construct payload
  payload = {
    assistantId: process.env.WATSON_ASSISTANT_ID,
    sessionId: request.headers.session_id,
    input: {
      message_type: "text",
      text: request.body.input,
    },
  };

  try {
    const message = await assistant.message(payload);
    response.json(message["result"]);
  } catch (error) {
    response.send("There was an error processing your request");
    console.log(error);
  }
});


// Export routes
module.exports = router;
```

Figure 3: Backend routing architecture.

3.2.3 Front-end

Ahead of working on the application's front-end, we decided to choose one of the architecture patterns that will efficiently organize the codebase, ensure cohesion between modules, and provide clear direction as the application grows. One commonly used pattern is model view controller (MVC), an architectural paradigm

used to develop user interfaces developed in the late 1970s in Xerox laboratories by Trygve Reenskaug (Reenskaug, 1979). The model, view, controller architecture assumes the separation of three essential interconnected parts: the model, which is responsible for receiving and managing the data; it is essentially the brain of the application. The view displays the data and receives data from the user. It is what the user sees when they interact with the application. Lastly, the controller takes care of data traffic within the application, and it serves as a middleman between the model and the view (Reenskaug, 1979).

The main advantage of the MVC architecture is the differentiation of the design layers in the Model, View, and Controller for easier code reuse and code maintainability (Pop et al., 2014). Code duplication is limited in MVC because it disconnects data and business logic from the display. Since each logical part is separated from the other, a modification in one location does not require making multiple changes in other parts of the system (Dissanayake, & Dias, 2017). Because of the MVC, multiple programmers can simultaneously develop different parts of applications, making development much quicker. (Pop et al., 2014). Due to the benefits mentioned above and the subsequent stage that introduced a state management library - Redux. Following this architecture seemed like a natural step.

Redux is a React library that helps you manage the application's state in a standardized and predictable manner (Redux, 2021c; React, 2021a). Redux acts as a container for storing the state of the application. In practice, this state is represented by an immutable data structure. In the event of a change in the state of any of the components, a copy of the Redux state is created, including the change, and the affected components are notified of the change (Redux, 2021c).

Along with advanced applications development, there is an issue with passing props from parent components to children components. Redux flawlessly eliminates this nuisance and helps to separate the application state management layer from the component code (Redux, 2021c). Being acquainted with the MVC pattern it is possible to translate Redux modules into easily understood pieces where the model is comparable to incorporated reducers, wrapped in a centralized Redux store. The controllers are the Redux actions, and the Views correspond to standard React components.

Basic Webpack and Babel configuration are needed to get started with Redux development (Redux, 2021c). Fortunately, the React team has already prepared a boilerplate app that seamlessly fits this project's scope. A client directory was created for clarity and the explicit distinction between the front-end and the application's back-end. The "*npx create-react-app*" command was executed, which bootstrapped a lightweight version of the React app (React, 2021b). Before anything else, all of the unnecessary pre-installed files appended to the application when the default create react app was installed were deleted, and the application was left with only essential components needed for this project. Server worker, tests, stylesheets files, and user interface elements related to React were removed. From that, we could focus on all of the features needed for the app's front-end.

However, in order to avoid running the back-end and front-end each time separately, a few changes to package.json within the back-end API folder were made.

Two new scripts commands were added, first the "react" script to make it possible to run client-side from the back-end folder level. The second one was a "dev" script that starts both client and server sides simultaneously with the help of a Concurrently library.

```
"scripts": {
  "start": "node server",
  "server": "nodemon server",
  "react": "npm start --prefix client",
  "dev": "concurrently \"npm run server\" \"npm run react\"",
  "heroku-postbuild": "cd client && npm install && npm run build"
},
```

Figure 4: Helper script to run multiple commands simultaneously.

Once the app is assembled and operated correctly on both front-end and back-end, initial client development dependencies were added, such as redux and redux-thunk that makes working with application state management more convenient (Redux-thunk, 2021). The redux development tools extension allows previewing the state inside the Chrome development tools (Redux DevTools, 2021; Chrome, 2021). Following this, the Axios library was introduced, which helps make back-end API requests (Axios, 2021).

After having all the relevant packages and tools installed, it was time to develop. Firstly we focused on defining our model; therefore, the process began from determining the reducers and the store. The core of redux is the store, which is essentially the file where the entire application's state is stored and managed in the object tree inside a single store object (Redux, 2021b). The "*store.js*" file was created right within the client directory. The dependencies were imported afterward; the store was connected to redux development tools. Finally, the initial "store" state was defined as an empty object.

Moreover, the middleware that helps build action creators was introduced (Redux, 2021a). The store itself was established with the assistance of the *"createStore,"* method where multiple parameters need to be passed, such as a combined collection of reducers, initial application state, and *"composeWithDevTools"* (Redux, 2021a). Ultimately, the store was exported to be available across the application.

```
// Import dependencies
import { createStore, applyMiddleware } from 'redux';
import thunk from 'redux-thunk';
import combineReducers from './reducers';

// Connect the application to redux development tools
import { composeWithDevTools } from 'redux-devtools-extension';

// Setup initial state
const initialState = {};

// Import middleware
const middleware = [thunk];

// Setup store
const store = createStore(
  combineReducers,
  initialState,
  composeWithDevTools(applyMiddleware(...middleware))
);

// Export store
export default store;
```

Figure 5: Redux store.

Successively, the reducers directory was created to handle all redux reducers within the app. Reducer is a pure function that performs the appropriate modifications and returns the new state when called with the current state and action (Redux, 2021b). Since it is a function, it can also be used as a callback function that "store" will run when an "action" is dispatched. This function takes two parameters: state and action. When someone calls "action," the store object calls the reducer function, passing the current state and "action" to it; the reducer function checks the type of action transferred to it and returns a new version depending on the type of the state object. Reducers will primarily determine how to update data across the application (Redux, 2021b).

Inside the" queue.js" located in the reducers directory, the initial state of the messages queue is being defined as an empty array. Depending on what happens within the application, there can be different options. To handle those cases watsonReducer function is needed. It takes state and action as parameters and later unpacks the type and payload values using the destructuring assignment syntax on the action parameter. We also define messages that we get from the state. When having those clarified, we moved on to switch statements that will determine the response to

different types. The switch operator performs a similar task as the conditional *"if"* statement; it enables certain decisions depending on a particular condition (Mozilla Developer Center, 2021b). In the case of *"INPUT_SUCCESS,"* we submit a new user message as the last object in the queue and return the state. In case of the error, only the last state is being returned. Once all the scenarios for the previously defined actions are determined based on whether the action was a success, it is time to incorporate them. Inside the reducers folder, an additional *"index.js"* file is being established where the *"CombineReducers"* function is imported from redux. *"CombineReducers"* supplies the way to merge all of the reducers and export them as one to the store (Redux, 2021a).

```
import {INPUT_SUCCESS, INPUT_FAIL, SESSION_SUCCESS, SESSION_FAIL, MESSAGE_SUCCESS, MESSAGE_FAIL} from './types';

// Initial state
const initialState = {
  messages: [],
};

// Update state
export default function watsonReducer(state = initialState, action) {
  const { type, payload } = action;
  let { messages } = state;

  switch (type) {
    case INPUT_SUCCESS:
      messages = [...messages, { message: payload, type: 'user' }];
      return {
        ...state,
        messages,
      };
    case INPUT_FAIL:
      return {
        ...state,
      };
    case SESSION_SUCCESS:
      sessionStorage.setItem('session', payload.session_id);
      return {
        ...state,
      };
    case SESSION_FAIL:
      return {
        ...state,
      };
    case MESSAGE_SUCCESS:
      messages = [
        ...messages,
        {
          message:
          payload.generic.length >=2
          ? payload.generic
          : payload.generic[0].response_type === 'text'
              ? payload.generic[0].text
              : [...payload.generic[0].suggestions],
          type: 'bot',
          msgType: payload.generic.length >=2 ? "MULTIPLE" : payload.generic[0].response_type,
        },
      ];
      return {
        ...state,
        messages,
      };
    case MESSAGE_FAIL:
      return {
        ...state,
      };
    default:
      return {
        ...state,
      };
  }
}
```

Figure 6: Redux reducers.

When all the reducers are specified, the process continues with describing the controllers. As mentioned earlier, analogous to the controllers are Redux actions as they define changes to the application's state and are the fundamental communication mechanisms that represent unique events in the system and carry specific information with them (Redux, 2021c). They are triggers that can be dispatched from the front-end to reach the reducers. In this application, all actions needed to operate the application correctly live inside the *"queue.js"* file inside the actions directory. Along the *"queue.js,"* there are *"types.js."* Various types are based on action results and display them inside redux development tools; therefore, types are the flags that inform whether the action was finalized successfully or not (Redux, 2021b).

In order to make the chatbot usable, we needed to define a minimum of three actions. One that will handle user messages, another that will ensure we have a session token, and last that will ensure user input is sent to Warson API.

Initially, an asynchronous *"userMessage"* function was created that does not consist of any API calls. The function includes a try-catch block to take care of exceptions. Based on the success of the action, either the *"INPUT_SUCCESS"* or *"INPUT_FAIL"* type is being dispatched.

Following, an action that handles API calls to the back-end and eventually creates a session token was defined. However, it is essential to set up the proxy inside the *"package.json"* file to ensure that when the app receives a request, it will be by default rerouted to our back-end API, and there is no need to specify HTTP (Mozilla Developer Center, 2021c). From now on, it is possible to reach the API; therefore, the asynchronous function called *"createSession"* is defined where the call is being dispatched inside the try-catch block. The Axios GET method initiates a call to the back-end API, storing the value in the *"response"* variable (Axios, 2021). Assuming that the request was successful, the session success is dispatched along with the payload that equals data from the response.

The last controller needed for this project is the *"sendMessage"* action that sends a user message to the back-end API. We create a request that captures a message and stores it inside the body that points to the Watson message API. Following with the assistance of Axios, a post request is made that requires two parameters, the Watson endpoint URL and object that includes message input (Axios, 2021). The *"MESSAGE_SUCCESS*" is dispatched; otherwise, an error message is loaded.

```
import axios from 'axios';
import {INPUT_SUCCESS, INPUT_FAIL, SESSION_SUCCESS, SESSION_FAIL, MESSAGE_SUCCESS, MESSAGE_FAIL} from './types';

// Function that handles user message
export const userMessage = (message) => async (dispatch) => {
  try {
    dispatch({ type: INPUT_SUCCESS, payload: message });
  } catch (error) {
    dispatch({ type: INPUT_FAIL });
  }
};

// Create a session - API CALL
export const createSession = () => async (dispatch) => {
  try {
    const response = await axios.get('/api/watson/session');
    dispatch({ type: SESSION_SUCCESS, payload: response.data });
  } catch (error) {
    dispatch({ type: SESSION_FAIL });
  }
};

// Send the message to the Watson Assistant - API CALL
export const sendMessage = (message) => async (dispatch) => {
  try {
    const body = { input: message };
    const response = await axios.post('/api/watson/message', body);
    dispatch({
      type: MESSAGE_SUCCESS,
      payload: response.data.output,
    });
  } catch (error) {
    dispatch({ type: MESSAGE_FAIL });
  }
};
```

Figure 7: Redux actions.

Ultimately, with all your controllers ready, it is time to focus on views. The prior pre-installed boilerplate app consists of the *"App.js"* file, a root component of each React app (React, 2021a). At present, it is just the placeholder component that needs to be adjusted and connected with the store. In order to achieve that, the *"Provider"* was imported from react-redux along with the "store" that was set up in previous steps. *"Provider"* serves as a wrapper of the entire app and allows binding the store to the application. It is an auxiliary component that allows injecting the store into the application (Redux, 2021a). Having this at the top level of the application allows all the children components to access the Redux store. If this step were omitted, none of the components willing to communicate with Redux would work properly, and we would notice an error in the console stating that the component is unable to find the store object in context.

Another matter that must be applied is an automatic action dispatch that eventually returns the session token and stores it inside browser session storage (Mozilla Developer Center, 2021e). To handle that *"useEffect"* hook is needed (React, 2021c). The *"useEffect"* hook on each call clears session storage in case a previous session token still exists inside the browser storage. Then dispatches the *"createSession"* action to appoint a new token. IBM Watson makes sessions invalid if inactivity with the chatbot is lengthier than five minutes (IBM, 2022d). Hence, *"useEffect"* hook ensures that the chatbot session token restores after the session time is longer, when

the application loads, on each browser refresh, or when a page is being opened in a new tab.

Last but not least, a lightweight version of the chat was created and imported inside the App component to ensure everything ran as expected before creating more complex user interface elements. It consisted solely of an input field and div element that included user and bot messages. In order to handle user messages, the *"useState"* hook was used that temporarily captivated user messages inside the variable *"message"* and function *"setMessage"* that permits updating that message (React, 2021c). However, this action was performed exclusively behind the scenes, and the user had no way to witness what was happening without accessing Chrome developer tools. The input field needs to be changed to visualize those changes inside the browser window. In order to accomplish it, the *"onChange"* method was used so whenever the event fires, new information is set by the *"setMessage"* function, which accesses the target value. Following *"handleKeyDown"* function was created that targeted the *"Enter"* key, therefore initiating the function as soon as this particular key was pressed (Mozilla Developer Center, 2021a). That permitted the visualization of the message in the form of a text above the field in the window.

Lastly, the *"userMessage"* action was imported, which guaranteed that the message would be passed to the Watson assistant. After successfully sending messages to the Watson Assistant started returning messages that were pushed to the messages queue. To display each of them inside the window browser, it was needed to loop through them with the support of the native javascript map method (W3 school, 2021b). However, first, it was essential to ensure that the app would not fail in case the array was empty. The queue was checked using the logical *"&&"* operator, and if the array successfully passed the check, the object message was returned, including the type of message that can be either bot or user and the message itself.

```
import React, { useState } from "react";
import PropTypes from "prop-types";

const Bot = ({ chat, userMessage, sendMessage }) => {
  const [message, setMessage] = useState("");

  const handleKeyDown = async (e) => {
    const code = e.keyCode || e.which;

    if (code === 13) {
      userMessage(message);
      sendMessage(message);
      setMessage("");
    }
  };

  return (
    <section>
      <div>
        {chat.length === 0 && chat.map((msg) => <div>{msg.message}</div>)}
      </div>
      <input
        onChange={(e) => setMessage(e.target.value)}
        onKeyPress={handleKeyDown}
        value={message}
      />
    </section>
  );
};

Bot.propTypes = {
  chat: PropTypes.array.isRequired,
  userMessage: PropTypes.func.isRequired,
  sendMessage: PropTypes.func.isRequired,
};

export default Bot;
```

Figure 8: Initial bot structure.

3.2.4 Functional improvements

Previous steps allowed us to create a prototype that lacked any styling attributes and could handle only simple text messages like the bot response. This stage aims to enhance chatbot capabilities and introduce a design system.

One of the first things that need to be done is introducing SASS into the application. SASS is a CSS preprocessor that permits the use of a special syntax which is then compiled into regular CSS interpreted by the browser. In practice, this means that we get access to methods that permit styling views quicker and more conveniently, the code base is more transparent and readable (SASS, 2021b). There are two different syntaxes available for SASS development. First is the older version of SASS that uses intended syntax instead of curly braces. The second syntax used in this project due to similarity to CSS is SCSS, where curly braces and semicolons are still available (SASS, 2021b).

The fastest way to get SASS working in this application is to use the NPM package manager. To install node-sass, we need to run the following command in a terminal:

npm install node-sass. This package, later on, will take all SCSS files and process them during the react builds. After installing node-sass a main.scss file was created that will bundle all the SCSS created (SASS, 2021b).

When writing the stylesheets using preprocessors, we attempted to adhere to the do not repeat yourself (DRY) principle. This pattern focuses on avoiding repetition of the same piece of code in multiple places, which increases the probability of making a mistake and increases the time needed to make modifications if necessary (Microsoft, 2021). Therefore, the process started with declaring variables that will be repeatedly used in the application. SCSS variables are defined units with specific values that are reusable throughout the document (SASS, 2021a). They are set using custom property notation starting from the *"$"* sign. It is possible to use them as simple constants to represent colors, sizes, or as operators in mathematical expressions (SASS, 2021a).

```scss
// Colors Variables

// Basic Colors
$color-white: #ffffff;
$color-light-grey: #EAECF0;
$color-dark-grey: #54595D;
$color-black: #000000;

// Accent Colors
$color-primary-light: #EAF3FF;
$color-primary: #3366CC;
$color-primary-dark: #2A4B8D;

// Utility Colors
$color-success: #00AF89;
$color-warning: #FFCC33;
$color-error: #DD3333;
```

Figure 9: SCSS variables.

After declaring the variables, it is time to create a mixin library used in all project documents. Like variables, mixins are created to avoid duplication in the code and make it easier for developers to make any subsequent modifications (SASS, 2021a). In our case, mixins allow us to define styles for different device widths and ensure that the application is responsive. Although the use of mixins does not reduce the volume of the CSS sheet in any way, it significantly improves the comfort of work and the time spent on writing styles.

```scss
@mixin respond($breakpoint) {
  @if $breakpoint == sm {
    @media (min-width: 576px) {
      @content;
    }
  }

  @if $breakpoint == md {
    @media (min-width: 768px) {
      @content;
    }
  }

  @if $breakpoint == lg {
    @media (min-width: 992px) {
      @content;
    }
  }

  @if $breakpoint == xl{
    @media (min-width: 1200px) {
      @content;
    }
  }

  @if $breakpoint == xxl{
    @media (min-width: 1300px) {
      @content;
    }
  }
}
```

Figure 10: SCSS mixins.

In the next stage, we focused on expanding the chat module. The component consists of three elements: static which is a chat header and two dynamic features, a footer with an input field, and the main chatbot stream to preview messages queue. The chatbot header had no functionality other than being a presentational component. Hence all that was required was to specify classes with styling rules based on previously prepared variables and mixins.

Subsequently, we focused on the chatbot footer that required styling and additional functionality. We desired to allow users to send a message not exclusively by pressing the "Enter" key but also by clicking on the airplane icon. To achieve that, it was necessary to add the HTML *<button/>* tag that is a clickable HTML element (W3 school, 2021a) but also the introduced *"onClick"* event (W3 school, 2021c). While creating this feature, it comes up that the app is not handling the case in which the user tries to send a blank message. Therefore, a conditional check was added, making the *"sendMessage"* function disabled if the input field was empty.

Following we moved to the most complex module in the entire app. The primary purpose of this component was not only to display different types of messages but also to be able to interact with them and execute actions instead of using the input field. Each time Watson returns the message, it returns the message type, which can

be either text, image, option, suggestion, or search (IBM, 2022b). Text is responsible for the standard response when the assistant understands the user's intent. Image response returns the URL to the place where an image is being hosted. The options type is returned when the assistant understands the intent, but multiple roads can be taken from there; therefore, the assistant gives a user multiple choices that can be selected. The suggestion is returned when the assistant doubts the user's intent; thus, it returns multiple suggestions that might be matched. Furthermore, search is returned where premium search skill is utilized (IBM, 2022b).

As the first step, the focus was put on the most common type of message - text. By default, this message type is easier to handle; however, our intention was to expand its capabilities with different sup-types to allow responses that included links, code snippets, or video tutorials.

Of the three, IBM Watson Assistant partially supports only hyperlinks. Since the assistant's response is always a string, the hyperlinks are returned in markdown format where the link label is wrapped within square brackets, and the path itself is within parenthesis (Markdown Guide, 2021a). To make the app aware that the returned string consists of a link, each of them are being parsed with a Regex pattern (Mozilla Developer Center, 2021d). This process allows us to pull out a link and insert it between the HTML *<a>* tag. The next step was to differentiate the links between those that should redirect to another tab and those that should show video preview. From the perspective of the app, there is no difference between them. However, YouTube video URLs have a specific pattern that, when defined, can exclude video IDs with the help of Regex (Regexr, 2021). Lastly, we needed to determine how the app would comprehend that the response included a code snippet. Inspired by the IBM solution for hyperlinks, we settled to use markdown tags and wrap code snippets in them (Markdown Guide, 2021b). Once again, we could define the message as a code snippet with the Regex pattern. As a result, it was possible to differentiate four types of data: plain text, code snippets, hyperlinks, and videos.

Finally, the chat stream was made interactive so it could cover the following message types: options and suggestions. In order to do so, a separate component was created as options and suggestions differ both semantically and visually from regular text messages. Each option and suggestion needed to be clickable; consequently, they needed to be wrapped by a *<button/>* element. Button label was set based on what the assistant responded to; the same parameter was passed to the *"sendMessage"* function when the user pressed one of the options. Hence the last necessary functionality was added to a chatbot.

3.3 Deployment

The next step after building the chatbot is making it available to the users. There are a number of options for chatbot deployment depending on the specific need of the interaction. A channel is a communication passage where the chatbot is available for the user, such as a social media platform or custom website. The chatbot can be deployed to one or multiple channels. IBM provides integration for deployment to different channels, which work as an adapter that allows users to communicate with

the chatbot. The most popular integrations include web chat integration, phone integration, FacebookMessenger, Slack, Intercom, Twilio, Whatsapp, or custom client integration. Connecting chatbots to the custom website is possible through the IBM Watson API by providing correct API keys for the development and production environments that include Watson Assistant API key, URL, and id (IBM, 2021f).

Since the chatbot is a full-stack app that includes a front-end and a back-end integrated with the IBM Watson Assistant API, the deployment process needs to support building such a solution. As we have already used Git for version control in the development process, we decided to use a popular platform that enables developers to build, run, and operate applications in the cloud called Heroku (Heroku, 2021). The app is deployed by pushing the code to the Heroku-hosted remote associated with the application. The *heroku create* command creates a new empty app on Heroku and an associated Git repository. Running this command directly from the app's root directory automatically replaces the empty Heroku Git repository with the local repository. When the remotes are set up correctly, the application can be deployed using *git push heroku main* command to push the code from the local repository's main branch to the Heroku remote. This command can be used each time to deploy the latest version of the code to Heroku (Heroku, 2021). The process is straightforward; however, the setup required some additional adjustments. The first troubleshooting was necessary during the deployment process due to how the app is built with the client code nested in the app's folder with its package manager. The solution to this issue was adding a line to the main package.json that tells Heroku how to post-build the app by accessing the client folder, installing the dependencies, and building the app. The next issue that may occur in this type of application setup is a problem with running deployed applications in the browser. That is because Heroku does not know how to run the application at this point since locally, the app is initiated by the command node server or with the help of the node package manager by the command npm start. To solve this problem, the Procfile is required with the instructions for Heroku on how to run the app. In order to update the application, a new commit to GitHub is required for the app to run correctly in the browser. After deployment, the chatbot is available to everyone in production and can be evaluated by the users.

4 Chatbot evaluation method

As standardized methods to assess user experience with conversational interfaces are currently unavailable (Borsci, 2021), this research uses popularized methods for evaluating usability (Laubheimer, 2018) adapted for user interaction with a chatbot (Holmes et al., 2019). The study participants are asked to perform four tasks connected to Wikipedia editing in a concurrent think-aloud test, and the chatbot's usability is evaluated in the post-task and post-test questionnaires.

The tests were conducted as follows:

- Participants signed the consent form.

- Participants were given access to Wiki Chatbot.
- Test instructions were presented to the participants.
- Participants were audio-recorded, fulfilling the task. Participants talked through what they were trying to do during the task and what they could see on the screen.
- Participants responded to a post-task Single Ease Question. The questions were formed "Overall, this task was:" and answered by selecting a score on a seven point ranking from "very difficult" to"very easy."
- To collect qualitative data, participants were asked to provide feedback for why they answered with the specific score in the Single Ease Question.
- Finally, each participant completed a post-test usability survey Chatbot Usability Questionnaire (CUQ) developed for the chatbot usability evaluation (Holmes et al., 2019).

The following sections provide information about the study participants, research design, setup, procedure, and analysis.

4.1 Participants

Eleven people participated in the study, with 55% male and 45% female participants. The average age of the testers equaled 37, with the youngest participant 24 years old and the oldest 49 years old. The study was conducted with English-speaking participants from the United States, Canada, and Great Britain. Four testers came from the United States, two from Canada, and five from Great Britain. All the tests were conducted on the desktop devices with seven participants using Windows operating system, three macOS, and one Linux.

4.2 Research design

When selecting a research method for user experience design, we consider a three-dimensional framework described by Nielsen Norman Group that reflects on the following axes of research: attitudinal versus behavioral, qualitative versus quantitative, and the context of use (Rohrer, 2014). The distinction between attitudinal and behavioral dimensions essentially contrasts people's actions and behavior. While most usability studies rely on user behavior, some attitudinal research methods, including card-sorting, can provide insights about user mental models, or surveys can measure and classify attitudes to discover important user experience issues. Behavioral methods focus on gathering insights from users interacting with the service. However, some of the most popular methods, such as usability and field studies, lie between the attitudinal and behavioral research methods. They leverage self-reported and behavioral data (Rohrer, 2014).

Another dimension compares qualitative and quantitative research methods. The understanding of the qualitative method goes beyond the open-ended questions and generates data about attitudes and behaviors on observing the users directly, while quantitative studies gather the data indirectly, often through surveys or analytic tools.

Insights from qualitative studies are usually collected by observing how people interact with software and whether it fulfills their needs. The data analysis process is typically not mathematical. In contrast, insights from quantitative studies are usually based on statistical analysis since the data collection instrument captures a substantial amount of data that can be coded numerically. Considering the nature of each approach, the qualitative methods are more suited for searching for a cause of the problem or possible improvements, while quantitative methods answer how much or how many types of questions (Rohrer, 2014).

The final distinction is the context of use that considers how people use the product in the study. The context of use can be described as natural or near-natural, scripted, hybrid, or not using the product during the study. The goal of the research in natural or near-natural context is to minimize the interference from the study to understand the attitudes and behavior close to reality. A scripted study collects the insights on the specific aspects of the usage, for example, a newly implemented flow. The research where the service is not used is conducted to gather issues beyond usefulness and usability, such as cultural behavior. Most methods can move within different dimensions to satisfy the product development objectives (Rohrer, 2014).

The time dimension is another crucial distinction during the selection of research method, and it reflects on the phases of product development, including strategy, execution, and assessment. Strategizing means considering new ideas and opportunities reflecting on the business objectives at the early stage of product development. The goal is to explore and select new directions. The research methods for this phase can be both qualitative and quantitative and typically include field studies, diary studies, surveys, analytics, or data mining. Executing describes primarily qualitative research methods such as card sorting, field studies, participatory design, or usability studies. This is when the decisions are made and involves continually improving the design direction. The objective is to minimize the risks, inform and optimize the designs, and improve usability. The final phase is assessment when the product is available for real users, and the summative measurements can be done. The goal is to measure the performance of the product. The quantitative approach to assessment usually includes usability benchmarking, surveys, or A/B testing (Rohrer, 2014).

This study's user experience research method relies on a behavioral, qualitative approach in the context of a scripted study that falls into the timeframe of product execution with the focus on software evaluation and improvements. The chatbot evaluation process is conducted in the form of a usability test where participants are invited to perform a series of tasks in a remote, unmoderated session consisting of the think-aloud test, single ease question, and system usability score questionnaire adapted for chatbots development.

In the think-aloud test, the participants use the chatbot while continuously thinking aloud; therefore, they verbalize their impressions as they navigate through the interface (Nielsen, 2012). The procedure for conducting a test consists of recruiting participants, distributing representative tasks to be performed, and letting users express their thoughts. The think-aloud test benefits include cost-efficiency, robustness, flexibility, credibility, and easy setup and learning process.

Cost-effectiveness and robustness are great values of qualitative methods such as thinking aloud; however, the method is not sufficient for drawing up statistical data. Other disadvantages of this testing approach include putting participants in a strange situation, the possibility of filtered statements as users do not share their thoughts without reflection, and biasing user behavior through the moderator's insights.

After completing each task, the participants are asked to evaluate their experience in a post-task questionnaire. Each completed mission is evaluated with the post-task question designed as the Single Ease Question (SEQ). Measuring satisfaction with SEQ brings the benefits of recognizing the most problematic areas of the product and collecting fresh feedback and reactions from the users. The user's perception of usability is captured with the simple formula, including the sentence "Overall, this task was:." The participant evaluates the task from "very difficult" to "very easy" and gives it a number from one to seven. Moreover, the qualitative data may be gathered by asking the testers why they gave such a score to enrich the quantitative results (Laubheimer, 2018).

The Chatbot Usability Questionnaire (CUQ) was created to measure the user experience of chatbots with a special focus on usability. A conventional usability metric called System Usability Scale was the base for the questionnaire further modified to address the needs of conversational design. Chatbots are evaluated with CUQ concerning the user experience guidelines, including onboarding, understanding, navigation, intelligence, personality, and error handling (Holmes et al., 2019). The questionnaire consists of 16 questions instead of 10 to accommodate all chatbot user experience features, but it is still comparable to the system usability score. Testers evaluate the level of agreement with the individual statement on a five points scale beginning from "strongly disagree" to "strongly agree." (Holmes et al., 2019). Table X presents all statements about user experience with chatbots:

Table 3: Chatbot Usability Questionnaire (Holmes et al., 2019)

	Question
1	The chatbot’s personality was realistic and engaging.
2	The chatbot seemed too robotic.
3	The chatbot was welcoming during the initial setup.
4	The chatbot seemed very unfriendly.
5	The chatbot explained its scope and purpose well.
6	The chatbot gave no indication as to its purpose.
7	The chatbot was easy to navigate.
8	It would be easy to get confused when using the chatbot.

9	The chatbot understood me well.
10	The chatbot failed to recognize a lot of my inputs.
11	Chatbot responses were useful, appropriate, and informative.
12	Chatbot responses were irrelevant.
13	The chatbot coped well with any errors or mistakes.
14	The chatbot seemed unable to handle any errors.
15	The chatbot was very easy to use.
16	The chatbot was very complex.

4.3 Setup

Participants were invited to the chatbot evaluation through the Userbrain portal (Userbrain, 2021a). The testers were briefed about the study objective and that the research was voluntary and anonymous. They agreed to record the test according to the terms and conditions of the platform (Userbrain, 2021b).

4.4 Procedure

The testers were invited to participate in the study through the Userbrain portal (Userbrain, 2021a). The test consists of two parts. Users are asked to think aloud during the four tasks and later during the questionnaire consisting of sixteen short statements. Each mission ends with a post-task assessment to measure satisfaction and is carried out with the Single Ease Question (Laubheimer, 2018) and an open question for collecting qualitative feedback. The system captures the screen recording of each participant and their answers to each question while the users complete the tasks. The collected metrics help assess the usability and user experience of the software. The questionnaire used to collect post-test results is called Chatbot Usability Questionnaire (Holmes et al., 2019). The study took place on 20 December 2021. All the responses are gathered in the form of a screen recording with the voice-over of the participants. The shortest recording took 4 minutes 59 seconds, and the longest 22 minutes 59 seconds. On average, the completion time of the test equaled 12 minutes 7 seconds. The accumulated total time of the study reached 2 hours, 13 minutes, and 27 seconds.

4.5 Analysis

For the analysis of the collected data, the research leverages thematic analysis. It is a method of analyzing qualitative data often used on textual content such as

transcripts. The objective of the thematic analysis is to closely examine collected data to identify themes, ideas, and patterns emerging repeatedly (Nowell et al., 2017). A popular approach to thematic analysis follows six steps established by Clarke and Braun (2016), including familiarization, coding, searching for themes, reviewing themes, defining and naming themes, and synthesizing collected data in the written report (Clarke, & Braun, 2016).

The advantage of the thematic analysis is the flexibility it offers researchers to interpret the data and organize it easily by using broad themes. On the other hand, it carries the risk of missing the essential nuances in the data, focusing on subjective interpretations relying on the researcher's judgments. Therefore, it is important to reflect carefully on the proposed interpretation of the data set (Clarke, & Braun, 2016). In the following sections, the subsequent phases of thematic analysis are described.

4.5.1 Familiarization

The first step of thematic analysis is getting accustomed to the data collected in the study. A thorough understanding of the data is necessary before analyzing individual features. A common approach to familiarization is the careful process of transcribing collected audio material. It requires a scrupulous verbatim transcription of all verbal utterances. The core of this exercise is collecting data that is true to its original nature and not only retaining necessary information (Clarke, Braun, & Hayfield, 2015).

Engaging in the analysis may happen over the self-collected data through interactive means such as interviews, podcasts, or different recordings. Therefore, the analysis starts with prior knowledge of the data and initial analytic considerations. Regardless of the origins of the data, the researcher must immerse in the data to the extent that they are truly familiar with the contents. This kind of immersion usually requires repeated data reading to search for meanings and patterns. However, the task is time-consuming and tempting to quickly read through or neglect some of the fragments it is not advised. This phase of thematic analysis should focus on collecting initial thoughts and notes before coding in subsequent phases. Once this is over, the formal coding process begins and continues through the entire analysis (Clarke, Braun, & Hayfield, 2015).

To ensure that we immerse properly in the data collected in the think-aloud test, we first watched recorded screen-sharing sessions collecting notes, then created transcripts, and finally reshaped them a few times to explore different perspectives. Apart from the simple transcript following the order of each recording, we created separate files for each task before coding for more precise comparison. Furthermore, for the coding purpose, we reshaped the transcripts again into tables divided by timestamps to synthesize the findings from each piece of content efficiently.

While working with verbal content, the speech needs to be transcribed into written form to conduct thematic analysis. Transcribing itself is an opportunity for familiarization with the collected information. Some researchers argue that transcribing is the crucial phase of interpreting qualitative data that creates meanings

rather than just a mechanical act of writing down the spoken texts (Bird, 2005). There are no strict guidelines for the transcription in thematic analysis apart from retaining the information collected in the study according to its true nature. In some cases, transcribing the data saves time as it allows actual familiarization with the contents requiring additional time if the data has been already provided. Therefore, the transcriptions themselves can be used in the analysis to inform the early stages of considerations (Clarke, Braun, & Hayfield, 2015).

We use the OtterAI tool (OtterAI, 2021) for automatic transcript generation that can be easily reviewed by the researcher and adjusted to capture all verbal and nonverbal utterances during the transcription process. Using the software speeds up the process; however, it still requires careful re-reading and re-listening of the collected data and collection of potentially essential pieces of information. While reviewing the generated transcript, the researchers collect user feedback from each task.

4.5.2 Coding

The next step in the thematic analysis is coding the data. Codes are primary segments of raw information that can be evaluated in a meaningful way to describe the phenomenon (Clarke, & Braun, 2016). Coding essentially means highlighting the sections in text and defining labels for the selected fragments. The process should be thorough, each transcript should be coded, and every potentially important or interesting sentence labeled (Clarke, Braun, & Hayfield, 2015).

The coding begins when the researchers are familiar with the data and have generated the initial ideas about the data and what makes it interesting for the analysis. The next step in the process is the production of initial codes. The codes can be organized in meaningful groups; however, the themes that emerge from them are often even broader. The codes themselves do not require interpretative analysis in the next phase when the themes are generated. Coding depends to some extent on the perspective of the thematic analysis. Some themes may be more data-driven while others theory-driven (Clarke, Braun, & Hayfield, 2015).

Therefore, the codes may emerge from the interpretation of the data or predefined assumptions deriving from theories and literature. It is essential to work systematically through the entire data set, give equal attention to each piece of information, and identify the emerging patterns early in the process. A popular approach to the manual coding process is highlighting the content with color and labeling it with the same color code. The codes can be defined early and then sorted into different extracts or assigned gradually. Regardless of the method, towards the end of the coding process, all the extracts should be collected with corresponding code (Clarke, Braun, & Hayfield, 2015).

Clarke, Braun, and Hayfield provide advice to code for multiple potential themes and patterns, keep the context of each extract, and code individual extracts into as many different patterns as they fit into not only one category. The outcome of the

coding process is the thematic map that shows different relations between extracts and supports the final narrative of thematic analysis (Clarke, Braun, & Hayfield, 2015).

4.5.3 Generating themes

After coding, it is time for generating themes. The themes are the patterns identified from the codes and grouped together. At this stage, it is essential to assess the quality of the codes so the more vague or irrelevant codes can be discarded. Some of the codes can become themes on their own. The generated themes can be flexible as they are only potential to be reviewed in the next phase of the thematic analysis (Clarke, Braun, & Hayfield, 2015).

This phase starts with the collected data and the codes created in the previous step. The focus moves from the individual codes to resolving potential themes. Essentially, the analysis starts with considering how each code combines with the others and creates a theme. Visual representations such as tables, mind-maps, or posters may help sort the codes into themes. That is when the analysis focuses on the relationships between codes, themes, and different levels. Some of the initial codes can be transformed into main themes, sub-themes, or completely discarded. At this point, some of the code may not belong anywhere, and it is acceptable to sort them into the theme of miscellaneous items and revisit them at a later stage (Clarke, Braun, & Hayfield, 2015).

Potential themes should be collected at the end of this phase, and all extracts of data coded according to them. The sense of the significance of each theme should start emerging. Although, the researchers should not abandon any ideas without further review. However intuitive, generating themes based on the problem statement may lead to subjective analysis of the collected data (Clarke, Braun, & Hayfield, 2015). Therefore, in the next phase of thematic analysis, we review the selected themes.

4.5.4 Reviewing themes

When reviewing the themes, it is essential to compare them to the raw data collected in the study to ensure that they accurately represent the information. At this point, the researchers should modify problematic themes by combining them, splitting them up, discarding, or creating entirely new ones. As a result of this phase, the researchers should have a clear idea of the different themes, how they work together, and what story they tell about the data. The final thematic map should be created to show the relationship between different themes (Clarke, Braun, & Hayfield, 2015).

This phase of the thematic analysis starts with refining the generated themes. It becomes evident that some of the proposed themes are not supported enough, whereas others could be merged or further broken down. The information gathered in each theme should be coherent, and the distinctions between themes should be clear. The process involves two stages of reviewing and refining the themes. The first one is focused on reviewing the code extracts and requires reading all extracts within the team again to ensure that they create a consistent pattern. Suppose the extracts are not

sorted coherently; they should be revisited. The particular extracts might be in the wrong theme, should be discarded entirely, or a theme itself might be problematic. Once all the extracts are correctly sorted, the analysis may move on to the next level that involves similar refinement but about the entire data collection. A thematic map should be created to show the relationship between different themes (Clarke, Braun, & Hayfield, 2015).

At this point, the validity of individual themes is considered regarding the theoretical and analytical approach. The collected data should be revisited to ensure that the themes represent the data set and code the remaining information that could be missed in the previous stages. The necessity of re-coding is natural and expected as it is a continuous organic process. The need for re-coding should be assessed based on the thematic map developed so far. The process continues until the thematic map is satisfactory and all the relevant themes are covered (Clarke, Braun, & Hayfield, 2015).

4.5.5 Defining and naming themes

We can define and name the final list of the themes available. Defining themes focuses on formulating precisely what they represent and how they help understand the information collected in the study. Naming involves creating easily understandable and concise names that give a reader a sense of the theme. (Clarke, Braun, & Hayfield, 2015). In this stage of thematic analysis, we came up with the following themes: positive chatbot user experience and negative chatbot user experience.

4.5.6 Producing the report

The final phase of thematic analysis is writing the report about the developed themes. The goal of the task is to present the story that truthfully represents the collected data and convinces the reader about the validity and contribution of the analysis. The analysis should provide a coherent, logical, and interesting story deriving from the data based on the developed themes. Choosing particularly eloquent examples that capture the essence of the argument should support selecting the themes and provide sufficient evidence without redundant complexity. The extracts from transcripts should be easily identifiable as representatives of the issue and embedded in the analytic narrative. The story should go beyond simply describing the findings yet present arguments for established research questions (Clarke, Braun, & Hayfield, 2015). In the results section, we attempt to compose such a narrative to fulfill the requirements of thematic analysis.

5 Results

The following passages present the findings from the development and evaluation of the chatbot solution for Wikipedia editors. The insights contain the results about the usability of the chatbot and the necessary improvements needed to enhance the editor experience on Wikipedia. The qualitative nature of the study is supported by numerical data collected through the Single Ease Question and Chatbot Usability Questionnaire. The research culminates with the thematic analysis of the data collected in the think-aloud test. Full study transcripts of all participants mentioned by name and timeframe are available in the Appendix.

5.1 Single Ease Question

The think-aloud test started with the following task: "You opened an account on Wikipedia years ago, and you want to change your silly password. Ask the chatbot for help". The participants perceived the usability of the chatbot on average at 6.81 out of 7 points from "very difficult" to"very easy." The numerical value and user feedback describe the task as very easy. All the users succeeded in finding the answer, with some of them using the predefined categories (Rich, 0:42, Jessie, 0:58) and others simply typing their question (Ian, 0:30, Tomeo, 0:41). However, one of the participants pointed out that it was confusing not to see an appropriate category in the predefined options. The possibility of using the input field was not clear enough (Latasha, 1:30). We could observe how people learn to interact since it was the first contact with the chatbot. None of the users had problems opening the chat by clicking the button "Let's chat," one of the participants appreciated that the chat view opened in the middle of the screen instead of the bottom right corner (Israel, 1:26). Some participants expressed difficulties understanding how to initiate the contact with the chatbot (Israel, 1:26) for example, one of them kept typing the word "help" before asking any specific question (Jacob, 0:52), and another asked themselves how do I speak to this chatbot (Thara, 0:57). Some of the most common adjectives describing the first interaction included easy (Israel, 3:34) and quick (Elisabeth, 3:01).

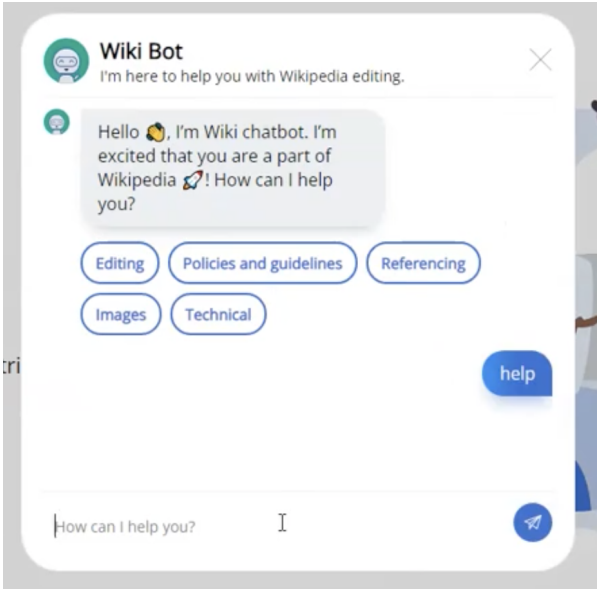

Figure 11: A participant typing the word "help" to initiate the conversation.

The second mission included a more complicated task to find out what to do about the content used without permission allowed testers to take different paths to learn about copyright violations or plagiarism. The task achieved the perceived usability score of 5.36; however, 36% of the participants evaluated it as very easy. The open question answers showed that the users were getting more confused with a more complex task. Ten out of eleven people completed the task according to the script. Some of them appreciated predefined options (Tomeo, 2:28), but others could not categorize their intent to any option (Steven, 5:16, Tokoni, 1:21). With the task requiring more interaction, some user interface issues emerged, such as the inability to clear the chat without refreshing the page (Steven, 3:55). The most alarming feedback was the doubt whether the answer from the chatbot was helpful (Jacob, 1:47). Moreover, three testers encountered serious difficulties either understanding the task (Steven, 5:16), Wikipedia policies regarding copyright (Israel, 4:35), or categorization (Latasha, 2:04). The difficulties expressed by the participants saying for example, "I'm having trouble with this one, to be honest" (Steven, 5:16) or "I can tell you this, that this task was hard for me." (Israel, 7:40) may indicate that real users would feel frustrated trying to find information about policies and guidelines.

The third task corresponded to the category of referencing, and the participants were asked to find out how to add a footnote to the article they were writing. The task received a perceived usability score of 5.54. Although the correct answer according to the script was covered by the answer to "How do I cite sources," a few users felt that they could receive the information regarding footnotes in a more general article regarding editing, which is also correct (Rich, 4:09, Ian, 2:41, Jessie, 2:44). Some of

the participants tested multiple options offered by the chatbot, including the article about editing, a more specific article about referencing for beginners, or a code snippet offered by the chatbot (Ian, 2:41). The respondents appreciated the multiple options of solving the problem as they experimented with the chatbot (Steven, 7:11, Tomeo, 3:25). In this task, the participants started to learn how to use the chatbot in the most effective way (Israel, 8:51).

On the other hand, the participant noticed that the task was confusing because he lacked the mindset of the Wikipedia editor (Israel, 11:08). One of the users expressed that the more interactions with the chatbot, the more complex the task when they encountered the buttons offering three different options to choose from (Thara, 6:25). Some of the users assessed the interaction with the chatbot and the quality of the information provided and expressed the need for summarized knowledge rather than only redirecting to relevant articles or media (Latasha, 4:43).

The goal of the fourth mission was to add a new category on Wikipedia. The only task that could not be solved with predefined options offered in the chatbot and required typing the questions to the chatbot. The mission to add a category to an article received an overall score of 6.45. Multiple responses to an open question described the assignment as easy; moreover, the motive of multipath functionality of the chatbot was revisited (Ian). At this point, it can be observed that the participants have learned their ways to interact with the chatbot mentioning that they need to enter specific keywords (Tokoni, 3:07) while others keep typing full sentences in natural language (Elisabeth, 8:10).

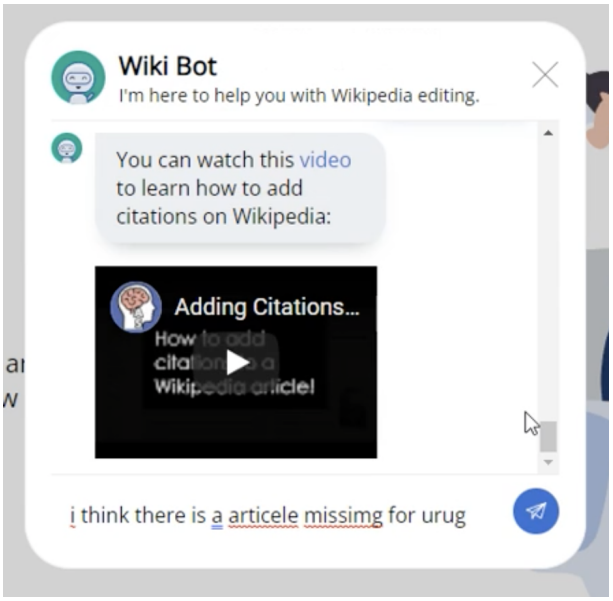

Figure 12: A participant typing the question in natural language (Elisabeth, 8:10).

Some of them mention that they have learned how to interact with the chatbot at this point, saying, "Okay, I'm getting used to the to the question that I have to make to get good answers." (Israel, 12:19) and continues when rating the task with "No, no, now I'm getting used to the chatbot, and I can tell you what it was easy. I will give. I'm going to give. I'm giving the rate six to this question because it was easy. But just because I'm getting used to use the chatbox and let's go for the next step." (Israel, 13:13).

5.2 Chatbot Usability Questionnaire

The last segment of the study was a Chatbot Usability Questionnaire (CUQ) composed of sixteen questions that were subsequently accumulated and calculated into the overall score defining the general usability of the chatbot. Eleven participants took part in the questionnaire. 82% of them agreed or strongly agreed that the chatbot's personality was engaging and realistic. On the other hand, 9% of participants strongly agreed that the chatbot was too robotic, although 63% disagreed with that statement. Overall, participants considered the chatbot welcoming during the initial setup 55% giving it the highest score. Respondents almost unanimously disagreed with the statement that the chatbot seemed very unfriendly, whereas 91% selected this option. 82% of the testers thought that the chatbot explained its purpose well, and 72% of them strongly disagreed that it did not demonstrate its purpose sufficiently. The participants found that the chatbot is easy to navigate, where 72% of testers gave the highest score. 64% of them strongly disagreed that it would be easy to get confused when using the chatbot. 34% of users strongly agreed that the chatbot understood them well, while all the others scored from four to five. 82% of respondents strongly disagreed with the statement that the chatbot failed to recognize many of the inputs. 91% of participants found the responses useful, appropriate, and informative and the same number disagreed regarding the irrelevance of the responses. 36% of the testers were uncertain whether the chatbot handled errors sufficiently; however, the remaining testers either agreed or strongly agreed. 55% strongly disagreed that the chatbot appeared unable to manage errors. 91% of respondents found the chatbot very easy to use and disagreed that it is complex. The chatbot earned the lowest usability score of 70.31 and the highest score of 95.31. The final mean CUQ score was 84.80, and the median equaled 85.94. The mean CUQ score is analogous to the System Usability Score (Holmes, 2019) and converts to the adjective description of excellent (Bangor, Kortum, & Miller, 2009).

As for the qualitative feedback provided by the participants regarding the Chatbot Usability Questionnaire, the chatbot's personality was assessed as realistic (Thara, 9:22), and some did not feel like they were talking to the bot (Rich, 8:11). The testers were divided regarding the robotic nature of the chatbot. Some people felt definitely aware that they were communicating with the chatbot (Tokoni, 4:22, Thara, 9:59); however still found it natural (Tokoni, 4:22) and intelligent (Thara, 9:59). More extreme responses described interaction with a chatbot as almost human-like (Tomeo, 5:36), while others felt like the chatbot had no personality at all (Latasha, 6:33).

All the testers agreed that the chatbot was welcoming during the initial setup and expressed opinions such as: "And I knew exactly what I was supposed to do after I read the introduction. So that's great." (Thara, 10:28). The same tester found the chatbot "humanizing" when asked about the unfriendliness and disagreed with the statement. Other participants supported this view by saying, "I didn't have the sense that it's not a person." (Israel, 16:10). The statement encouraged participants to share some additional feedback about the quality of messages as they should provide more specific information and context (Latasha, 7:11).

On the other hand, some participants did not expect to get in-depth information from the chatbot when asked about explaining the scope and purpose (Tokoni, 6:31), and others felt fairly comfortable with the provided information (Rich, 8:55). Although the numerical responses show that participants did not feel like the chatbot did not indicate its purpose, some did not record it stating its purpose (Tomeo, 6:31).

All participants agreed that the chatbot was easy to navigate with statements like "All I had to do is ask questions, and I got my answer. "(Thara, 12:52). When assessing the statement about easily getting confused when using the chatbot, one of the users tested the chatbot by typing a random mix of characters and noticed that "Yeah, even when it doesn't understand it's very direct about that." (Ian, 8:00).

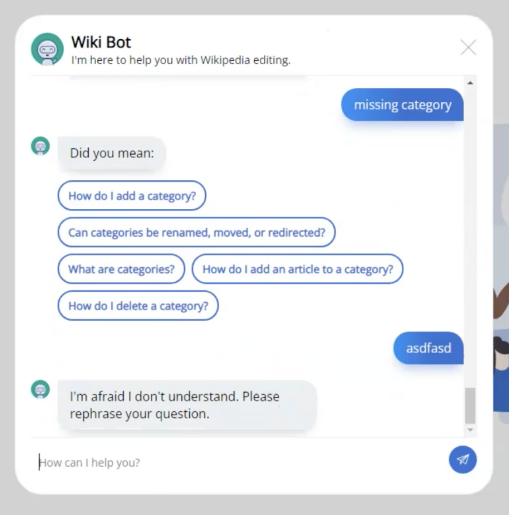

Figure 13: A participant testing the chatbot with a random content (Ian, 8:00).

The participants felt that the chatbot understood them well and described that "usually within two tries, whether it was putting in a keyword and selecting something within two actions, I was typically able to find what I was looking for." (Tokoni, 6:04). Furthermore, regarding failing to recognize user inputs, some of the users felt that even if it did not have a specific understanding, it would provide them with multiple options to choose from (Tomeo, 7:39). Similar statements were recorded when users were asked whether chatbot responses were useful, appropriate, and

informative, with users appreciating the number of options provided (Steven, 12:02). One of the participants stated that the chatbot provided links, information and showed empathy (Thara, 14:51). Moreover, when asked about the relevance of the responses, a user stated that he had the sense of a person behind the chat (Israel, 20:12).

The participants felt that the chatbot coped well with errors or mistakes, but some suggested adding a category that would allow them to ask something other than predefined options or state that none satisfies their needs (Ian, 10:19). Participants seemed to have issues with understanding the question "The chatbot coped well with any errors or mistakes." as they assumed it was about their own mistakes rather than the chatbot's functioning (Jacob, 6:28, Steven, 12:29, Latasha, 9:36).

All of the participants stated that the chatbot was easy to use and was not complex, with one of the testers explicitly explaining, "I thought it's a complex system, but I don't think it was complex to use." (Tomeo, 9:16).

5.3 Thematic analysis

In the following sections, we present the process of thematic analysis conducted on the data collected in the think-aloud test that covers familiarization, coding, generating themes, reviewing themes, and formulating the narrative.

5.3.1 Familiarization

The first step of thematic analysis is getting accustomed to the data collected in the study. The familiarization started from watching screen recordings and further moved on to the creation and initial interpretation of the transcripts from the study.

We used the OtterAI tool (OtterAI, 2021) for automatic transcript generation that the researcher can quickly review and adjust to capture all utterances during the transcription process. The initial feedback from the usability test tasks is presented in Table 4.

Table 4: Familiarization with transcript from usability test tasks.

Task	User feedback
You opened an account on Wikipedia years ago and you want to change your silly password. Ask the chatbot for help.	- "Personally, I thought that was very simple and very intuitive." (Tomeo, 0.41) - "So I think for the most part it is fairly easy, but because there are like limited with there's like with respect to editing, policies and guidelines, referencing, images, and technical, I wasn't really sure if I can put in some more thing that

	didn't fell within, or I didn't think fell within there." (Latasha, 1:30)
You are a blogger and you noticed that someone used your content on Wikipedia without permission. Use the chatbot to find out what to do about it.	- "I'm not sure if they helped me out here." (Jacob, 1:47) - "But I did really like that they have multiple options to make sure that I get the right information based on what I was asking." (Tomeo, 2:28) - "I just I think it would be easier if I could like start over." (Steven, 3:55) - "The answer wasn't in the sort of options at the top. So it was a case then of framing your question correctly." (Steven, 6:26) - "And all I had to do was ask my question. And it gave me a series of related topics that I could click on. And I was able to click on the first topic and it gave me the exact answer that I was looking for. So great." (Thara, 4:14)
You are writing an article about a football player and you think that there is a category missing for Uruguayan football players. Ask the chatbot if it can help you.	- "So it seems that there's multiple ways of solving this problem, depending on the kind of level of detail that I'm looking for." (Ian, 2:41) - "You gave me multiple different options that are most helpful based on my preferred method of use." (Tomeo, 3:25) - "But I was quite happy to sort of write out a question after the last task, and that seemed to work very, very well." (Steven, 7:11) - "No, no, now I'm getting used to the chatbot and I can tell you what it was easy." (Israel, 13:13) - "I don't mind necessarily being redirected to another page. But I feel like I prefer to kind of get a little bit more like a summary as to what I can do within the chatbot. And then if I needed further information, I can

	actually go to the article about that " (Latasha, 4:43)
You can see that one of the articles is missing a footnote. Find out how to add it.	- "It provides me this information in multiple ways here too is so that I can build up. I do like this that it offered kind of new multi-path directions in how I would find the solution" (Ian, 4:25)

While collecting the feedback from the test, we focused on capturing different aspects of user experience that could help answer the research questions and collecting contradictory or more extreme opinions to analyze them thoroughly in the following phases of thematic analysis. Table 5 below gathers some of the most interesting comments from the familiarization process with the transcript from the Chatbot Usability Questionnaire.

Table 5: Familiarization with transcript from Chatbot Usability Questionnaire

Chatbot Usability Questionnaires (Holmes et al., 2019)	Feedback
The chatbot's personality was realistic and engaging.	- "I didn't get a feel that I was actually talking to a bot, I could have very easily been talking to person on the other end." (Rich, 8:11) - "It was definitely realistic for a chatbot." (Thara, 9:22)
The chatbot seemed too robotic.	- "It is literally a robot AI thing so I don't expect it to tell me jokes, but it seems like the right amount of robotic and that it gave you what you needed but it didn't feel uncanny" (Tokoni, 4:22) - "It felt like I was talking to a person actually, to some extent" (Tomeo, 5:36). - "I definitely know that I'm talking to a bot. And when the bot says like, Did you mean and then gives me like a list of topics. I know it's definitely a bot, but it's not too bad. It's definitely a very intelligent bot." (Thara, 9:59)

	- "I didn't feel it had much of a personality." (Latasha, 6:33)
The chatbot was welcoming during the initial setup.	- "And I knew exactly what I was supposed to do after I read the introduction. So that's great." (Thara, 10:28)
The chatbot seemed very unfriendly.	- "Yeah, like, when I asked about copyright violations, it was like, We take this very seriously. It's not something a bot would say. It's very humanising. So that's great." (Thara, 11:00) - "I didn't have the sense that it's not a person." (Israel, 16:10) - "I would expect it within the chatbot to kind of get a lot more information and context. So I don't necessarily have to review a full page as to how to get to everything." (Latasha, 7:11)
The chatbot explained its scope and purpose well.	- "I felt fairly comfortable. I was at the right place to find out information. " (Rich, 8:55) - "It didn't go very in depth, but at the same time, I don't expect the chatbot to give me several paragraphs of explaining. I think it's very much, oh, here's the basics. You'll figure it out as you use it, which is how I expect the chatbot to work in the first place." (Tokoni, 4:57)
The chatbot gave no indication as to its purpose.	- "So I don't think it ever stated its purpose. But personally, I didn't think it needed to." (Tomeo, 6:31)
The chatbot was easy to navigate.	- "All I had to do is ask questions and I got my answer. " (Thara, 12:52)
It would be easy to get confused when using the chatbot.	- "Let's let's just type in some random thing and see what it does there. Yeah,

	even when it does understand it's very direct about that." (Ian, 8:00) - "I think for the most part, it was very intuitive." (Latasha, 8:01)
The chatbot understood me well.	- "Yes, usually within two tries, whether it was putting in a keyword and selecting something within two actions, I was typically able to find what I was looking for." (Tokoni, 6:04)
The chatbot failed to recognize a lot of my inputs.	- "If it didn't have a specific understanding, they would give me multiple solutions" (Tomeo, 7:39)
Chatbot responses were useful, appropriate, and informative.	- "And in most cases, he gave you more than one option, as well to you so that that was good." (Steven, 12:02) - "It provided links and information and it showed empathy." (14:51)
Chatbot responses were irrelevant.	- "I had this the sense of a live in person behind the chat." (Israel, 20:12)
The chatbot coped well with any errors or mistakes.	- "I think maybe you could add a it's something else category" (Ian, 10:19) - "They would give me suggested answers and suggested questions based on that, and gave me my information accordingly" (Tomeo, 8:24)
The chatbot seemed unable to handle any errors.	Participants have issues with understanding the question "The chatbot coped well with any errors or mistakes." as they assume it's about their own mistakes (Jacob, 6:28, Steven, 12:29, Latasha, 9:36).
The chatbot was very easy to use.	- "I strongly, strongly agree. There is nothing to say about it. Next, [...], why I said there is nothing to say it's because I am glad to use it is just definitely good." (Israel, 21:43) - "I felt it wasn't clear that I can kind of type in something" (Latasha, 10:04)

The chatbot was very complex.	- "I thought it's a complex system, but I don't think it was complex to use." (Tomeo, 9:16)

5.3.2 Coding

The next step in the thematic analysis is coding the data. We followed the popular convention described by Clarke, Braun, and Hayfield (2015) by placing the transcripts in tables and highlighting different pieces of information with colors later connected to the code labels. To ensure that all the pieces of information are covered in the coding process, all the irrelevant speech, such as the participants reading aloud the tasks or contents of the chatbot, was highlighted with light grey color. This way, we could quickly identify each highlight corresponding to the code. Additionally, each extract contains the approximate timeframe ensuring that we could get back to the recording and easily find the relevant fragment.

Table 6: Example of coded transcript extract from Tester 1 (Rich, 08:55).

Transcript extract	Codes
"The chatbot seemed very unfriendly. I'm gonna say strongly disagree I didn't get that vibe at all. The chatbot explained its scope and purpose. Well, yeah, right from the beginning, it said, you know, again, welcome. It was it clearly mentioned Wikipedia. And it wanted to know how I could help you and even gave you some sub menus, you know, edit, editing, policies, referencing images and technical. So I did. I felt fairly comfortable. I was at the right place to find out information."	- friendly - engaging, welcoming personality - predefined options functionality - feeling comfortable with using the chatbot - informative

After highlighting various phrases and color-coding corresponding labels, all the codes are grouped. It provides a condensed summary of the main findings gathered in the process and assigns common meaning deriving from the data collected (Clarke, Braun, & Hayfield, 2015). At this point, all the codes were loosely gathered in the common table to provide a clear overview of the initial coding outcomes. A short example table from one of the transcripts is presented below.

Table 7: Initial codes from Tester 1 grouped together (Rich)

- curious to use the chatbot - personality - engaging, welcoming - feeling comfortable with using the chatbot - gave indication about community and functionality - easy to use	- unsure about using chatbot - chatbot not explicit regarding its purpose - the chatbot could be less confusing - users confusing chatbot's errors with their own	- easy to find information - helpful information - provided relevant information
- no technical issues - user wants to start over - typing functionality - predefined options confused the user - predefined options functionality - no loops in the conversational flow - provide the predefined options also later in the flow	- open option if it doesn't fit in the predefined categories	

The initial groups were created following the research questions and attempting to assess the usability and user experience of the chatbot. Therefore, green codes represented positive user experiences and purple codes negative user experiences. At this point, a clear pattern regarding the informative character of the chatbot emerged and was grouped separately to determine later if it could create an independent theme. Orange code collected the extracts considering the existing functionality of the chatbot mentioned by the participants, while red code informed about the missing features.

5.3.3 Generating themes

After collecting the long list of the codes sorted into categories based on the initial idea, we searched the themes and patterns emerging from the collected data. At this stage, we still followed the direction drawn by the research questions regarding user experience and usability. We could see clear indications towards some of the user experience characteristics described by Morville (Morville, 2004), such as usefulness (Tokoni, 4:45), findability (Israel, 17:42), or accessibility (Steven, 10:36). Furthermore, participants noticed some of the features implemented to support the

usability of the chatbot (Jain et al., 2018) and suggested more improvements. During the process, we could see some loose ends of the generated themes as the informative character of the chatbot was still a prominent theme (Tokoni, 5:20, Jacob, 5:33) and some behavior of the users suggested the need for a separate theme for interactions that would cover feeling comfortable (Rich, 4:09) or unsure (Ian, 1:22) when using the software. However intuitive, generating themes based on the problem statement may lead to subjective analysis of the collected data (Clarke, Braun, & Hayfield, 2015). Therefore, in the next phase of thematic analysis, we review the themes selected in this step.

5.3.4 Reviewing themes

In the refinement phase of the previously prepared themes, we noticed that the interactions could be further divided into social and functional characteristics following the studies by (Chaves, & Gerosa, 2020; Brandtzaeg, & Følstad, 2017) instead of placing all the interactions together and keeping the separate theme only for information theme. In the process, we created the table attached below that consisted of four themes: user experience (Morville, 2004), usability features (Skorupska et al., 2020, Jain et al., 2018), social characteristics of human-chatbot interactions (Chaves, & Gerosa, 2020), and functional characteristics of human-chatbot interactions (Brandtzaeg, & Følstad, 2017). All the tables illustrating the process consist of the codes from only one transcript (Rich) to keep the report concise.

Table 8: Generating themes from code groups (Rich).

Codes	Themes
- personality - engaging, welcoming - gave indication about community and functionality - easy to use - chatbot not explicit regarding its purpose - the chatbot could be less confusing	User experience (Morville, 2004)
- no technical issues - user wants to start over - typing functionality - predefined options confused the user - predefined options functionality - no loops in the conversational flow - provide the predefined options also later in the flow	Usability features (Skorupska et al., 2020; Jain et al., 2018)

- open option if it doesn't fit in the predefined categories	
- curious to use the chatbot - unsure about using chatbot - personality - engaging, welcoming - not explicit regarding its purpose - the chatbot could be less confusing - users confusing chatbot's errors with their own - feeling comfortable with using the chatbot	Social characteristics of human-chatbot interactions (Chaves, & Gerosa, 2020)
- easy to find information - no technical issues - user wants to start over - no loops in the conversational flow - provided relevant information	Functional characteristics of human-chatbot interactions (Brandtzaeg, & Følstad, 2017)

Finally, as a result of continuous revision and refinement of the themes we concluded that in order to determine user experience aspects of the chatbot that need improvements we should formulate positive and negative user experiences emerging as themes from the study. In the table below presents consolidated codes from all the transcripts falling into two themes of positive and negative chatbot user experience.

Table 9: Reviewing themes from code groups.

Codes	Themes
- informative and relevant content - availability of information - provided guidance - provided resources - usability - usefulness - credibility - accessibility - findability - learnability - clear purpose and limitations - user engagement - user satisfaction	Positive chatbot user experience

- user awareness - personality - friendliness - understanding of utterances - intent recognition - use of predefined options - use of related topics - use of variety of media - multiple flow routes - clarifying questions - lack of technical issues - error-handling - instant responses - interface design	
- superficial information - necessity of rephrasing - language adjustments - user confusion - neutral or robotic personality - unexplicit regarding its purpose - excessive redirects - lack of miscellaneous option - lack of chat exit or restart - problems with natural language understanding - issues with error-handling - lack of the integration with ecosystem - functional bugs - lack of personalization	Negative chatbot user experience

In the final revision of the themes, we started composing a thematic map that would guide the thematic analysis. As usability is a part of user experience as a whole (Morville, 2004), we decided to analyze it as a sub-theme of user experience. Additionally, we divided both social and functional characteristics into positive and negative aspects affecting the user experience of the chabot for Wikipedia editors. Therefore, we could conclude with only two themes representing summarized positive and negative user experience emerging from the thematic analysis. The final thematic map is illustrated below.

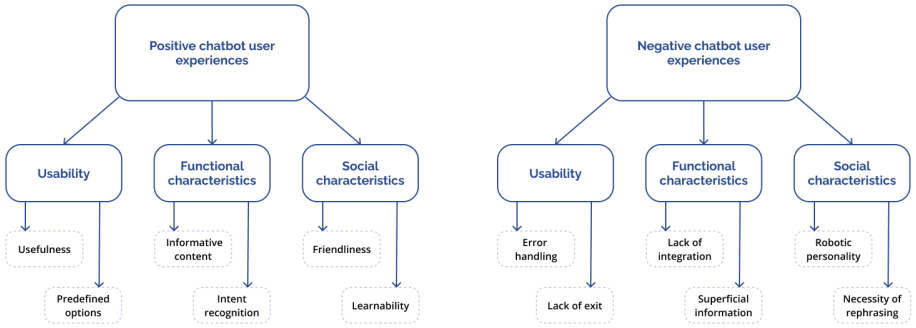

Figure 14: Thematic map of positive and negative user experiences with the chatbot.

5.3.5 Defining and naming themes

In this stage of thematic analysis, we came up with the following themes: positive user experiences and negative user experiences. On a more detailed level, each of the themes covers functional (Jiang, & Banchs, 2017; Maslowski et al., 2017) and social characteristics (Chaves, & Gerosa, 2020; Avula et al., 2018; Wallis & Norling, 2005) of human-chatbot interaction, as well as specific usability characteristics of the conversational agent (Borsci et al., 2021). The productivity, instant support, and assisting the editors in task completion by providing the necessary knowledge base should be functional characteristics of an efficient solution for Wikipedia editors (Brandtzaeg, & Følstad, 2017). The social characteristics should support the information search with proactivity, damage control, manners, and emotional intelligence (Chaves, & Gerosa, 2020). The usability features of the chatbot should cover easy access and appropriate location in the interface, the ability to create fair expectations, or recognition and facilitation of user goals (Borsci et al., 2021). Depending on the evaluation of the characteristics mentioned above, the sub-themes for positive and negative user experiences with the chatbot were developed.

5.3.6 Thematic analysis report

As developing chatbot solution for Wikipedia focuses on the domain of information search (Skorupska et al., 2020; Følstad et al., 2021; Wikipedia Contributors, 2021d), the social characteristics of human-chatbot interaction include proactivity, damage control, manners, and emotional intelligence (Chaves, & Gerosa, 2020). Therefore, in the thematic analysis to evaluate the user experience with the chatbot, we interpret the social interactions regarding generated themes of positive and negative user experiences.

The proactivity of the chatbot can be manifested by the conversation initiation, which was positively evaluated by participants, with some of them highlighting the

friendliness of the chatbot expressed in the use of emojis (Ian, 5:59). Moreover, the theme of topic suggestions represented by the predefined categories was recognized by testers (Thara, 4:14, Elisabeth, 1:40). The proactive characteristics of providing additional details and asking follow-up questions were visible when the chatbot offered different resources (Latasha, 8:44). Proactivity helps users realize that chatbot suggestions help in onboarding (Chaves, & Gerosa, 2020), which users expressed with sentences such as "Now I'm getting used to the chatbot, and I can tell you that it was easy." (Israel, 13:13). Some of the challenging aspects of proactivity can be relevance and timing (Chaves, & Gerosa, 2020), both especially important in the context of Wikipedia editing since the editors' motivation to use the chatbot is productivity (Brandtzaeg, & Følstad, 2017). Although the chatbot gained a high score in the Chatbot Usability Questionnaire regarding the relevance of the information, some alarming signals are emerging from the qualitative data. A few respondents expressed doubt whether the information provided by the chatbot was relevant (Jacob, 1:47; Latasha, 8:44), and one of them admitted that it is hard for them to judge the quality of the information as they are not experienced in Wikipedia editing (Israel, 11:08). In the case of timing and instant responses, we can observe that some participants notice how quickly they can get the information with the chatbot (Elisabeth, 3:01). The researchers (Chaves, & Gerosa, 2020) suggest leveraging conversational context to increase the productivity of the chatbot, for example, by providing a set of valuable topics which performed well in the chatbot evaluation. The users commented on this feature often by saying, "I thought it was very engaging gave me multiple solutions" (Tomeo, 3:25) or "The options were good, they were, the options were clear, and they related to the task at hand, and like I said before, but it is just very very straightforward and simple" (Steven, 0:53). However, some of the participants mentioned that predefined categories confused them (Israel, 1:26), and others had issues understanding whether they can use typing functionality if the predefined options are provided - "For me, I feel like I couldn't necessarily type in password just because those [predefined options] are here." (Latasha, 0:26).

The damage control was not appropriately evaluated as many participants did not recognize any errors during the conversation (Thara, 16:03). In contrast, others confused the chatbot's inability to handle errors with their own, for example, by saying, "That second question that I had trouble with, but that was my fault." (Steven, 12:15). The positive example of the chatbot regarding damage control can be observed when one of the users tried typing a string of random letters, and the chatbot reacted appropriately according to the user who said, "Yeah, even when does not understand, it is very direct about that." (Ian, 7:59). Moreover, some users perceived the chatbot as a software advanced enough to handle errors by saying, "I feel like the chatbot would handle errors pretty well." (Tokoni, 7:11). The available strategy for chatbots to manage difficult situations is politeness (Chaves, & Gerosa, 2017). In this regard, most of the participants agreed that the chatbot was friendly, with only one person stating that "I think it was, it was neutral. I did not think it was overly welcome, overly friendly" (Latasha, 7:00).

Although chatbots do not have a genuine ability to express feelings, it is essential to create a sense of emotional intelligence in their conversational style (Chaves, &

Gerosa, 2020). The software should recognize users' emotions, demonstrate empathy, and improve human-chatbot interactions (Chaves, & Gerosa, 2020). In regard to Wikipedia, the improved engagement deriving from interacting with users with emotional intelligence could support changing the new editors into capable and retained contributors (Bryant, Forte, & Bruckman, 2005). In the conversational design, the improved engagement is backed by the capability to sustain extended conversations, let users provide feedback, and use the interface components (Chaves, & Gerosa, 2020). Even though the chatbot is not capable of recognizing or expressing a lot of emotions at the moment, some of the participants mentioned its empathetic statements by saying "it showed empathy" (Thara, 14:51) or described them as humanizing - "Yeah, like, when I asked about copyright violations, it was like, we take this very seriously. It's not something a bot would say. It's very humanizing." (Thara, 11:00). Some of the participants had the feeling of talking to a person when interacting with the chatbot, saying, "I had this the sense of a live in-person behind the chat." (Israel, 20:12) or "Like I was talking to some real person." (Steven, 10:17). Other testers clearly stated that it is acceptable for the chatbot to act differently than a human by saying, "It is literally a robot AI thing, so I do not expect it to tell me jokes, but it seems like the right amount of robotic and that it gave you what you needed, but he did not feel uncanny if that makes sense." (Tokoni, 4:22). However, some areas still need to be improved as the users noticed repetitive answers - "I definitely know that I am talking to a bot. And when the bot says like, Did you mean and then gives me like a list of topics. I know it is definitely a bot, but it is not too bad. It is definitely a very intelligent bot." (Thara, 9:59) or no personality at all - I did not feel it had much of a personality." (Latasha, 6:20).

Følstad and Brandtzaeg (2020) describe the functional characteristics of the chatbot focusing on positive and negative user experiences. The positive experiences include usefulness, support in task finalization, or the perks of receiving general knowledge (Følstad and Brandtzaeg, 2020). In the context of Wikipedia editing, the chatbot should provide relevant information to help complete a specific task. The findings from the study suggest that the chatbot is perceived as useful, and the sub-theme of providing the information is prominent in the responses of many participants (Tokoni, 1:21, Tomeo, 3:25, Steven, 9:12). The participants recognize that the chatbot offers access to knowledge (Steven, 10:51), information (Tomeo, 3:25), specific examples (Thara, 7:28), and different learning resources (Jacob, 7:26). However, there are cases of doubt in the quality of the information provided (Jacob, 1:47) or describing it as "somewhat irrelevant" (Ian, 9:54).

The negative user experiences with chatbots described by the researchers focus on poor usability, particularly the issues with understanding utterances and intent recognition (Følstad, & Brandtzaeg, 2020). The overall usability of the chatbot for Wikipedia received a high score in both Single Ease Questions and Chatbot Usability Questionnaire; however, the qualitative data also supports this claim with the strong sub-theme for usability (Steven, 12:37, Jessie, 0:58). Participants describe it as being intuitive - "Personally, I thought that was very simple and very intuitive." (Tomeo, 0.41) and easy to use "I've used chatbots with a number of sites. I think this is one of the easiest ones. You're not constantly doing guesswork. (Tokoni, 5:28). The

participants generally appreciate the level of intent recognition with words like - "The fact that it was able to recognize that I wanted to change my password and provided me a direct link to it rather than just stating it, makes it, it did exactly what I wanted it to do." (Ian, 0:58) or "it did work out when I typed what I was looking for" (Elisabeth, 13:04).

The chatbot encountered challenges mentioned in the research specific to developing a solution for Wikipedia that suggests recovery techniques to improve user experience in case of chatbot's failure (Skorupska et al., 2020). The chatbot failed to recognize user input that was written descriptively, and even though it offered suggested topics as a reply, it performed with tiring repetitiveness (Thara, 9:59). Moreover, one of the participants encountered an issue with the recovery technique providing users with multiple response types. When trying to switch from one response to the other, the chatbot did not react, and only repeating the question and choosing one of the options again helped. The participant described it in their own words as follows - "Okay. Article. So they can even show civility to a code snippet. That's pretty useful. It's a code snippet. So the code snippet really helps. No, wait, no, hold on. That's not good. They were supposed to show us the citation in the code snippet, but it couldn't do that. Hold on, so that's interesting. Okay, there we go to now it works, so it's a bit of an error here, just FYI." (Jacob, 7:26).

Regarding the usability of the chatbot, the solution can be evaluated based on the quality attributes introduced in the related literature (Borsci et al., 2021). The ease to start the conversation (Borsci et al., 2021) by the chatbot was evaluated positively by multiple participants (Tokoni, 0:54, Thara, 10:54), with two of them mentioning the welcoming personality based on the use of emojis Ian, 5:59, Tomeo, 5:49). Access to the chatbot was easy for the participants as none of them had issues with starting the chat by clicking the "Let's chat" button (Elisabeth, 1:40). Moreover, one of the participants (Israel, 1:26) explicitly appreciated the location of the chat view in the interface (Borsci et al., 2021). In the view of some participants, the chatbot created fair expectations regarding its capabilities (Bostrom, & Yudkowski, 2014) as it expressed not understanding some inputs and asked for rephrasing (Ian, 8:16). The chatbot can preserve some of the contexts from the conversation, for example, choosing a preferred variant of information (Jacob, 7:26); however, it is not fully capable of maintaining a themes conversation (Radziwill, & Benton, 2017). The chatbot is not able to sustain the reference to the system (Staven, 2017) as it is not integrated with Wikipedia as of now. On the other hand, many participants still evaluated chatbot responses directing them to specific spaces on Wikipedia as useful (Thara, 5:47). All participants claimed that the chatbot was able to recognize their intents (Ian, 0:58), and the information provided was relevant (Tomeo, 8:14). However, one of the testers had doubts whether the advice from the chatbot was beneficial (Jacob, 1:47), and others noticed that the responses get "somewhat irrelevant" (Ian, 9:54) when the chatbot encounters issues with understanding the natural language (Wilson et al., 2017). While some users (Thara, 7:28) recognized that the chatbot was able to respond with quality support without adding too much information (Borsci et al., 2021), one user felt like this amount of information was insufficient and would rather receive all the knowledge from the chatbot that external

page (Latasha, 8:44). As stated before, the users (Ian, 8:16) recognized that the chatbot could respond appropriately when it is not equipped to handle the conversation or request (Morrisey, & Kirakowski, 2013; Vetter, 2002). All the participants agreed that the chatbot was friendly (Tomeo, 6:02) and therefore carried responses with acceptable manners (Morrisey, & Kirakowski, 2013). During the study, we could observe the perceived conversational capability (Tokoni, 6:49) and speed of answer, which provide chatbot with the capability of delivering informative responses (Jessie, 0:58, Elisabeth, 3:01) in a timely manner (Borsci et al., 2021). The chatbot's attribute of meeting neurodiverse needs (Radziwill, & Benton, 2017) could be briefly expressed by the user stating, "The chatbot seemed very unfriendly, definitely not... managed to get the answers that I wanted. And I'm quite sensitive. If somebody has been unfriendly or something, you know, like that. And I didn't feel like that at all." (Steven, 10:36), which results in an accessible environment for people with diverse needs regardless of their age, health condition, or well-being (Borsci et al. 2021). Finally, at this point, the chatbot is not capable of managing and adapting to different conversational styles of the users (Meira, & Canuto, 2015), and this aspect requires further improvements.

As Kvale et al. (2020) formulate the characteristics of successful and unsuccessful chatbot dialogs in chatbot training, their work can be used to evaluate the solution for Wikipedia. The essential findings for successful conversational flow refer to the chatbot's linguistic elements mirroring the user's vocabulary, and this study shows the need for further improvements as chatbot has limited understanding of natural language, for example, not recognizing the phrase "stolen" as "used without permission" (Ian, 10:45) or having trouble in understanding the user typing full sentences (Thara, 1:43). The chatbot recognized user input regardless of the misspelling or mistakes (Elisabeth, 13:04). However, as mentioned before, it is not integrated with the ecosystem of Wikipedia (Kvale et al., 2020). While intent recognition issues or missing content typically motivated the suggestion for improvements in the related research (Kvale et al., 2020), the proposals from the participants covered the volume of the information provided (Latasha, 8:44) or the ability to choose the predefined option formulated as "Other" (Ian, 10:19).

Overall, the positive user experience regarding the usability of the chatbot raised in the study includes essential functionality such as understanding utterances (Thara, 14:32) and intent recognition (Ian, 0:58). Additionally, participants mentioned features such as typing (Rich, 6:33), predefined options (Jessie, 0:58), links (Tomeo, 0:41), suggested topics (Jacob, 5:23), multi-path functionality (Elisabeth, 13:04), or interactive resources (Latasha, 8:44).

Moreover, the qualitative data can be interpreted following the user experience characteristics such as usability (Jakob, 0:32), usefulness (Tomeo, 3:25), credibility (Tokoni, 4:22), or accessibility (Steven, 10:36).

The positive social characteristics of user experience represented in the study refer to users learning how to use the application (Israel, 12:19, Steven, 11:16), feeling comfortable (Tokoni, 5:28) and curious (Rich, 4:09) when using the chatbot, appreciating the welcoming (Thara, 10:24) and friendly (Tomeo, 6:02) features. Some

users recognize the chatbot's personality (Israel, 16:12) and appreciate natural interaction (Tomeo, 5:36).

The positive functional characteristics represented in the qualitative data collected in the study focused on providing helpful information (Steven, 10:07), relevant (Israel, 19:20), and can be validated in the ecosystem (Thara, 5:47). Moreover, testers appreciate the flexibility of the chatbot (Tomeo, 7:15), provided guidance (Thara, 10:24), and the lack of technical issues (Rich, 1:53) or loops in its functionality (Rich, 10:08). They expect the chatbot to cope well with potential errors (Tokoni, 6:49).

On the other hand, the negative user experience concerning usability refers to the issues with error handling (Jacob, 7:58) or misunderstanding user input (Ian, 10:45). The users lack the functionality to clear the chat and start over (Steven, 3:58) or the possibility to select a miscellaneous category so the chatbot could guide them better (Latasha, 1:30). Some of them would like to see the predefined options later in the conversational flow (Rich, 10:29).

The negative social characteristics of user experience are reflected in the feeling of confusion when interacting with the chatbot through predefined options (Latasha, 8:13) or keywords (Thara, 13:02). To some participants, the chatbot seems not human enough (Thara, 13:02), too robotic (Latasha, 6:20), or neutral (Ian, 5:59). One of the participants suggested that the chatbot is missing personalization (Latasha, 6:39).

While the negative functional characteristics include the necessity of rephrasing (Steven, 6:26), figuring out the most helpful category (Latasha, 8:13), or using keywords (Tara, 13:02) to adjust the language to the chatbot (Tokoni, 6:16). Some of the participants had doubts whether the chatbot provides enough information (Israel, 8:51), relevant knowledge (Jacob, time), or synthesis of the most important aspects (Latasha, 5:11).

In the following chapter, we summarize the findings from the study, describe the usability and user experience evaluation, and seek ways to improve the user experience aspects of the current chatbot solution for Wikipedia editors.

6 Discussion and Conclusion

The research examines how to develop and evaluate a conversational interface to improve editor experience on Wikipedia. The process corresponds to the research directions described by Følstad et al. (2021) as chatbot frameworks and platforms and chatbot user experience and design (Følstad et al., 2021). Chatbot frameworks and platforms focus mainly on developing conversational agents, including designing, developing, and deploying the solution we followed in this research. User experience and design encourage initiatives evaluating chatbot applications and improving user experience (Følstad et al., 2021). This study describes the creation and evaluation of the chatbot solution designed to improve editor experience on Wikipedia. The findings from the study support the research questions raised in the paper as follows:

RQ 1: How usable is the chatbot solution for Wikipedia editors?

- In the Chatbot Usability Questionnaire, the application achieved a mean score of 84.80, corresponding to excellent usability. The participants provided high scores for the sentences describing the chatbot as easy to use and navigate, while the chatbot's responses were helpful, appropriate, and informative. On the other hand, the respondents disagreed that the chatbot failed to recognize many of their inputs or that it would be easy to get confused when using the chatbot. The sentence about the chatbot's personality being realistic and engaging uncovered the necessity for improvements in this area.
- In the thematic analysis of qualitative data collected in the study, the positive user experience regarding usability included considerations about understanding utterances, intent recognition, features such as typing, predefined options, links, suggested topics, multi-path functionality, or interactive resources.
- The negative user experience concerning usability refers to error-handling or misunderstanding user input. The users lack the functionality to clear the chat or select a miscellaneous category. The predefined options were missing at the later stages of conversational flow.

RQ 2: What user experience aspects of the chatbot need to be improved?

- Based on the Chatbot Usability Questionnaire chatbot's personality should be improved as, at times, it was perceived as neutral or robotic.
- The thematic analysis described the negative social characteristics of user experience reflected in the feeling of confusion when interacting with the chatbot. The qualitative data confirmed the numerical data collected in the Chatbot Usability Questionnaire that highlighted the chatbot's personality issues. The chatbot seems not human enough, too robotic, or neutral to some participants. One of the participants suggested that the chatbot should be more personalized.
- The negative functional characteristics include the necessity of rephrasing or using keywords to adjust the language to the chatbot. The chatbot provided questionable information or inefficient synthesis of the essential information.
- The overall results suggest the necessity for improvements in the development of the chatbot's personality, the content selection that would accommodate the needs of Wikipedia editors, and further evaluation of the chatbot with Wikipedia editors.

Developing and evaluating a chatbot solution to enhance the contributor experience on Wikipedia demands a comprehensive analysis of chatbot user experience, human-chatbot interaction, and usability. This research covers the solution's initial design, development, and deployment. Based on the findings on the usability and improvements in user experience, further action is necessary to create a solution that is genuinely beneficial to the Wikipedia editors.

7 References

Abdellatif, A., Costa, D., Badran, K., Abdalkareem, R., & Shihab, E. (2020). A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering. https://arxiv.org/pdf/2012.02640.pdf

Adam, M., Wessel, M., & Benlian, A. (2020). AI-based chatbots in customer service and their effects on user compliance. Electronic Markets. (pp. 427-445). https://doi.org/10.1007/s12525-020-00414-7

Adiwardana, Luong, M.-T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R., Yang, Z., Kulshreshtha, A., Nemade, G., Lu, Y., & Le, Q. V. (2020). Towards a Human-like Open-Domain Chatbot. (pp. 1-38) https://arxiv.org/abs/2001.09977

Agarwal, & Wadhwa, M. (2020). Review of State-of-the-Art Design Techniques for Chatbots. SN Computer Science. (pp. 1-12). https://doi.org/10.1007/s42979-020-00255-3

Arora, C., Sabetzadeh, M., Nejati, S., & Briand, L. (2019). An Active Learning Approach for Improving the Accuracy of Automated Domain Model Extraction. ACM Transactions on Software Engineering and Methodology. (pp. 1-34). https://doi.org/10.1145/3293454

Ashktorab, Z., Jain, M., Liao, Q. V., & Weisz, J. D. (2019). Resilient Chatbots. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. (pp. 1-12). https://doi.org/10.1145/3290605.3300484

Avula, S., Chadwick, G., Arguello, J., & Capra, R. (2018). SearchBots. Proceedings of the 2018 Conference on Human Information Interaction. (pp. 52-61). https://doi.org/10.1145/3176349.3176380

Axios. (2021). Axios. https://axios-http.com/docs/intro

Baez M, Daniel F, Casati F. (2019). Conversational web interaction: proposal of a dialog-based natural language interaction paradigm for the Web. (pp. 94–110) http://dx.doi.org/10.1007/978-3-030-39540-7_7

Bangor, A., Kortum, P., & Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies. (pp. 114-123). https://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Bangor_May2009.pdf

Bird, C. M. (2005). How I Stopped Dreading and Learned to Love Transcription. (pp. 226-248). https://doi.org/10.1177/1077800404273413

Böhm, S., Eißer, J., Meurer, S., Linnyk, O., Kohl, J., Locke, H., Novakovskij, L., & Teetz, I. (2020). Intent Identification and Analysis for User-centered Chatbot Design: A Case Study on the Example of Recruiting Chatbots in Germany. (pp. 34–43).

Borsci, Malizia, A., Schmettow, M., Van Der Velde, F., Tariverdiyeva, G., Balaji, D., & Chamberlain, A. (2021). The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. Personal and Ubiquitous Computing. (pp. 1-16) https://doi.org/10.1007/s00779-021-01582-9

Brandtzaeg, & Følstad, A. (2017). Why People Use Chatbots. Internet Science. (pp. 377-392). https://doi.org/10.1007/978-3-319-70284-1_30

Brockman, G. (2021). OpenAI API. OpenAI. https://openai.com/blog/openai-api/

Bryant, S., L, Forte, A., Bruckman, A. (2005). Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. https://www.cc.gatech.edu/~asb/papers/bryant-forte-bruckman-group05.pdf

Budiu, R. (2018). The User Experience of Chatbots. Nielsen Norman Group. https://www.nngroup.com/articles/chatbots/

Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., O'Neill, S., Armour, C., & McTear, M. (2019). Assessing the Usability of a Chatbot for Mental Health Care. Internet Science. (pp. 121-132). https://doi.org/10.1007/978-3-030-17705-8_11

Canonico, M., & Russis, L.D. (2018). A Comparison and Critique of Natural Language Understanding Tools. (pp. 1-15).

Chandar, P., Khazaeni, Y., Davis, M., Muller, M., Crasso, M., Liao, Q. V., Shami, N. S., & Geyer, W. (2017). Leveraging Conversational Systems to Assists New Hires During Onboarding. (pp. 381–391). https://doi.org/10.1007/978-3-319-67684-5_23

Chaves, A. P., & Gerosa, M. A. (2020). How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. (pp. 1-30). International Journal of Human–Computer Interaction. https://doi.org/10.1080/10447318.2020.1841438

Cheng, & Jiang, H. (2020). How Do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use. Journal of Broadcasting & Electronic Media. (pp. 592–614). https://doi.org/10.1080/08838151.2020.1834296

Chow, & Xu, L. (2021). Chatbot for Healthcare and Oncology Applications using Artificial Intelligence and Machine Learning. https://doi.org/10.2196/27850

Chrome. (2021). Chrome DevTools. https://developer.chrome.com/docs/devtools/

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. Future Generation Computer Systems. (pp. 92, 539–548). https://doi.org/10.1016/j.future.2018.01.055

Clarke, V., Braun, V. & Hayfield, N. (2015). Thematic analysis. In J. A. Smith (Ed.), Qualitative psychology: A practical guide to research methods. (pp. 222-248). London: Sage.

Clarke, V., & Braun, V. (2016). Thematic analysis. The Journal of Positive Psychology. (pp. 297-298). https://doi.org/10.1080/17439760.2016.1262613

Concurrently. (2021). Concurrently. https://www.npmjs.com/package/concurrently

Crovari P., Pidó S., Garzotto F., Ceri S. (2021) Show, Don't Tell. Reflections on the Design of Multi-modal Conversational Interfaces. In: Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2020. (pp. 64-77). https://doi-org.zorac.aub.aau.dk/10.1007/978-3-030-68288-0_5

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. (pp. 475-487). https://doi.org/10.1006/imms.1993.1022

Dawson, C., & Dawson, R. (2013). Software Development Process Models: A Technique for Evaluation and Decision-Making. Knowledge and Process Management. (pp. 42-53). https://doi.org/10.1002/kpm.1419

DeDeo, S. (2016). Conflict and Computation on Wikipedia: A Finite-State Machine Analysis of Editor Interactions. (pp. 1-23) https://doi.org/10.3390/fi8030031

Dissanayake, N., Dias, K. (2017). Balanced Abstract Web-MVC Style: An Abstract MVC Implementation for Web-based Applications. GSTF Journal on Computing. (pp. 27-41). 10.5176/2251-3043_5.3.375.

Dolbir, Dastidar, T., & Roy, K. (2021). NLP is Not enough - Contextualization of User Input in Chatbots. https://arxiv.org/pdf/2105.06511.pdf

Duncker, D. (2020). Chatting with chatbots: Sign making in text-based human–computer interaction. Sign Systems Studies, 48(1). (pp. 79–100). https://doi.org/10.12697/SSS.2020.48.1.05

Express. (2021). Express. https://expressjs.com/

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. (pp. 132; 138-161). https://doi.org/10.1016/j.ijhcs.2019.07.009

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender Bias in Chatbot Design. Chatbot Research and Design. (pp. 79–93). https://doi.org/10.1007/978-3-030-39540-7_6

Fichter, D., J. Wisniewski. (2017). Chatbots Introduce Conversational User Interfaces. (pp. 56-58). Online Searcher.

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. Internet Science. (pp. 145–156). https://doi.org/10.1007/978-3-030-17705-8_13

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: findings from a questionnaire study. Quality and User Experience. https://doi.org/10.1007/s41233-020-00033-2

Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer Von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. (pp. 2915–2942). https://doi.org/10.1007/s00607-021-01016-7

Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. (pp. 511-527). https://doi.org/10.1177/0306312717692172

Frost B,. (2016). Atomic Design. (pp. 8-64) https://atomicdesign.bradfrost.com/table-of-contents/

Gabrielli, Rizzi, S., Bassi, G., Carbone, S., Maimone, R., Marchesoni, M., & Forti, S. (2021). Engagement and Effectiveness of a Healthy-Coping Intervention via Chatbot for University Students During the COVID-19 Pandemic: Mixed Methods Proof-of-Concept Study. (pp. 1-19) https://doi.org/10.2196/27965

Gibson, A., McCauley, C., Mulvenna, M., Ryan, A., Laird, L., Curran, K., Bunting, B., Ferry, F., & Bond, R. (2016). Assessing usability testing for people living with dementia. Proceedings of the 4th Workshop on ICTs for Improving Patients Rehabilitation Research Techniques - REHAB '16, 25–31. https://doi.org/10.1145/3051488.3051492

Gluza, W., & Turaj, I (2021a). Improving editor experience on Wikipedia [Unpublished manuscript]. Information Science, Aalborg University.

Gluza, W., & Turaj, I (2021b). Designing a chatbot to improve editor experience on Wikipedia [Unpublished manuscript]. Information Science, Aalborg University.

Gluza, W., & Turaj, I (2021c). Developing a chatbot to improve editor experience on Wikipedia [Unpublished manuscript]. Information Science, Aalborg University.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Computers in Human Behavior. (pp. 304–316). https://doi.org/10.1016/j.chb.2019.01.020

Google Cloud. (2021). Dialogflow. https://cloud.google.com/dialogflow

Goot M.J., Hafkamp L., Dankfort Z. (2021) Customer Service Chatbots: A Qualitative Interview Study into the Communication Journey of Customers. In: Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2020. Lecture Notes in Computer Science, vol 12604. Springer, Cham. (pp. 190-204) https://doi-org.zorac.aub.aau.dk/10.1007/978-3-030-68288-0_13

Halfaker, A., Kittur, A., Riedl, J. (2011). Don't bite the newbies: How reverts affect the quantity and quality of Wikipedia work. (pp. 163-172). New York, NY: Association for Computing Machinery.

Halfaker, A., Geiger, R., S., Morgan, J., T., Riedl, J., Forte, A., Lampe, C. Wellman, B. (2013). The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline (pp. 664-688). Beverly Hills, CA: Sage Publications. https://doi-org.zorac.aub.aau.dk/10.1177%2F0002764212469365

Hao, K. (2021). We read the paper that forced Timnit Gebru out of Google. Here's what it says. MIT Technology Review. https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

Hargittai, E., & Shaw, A. (2014). Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. Information, Communication & Society. (pp. 424–442). https://doi.org/10.1080/1369118x.2014.957711

Heroku. (2021). Deploying with Git | Heroku Dev Center. https://devcenter.heroku.com/articles/git

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior.(pp. 49, 245–250). https://doi-org.zorac.aub.aau.dk/10.1016/j.chb.2015.02.026

Hobert S., Meyer von Wolff, R. (2019). Say hello to your new automated tutor-a structured literature review on pedagogical conversational agents. (pp. 301-313). https://aisel.aisnet.org/wi2019/track04/papers/2/

Hobert, S., & Berens, F. (2020). Small Talk Conversations and the Long-Term Use of Chatbots in Educational Settings – Experiences from a Field Study. Chatbot Research and Design, 260–272. https://doi.org/10.1007/978-3-030-39540-7_18

Höhn, & Bongard-Blanchy, K. (2021). Heuristic Evaluation of COVID-19 Chatbots. (pp. 131–144). Springer International Publishing. https://doi.org/10.1007/978-3-030-68288-0_9

Holmes,S., Moorhead, A., Bond, B., Zheng, H., Coates, V., Mctear, M. 2019. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? (pp. 207-214). https://doi.org/10.1145/3335082.3335094

IBM. (2021a). IBM Cloud Docs. https://cloud.ibm.com/docs/assistant?topic=assistant-assistant-add

IBM. (2021b). IBM Cloud Docs. https://cloud.ibm.com/docs/assistant?topic=assistant-skill-add

IBM. (2021c). IBM Cloud Docs. https://cloud.ibm.com/docs/assistant?topic=assistant-skills-choose

IBM. (2021d). IBM Cloud Docs. https://cloud.ibm.com/docs/assistant?topic=assistant-actions-overview

IBM. (2021e). IBM Cloud Docs. https://cloud.ibm.com/docs/assistant?topic=assistant-skill-actions-add

IBM. (2021f). IBM Cloud Docs. https://cloud.ibm.com/docs/assistant?topic=assistant-deploy-web-chat

IBM. (2021g). IBM Watson. https://www.ibm.com/watson

IBM. (2022a). IBM API Docs. https://cloud.ibm.com/apidocs/assistant/assistant-v2

IBM. (2022b). Handling Watson Assistant response types. https://www.ibm.com/docs/en/voice-gateway?topic=states-handling-watson-assistant-response-types

IBM. (2022c). SDK handbook. https://cloud.ibm.com/docs/sdk-handbook?topic=sdk-handbook-node

IBM. (2022d). Settings. https://cloud.ibm.com/docs/assistant?topic=assistant-assistant-settings

International Organization for Standardization. (2019). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. (pp. 1-40) (ISO/DIS Standard No. 9241-210).

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. Proceedings of the 2018 Designing Interactive Systems Conference. (pp. 895–906). https://doi.org/10.1145/3196709.3196735

Janssen, A., Passlick, J., Rodríguez Cardona, D., & Breitner, M. H. (2020). Virtual Assistance in Any Context. A Taxonomy of Design Elements for Domain-Specific Chatbots. Business & Information Systems Engineering. (pp. 211–225). https://doi.org/10.1007/s12599-020-00644-1

Jiang, R., & Banchs, E. R. (2017). Towards improving the performance of chat oriented dialogue systems. (pp. 23–26). In 2017 international conference on asian language processing.

Jovanovic, M., Baez, M., & Casati, F. (2021). Chatbots as Conversational Healthcare Services. (pp. 44-51). https://doi.org/10.1109/mic.2020.3037151

Kelleher, T. (2009). Conversational Voice, Communicated Commitment, and Public Relations Outcomes in Interactive Online Communication. (pp. 172-188). https://doi.org/10.1111/j.1460-2466.2008.01410.x

Kobayashi, M., S. Arita, T. Itoko, S. Saito, and H. Takagi. (2015). Motivating multigenerational crowd workers in social-purpose work. (pp. 1813–1824).

Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.

Kocaballi, A. B., Laranjo, L., & Coiera, E. (2019). Understanding and Measuring User Experience in Conversational Interfaces. Interacting with Computers. (pp. 192–207). https://doi.org/10.1093/iwc/iwz015

Kuniavsky, M., Goodman, E., & Moed, A. (2012). Observing the user experience a practitioner's guide to user research (2nd ed.). Chapter 17: Research into Action: Representing Insights as Deliverables. (pp. 479-529). Morgan Kaufmann.

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2020). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. Chatbot Research and Design. (pp. 187-200). https://doi.org/10.1007/978-3-030-39540-7_13

Laban, G., & Araujo, T. (2020). The Effect of Personalization Techniques in Users' Perceptions of Conversational Recommender Systems. https://doi.org/10.1145/3383652.3423890

Laubheimer, P. (2018). Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests. Nielsen Norman Group. https://www.nngroup.com/articles/measuring-perceived-usability/

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. (pp. 95-105). https://doi.org/10.1016/j.ijhcs.2017.02.005

Lee, M., S. Ackermans, N. van As, H. Chang, E. Lucas, and W. I. (2019). Caring for Vincent: A Chatbot for Self-Compassion. (pp. 1-13) In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Liebl, L. (2019). 6 Steps for Creating a Smooth Chatbot Conversation Flow. Userlike Live Chat. https://www.userlike.com/en/blog/chatbot-conversation-flow

Liebrecht, C., & van Hooijdonk, C. (2020). Creating Humanlike Chatbots: What Chatbot Developers Could Learn from Webcare Employees in Adopting a Conversational Human Voice. Chatbot Research and Design. (pp. 51-64). https://doi.org/10.1007/978-3-030-39540-7_4

Lin, & Chang, D. (2020). Enhancing Post-secondary Writers' Writing Skills with a Chatbot: A Mixed-Method Classroom Study. (pp. 78-92). Educational Technology & Society.

Littlejohn, A., Hood, N., Rehm, M., McGill, L., Rienties, B., & Highton, M. (2019). Learning to become an online editor: the editathon as a learning environment. Interactive Learning Environments. (pp 1–14). https://doi.org/10.1080/10494820.2019.1625557

Liu, & Dong, Z. (2019). A Study of User Experience in Knowledge-Based QA Chatbot Design. Intelligent (pp. 589-593). https://doi.org/10.1007/978-3-030-11051-2_89

Lubbe, & Ngoma, N. (2021). Useful chatbot experience provides technological satisfaction: An emerging market perspective. (pp 1-8). https://doi.org/10.4102/sajim.v23i1.1299

Luger E., Sellen A. (2016). Like having a really bad PA - the gulf between user expectation and experience of conversational agents. (pp. 5286-5297). In: Proceedings of CHI 2016.

Mann, C. (2021). Can Conversing with a Computer Increase Turnout? Mobilization Using Chatbot Communication. Journal of Experimental Political Science, 8(1). (pp. 51-62). doi:10.1017/XPS.2020.5

Markdown Guide. (2021a). Basic syntax. https://www.markdownguide.org/basic-syntax/

Markdown Guide. (2021b). Extended syntax. https://www.markdownguide.org/extended-syntax/

Maslowski, I., Lagarde, D., & Clavel, C. (2017). In-the-wild chatbot corpus: From opinion analysis to interaction problem detection. http://biblio.telecom-paristech.fr/cgi-bin/download.cgi?id=17387

McDonnell, M., & Baxter, D. (2019). Chatbots and Gender Stereotyping. Interacting with Computers. (pp. 116-121). https://doi.org/10.1093/iwc/iwz007

McGrath, C. (2017). Chatbot Vocabulary. https://tangowork.com/chatbot-vocabulary/

McTear, M. (2020). Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots. (pp. 1) Morgan & Claypool Publishers.

Meira, M, 0., and A. M. P. Canuto. (2015). Evaluation of emotional agents' architectures: An approach based on quality metrics and the influence of emotions on users. In Proceedings of the World Congress on Engineering 1.

Microsoft. (2021). Write extensible code. https://docs.microsoft.com/en-us/dynamics365/fin-ops-core/dev-itpro/extensibility/writing-extensible-code

Modrzejewski, M., & Rokita, P. Ł. (2018). Graphical Interface Design for Chatbots for the Needs of Artificial Intelligence Support in Web and Mobile Applications. Computer Vision and Graphics. (pp. 48–56). https://doi.org/10.1007/978-3-030-00692-1_5

Moore, R. J., Arar, R., Ren, G. J., & Szymanski, M. H. (2017). Conversational UX Design. (pp. 492–497). https://doi.org/10.1145/3027063.3027077

Morabito, V. (2016). The Future of Digital Business Innovation: Trends and Practices. (pp. 3-21) https://doi.org/10.1007/978-3-319-26874-3

Morrissey, K., &. Kirakowski, J. (2013). Realness in chatbots: Establishing quantifiable criteria. (pp. 87-96). In: International Conference on Human-Computer Interaction.

Morville, P. (2004). User Experience Basics. https://www.usability.gov/what-and-why/user-experience.html

Mozilla Developer Center. (2021a). Document: keydown event. https://developer.mozilla.org/en-US/docs/Web/API/Document/keydown_event

Mozilla Developer Center. (2021b). Making decisions in your code - conditionals. https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Building_blocks/conditionals

Mozilla Developer Center. (2021c). Proxy servers and tunneling. https://developer.mozilla.org/en-US/docs/Web/HTTP/Proxy_servers_and_tunneling

Mozilla Developer Center. (2021d). Regular Expressions. https://developer.mozilla.org/en-US/docs/Web/JavaScript/Guide/Regular_Expressions

Mozilla Developer Center. (2021e). Session storage. https://developer.mozilla.org/en-US/docs/Web/API/Window/sessionStorage

Murillo, B., Sang, J., & Paz, F. (2018). Heuristic Evaluation and Usability Testing as Complementary Methods: A Case Study." Design, User Experience, and Usability: Theory and Practice. (pp. 470-478).

Ndukwe, Daniel, B. K., & Amadi, C. E. (2019). A Machine Learning Grading System Using Chatbots. (pp. 365-368). https://doi.org/10.1007/978-3-030-23207-8_67

Nielsen, J. (2012). Thinking Aloud: The #1 Usability Tool. Nielsen Norman Group. https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool

Node. (2021). Node. https://nodejs.org/en/about/

Nodemon. (2021). Nodemon. https://nodemon.io/

Norman, D. (2019). The Four Fundamental Principles of Human-Centered Design and Application. https://jnd.org/the-four-fundamental-principles-ofhuman-centered-design/

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis. International Journal of Qualitative Method. 160940691773384. https://doi.org/10.1177/1609406917733847

NPM. (2021). NPM. https://www.npmjs.com/

OtterAI. (2021). Otter Voice Meeting Notes. https://otter.ai

Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. (pp. 1549-1565). https://doi.org/10.1002/cae.22326

Perez-Soler, S., Juarez-Puerta, S., Guerra, E., & de Lara, J. (2021). Choosing a Chatbot Development Tool. (pp. 94-103). https://doi.org/10.1109/ms.2020.3030198

Plantak Vukovac, Horvat, A., & Čižmešija, A. (2021). Usability and User Experience of a Chat Application with Integrated Educational Chatbot Functionalities. In Learning and Collaboration Technologies: Games and Virtual Environments for Learning (pp. 216–229). Springer International Publishing. https://doi.org/10.1007/978-3-030-77943-6_14

Plotnikova, V., Dumas, M., & Milani, F. (2020). Adaptations of data mining methodologies: a systematic literature review. https://doi.org/10.7717/peerj-cs.267

Pop, D., Altar, A. (2014). Designing an MVC Model for Rapid Web Application Development. (pp. 1172-1179). https://doi.org/10.1016/j.proeng.2014.03.106

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. (pp. 1-12). https://doi.org/10.1145/3173574.3174214

Postman. (2021). Postman. https://www.postman.com/

Radziwill, N., and Benton, M. (2017). Evaluating Quality of Chatbots and Intelligent Conversational Agents. (pp. 25-36). Software quality professional

Ramos, R. 2017. Screw the Turing test - Chatbots don't need to act human. https://venturebeat.com/2017/02/03/ screw-the-turing-test-chatbots-dont-need-to-act-human/.

Rasa, (2021). Rasa. https://rasa.com/

React. (2021a). React. https://en.reactjs.org/

React. (2021b). React. https://en.reactjs.org/docs/create-a-new-react-app.html

React. (2021c). Hooks. https://reactjs.org/docs/hooks-intro.html

Reboot. (2017). New Editor Experiences. https://upload.wikimedia.org/wikipedia/commons/0/08/New_Editor_Experiences_ summary_of_findings%2C_August_2017.pdf

Redux. (2021a). Redux - API reference. https://redux.js.org/tutorials/fundamentals/part-1-overview

Redux. (2021b). Redux - fundamentals https://redux.js.org/tutorials/fundamentals/part-1-overview

Redux. (2021c). Redux - getting started. https://redux.js.org/introduction/getting-started

Redux DevTools. (2021). Redux DevTools. https://github.com/zalmoxisus/redux-devtools-extension

Redux-thunk. (2021). Redux-thunk. https://github.com/reduxjs/redux-thunk

Reem Alotaibi, Ahlam Ali, Haya Alharthi, & Renad Almehamdi. (2020). AI Chatbot for Tourist Recommendations: A Case Study in the City of Jeddah, Saudi Arabia. (pp. 18-30). https://doi.org/10.3991/ijim.v14i19.17201

Reenskaug, T. (1979). Model View Controller. https://folk.universitetetioslo.no/trygver/2007/MVC_Originals.pdf

Regexr. (2021). Regex for matching and capturing YouTube URLs with video code. https://regexr.com/3dj5t

Rietz, T., Benke, I., & Maedche, A. (2019). The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance. (pp.3-17). Wirtschaftsinformatik.

Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Stern, M. J. (2015). Digital inequalities and why they matter. Information, Communication & Society (pp. 569-582). https://doi.org/10.1080/1369118X.2015.1012532

Rohrer, C. (2014). When to Use Which User-Experience Research Methods. Nielsen Norman Group. https://www.nngroup.com/articles/which-ux-research-methods/

Roller, Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Shuster, K., Smith, E. M., Boureau, Y.-L., & Weston, J. (2020). Recipes for building an open-domain chatbot. (pp. 1-25) https://arxiv.org/abs/2004.13637

Ruane E., Farrell S., Ventresque A. (2021) User Perception of Text-Based Chatbot Personality. In: Følstad A. et al. (eds) Chatbot Research and Design. CONVERSATIONS 2020. (pp. 32-47). https://doi-org.zorac.aub.aau.dk/10.1007/978-3-030-68288-0_3

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., Mctear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? (pp. 207-214). https://doi.org/10.1145/3335082.3335094

Sanchez-Adame, Mendoza, S., Urquiza, J., Rodriguez, J., & Meneses-Viveros, A. (2021). Towards a Set of Heuristics for Evaluating Chatbots. (pp. 2037–2045). https://doi.org/10.1109/TLA.2021.9480145

SASS. (2021a). SASS - documentation. https://sass-lang.com/documentation

SASS. (2021b). SASS - installation. https://sass-lang.com/install

Schade, A. (2013). Remote Usability Tests: Moderated and Unmoderated. Nielsen Norman Group. https://www.nngroup.com/articles/remote-usability-tests/

Schuurmans, J., Frasincar, F., & Cambria, E. (2020). Intent Classification for Dialogue Utterances. (pp. 82-88). https://doi.org/10.1109/mis.2019.2954966

Sharp, H., Preece, J., & Rogers, Y. (2019). Interaction design : beyond human-computer interaction. Chapter 12: Design, prototyping, and construction. (pp. 421-470).

Shaw, A., & Hargittai, E. (2018). The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing. (pp. 143–168). https://doi.org/10.1093/joc/jqx003

Singh, Ramasubramanian, K., & Shivam, S. (2019). Chatbot Development Essentials. In Building an Enterprise Chatbot (pp. 35–53). https://doi.org/10.1007/978-1-4842-5034-1_3

Skjuve, M., Følstad, A., Fostervold, K., & Brandtzaeg, P. (2021). My Chatbot Companion - a Study of Human-Chatbot Relationships. International Journal of Human-Computer Studies. https://doi.org/10.1016/j.ijhcs.2021.102601

Skorupska, K., Warpechowski, K., Nielek, R., & Kopeć, W. (2020). Conversational Crowdsourcing for Older Adults: a Wikipedia Chatbot Concept. https://doi.org/10.18420/ecscw2020_ep05

Smestad, & Volden, F. (2019). Chatbot Personalities Matters: Improving the User Experience of Chatbot Interfaces. (pp. 170–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-17705-8_15

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. https://doi.org/10.18653/v1/p19-1355

Syvänen, S., & Valentini, C. (2020). Conversational agents in online organization–stakeholder interactions: a state-of-the-art analysis and implications for further research. (pp. 339–362). https://doi.org/10.1108/jcom-11-2019-0145

Ta, Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. https://doi.org/10.2196/16235

Terblanche, N. (2020) A design framework to create Artificial Intelligence Coaches, International Journal of Evidence Based Coaching and Mentoring. (pp. 152-165). DOI: 10.24384/b7gs-3h05

Thapa, S. (2021). Investigating the implication for business innovation of customer encounters with AI=based chatbots. Aalborg Universitet. (pp.1-63) https://kbdk-aub.primo.exlibrisgroup.com/permalink/45KBDK_AUB/n4l1aj/alma9921003598705762

Thomas, C., & Bevan, N. (1995). Usability context analysis: a practical guide. (pp. 8-95). https://repository.lboro.ac.uk/articles/report/Usability_context_analysis_a_practical_guide/9353600

To, Green, C., & Vandelanotte, C. (2021). Feasibility, Usability and Effectiveness of a Machine Learning Based Physical Activity Chatbot: a Quasi-Experimental Study. https://doi.org/10.2196/28577

Torres, F, W., & Martins, L. (2018). Accessibility in Chatbots: The State of the Art in Favor of Users with Visual Impairment. In Advances in Usability, User Experience and Assistive Technology (pp. 623–635). Springer International Publishing. https://doi.org/10.1007/978-3-319-94947-5_63

Tosi, F. (2020). From User-Centred Design to Human-Centred Design and the User Experience. (pp. 47–61). In Design for Ergonomics

Tran, O. T., & Luong, T. C. (2020). Understanding what the users say in chatbots: A case study for the Vietnamese language. (pp. 1-10) https://doi.org/10.1016/j.engappai.2019.103322

Userbrain. (2021a). Userbrain - user testing made simple. https://www.userbrain.com/en/

Userbrain. (2021b). Userbrain - terms and conditions. https://tester.userbrain.com/terms

Van den Broeck, Z, B., & Poels, K. (2019). Chatbot advertising effectiveness: When does the message get through? (pp. 150–157). Computers in Human Behavior.

Vanichvasin. (2021). Chatbot Development as a Digital Learning Tool to Increase Students' Research Knowledge. (pp. 44-56). International Education Studies.

Vetter, M. (2002). Quality aspects of bots. In Software Quality and Software Testing in Internet Times. (pp. 165-184). Berlin, Germany: Springer Berlin Heidelberg.

Vijayakumar, B., Höhn, S., & Schommer, C. (2019). Quizbot: Exploring Formative Feedback with Conversational Interfaces. Technology Enhanced Assessment. (pp. 102-120). https://doi.org/10.1007/978-3-030-25264-9_8

W3 school. (2021a). Button. https://www.w3schools.com/tags/tag_button.asp

W3 school. (2021b). JavaScript Array map(). https://www.w3schools.com/jsref/jsref_map.asp

W3 school. (2021c). OnClick event. https://www.w3schools.com/jsref/event_onclick.asp

Wallis, P., & Norling, E. (2005). The trouble with chatbots: Social skills in a social world. (pp. 29–38). In Proceedings of the joint symposium on virtual social agents

WCAG. (2008). Web Content Accessibility Guidelines (WCAG) 2.0. https://www.w3.org/TR/WCAG20/

Wezel, M. M. C., Croes, E. A. J., & Antheunis, M. L. (2021). "I'm Here for You": Can Social Chatbots Truly Support Their Users? (pp. 96-113). https://doi.org/10.1007/978-3-030-68288-0_7

Wikimedia Foundation. (2019). Research Report Nº 1. https://research.wikimedia.org/report.html

Wikimedia Foundation. (2021). Wikipedia celebrates 20 years of free, trusted information for the world. https://wikimediafoundation.org/news/2021/01/14/wikipedia-celebrates-20-years/

Wikimedia Product. (2018). Contribution taxonomy: Master inventory [Dataset]. https://docs.google.com/spreadsheets/d/1jTJaYnLXrsqfplKaMcKSWR-9qpKP1xOu1gqRq4C8rhw/edit#gid=0

Wikimedia Product. (2019). Wikimedia Product/Contribution taxonomy - MediaWiki. https://www.mediawiki.org/wiki/Wikimedia_Product/Contribution_taxonomy

Wikipedia contributors. (2021a). Wikipedia:FAQ/Contributing. Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:FAQ/Contributing

Wikipedia contributors. (2021b). Wikipedia:FAQ/Editing. Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:FAQ/Editing

Wikipedia contributors. (2021c). Growth/Focus on help desk - MediaWiki. https://www.mediawiki.org/wiki/Growth/Focus_on_help_desk#Experiment_results_(September_2019)

Wikipedia Contributors. (2021d). Growth/Focus on help desk - MediaWiki. MediaWiki. https://www.mediawiki.org/wiki/Growth/Focus_on_help_desk#Experiment_results_(September_2019)

Wikipedia contributors. (2021e). Help:Introduction/All - Wikipedia. Wikipedia. https://en.wikipedia.org/wiki/Help:Introduction/All

Wikipedia contributors. (2021f). History of Wikipedia. Wikipedia. https://en.wikipedia.org/wiki/History_of_Wikipedia

Wikipedia contributors. (2021g). Wikipedia:Copyrights - Wikipedia. Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:Copyrights

Wikipedia contributors. (2021h). Wikipedia:Help desk - Wikipedia. Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:Help_desk

Wilson, H. J., P. R. Daugherty, and N. Morini-Bianzino. (2017). Will AI create as many jobs as it eliminates? MIT Sloan Management Review. Available at: http://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates/.

Wochinger T., (2019a). Rasa NLU in Depth: Part 1 – Intent Classification. https://blog.rasa.com/rasa-nlu-in-depth-part-1-intent-classification/

Wochinger T., (2019b). Rasa NLU in Depth: Part 2 – Entity Recognition https://blog.rasa.com/rasa-nlu-in-depth-part-2-entity-recognition/

Wolf, D, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Rush, A. M. (2019). HuggingFace's Transformers: State-of-the-art Natural Language Processing. (pp. 1-8). https://arxiv.org/abs/1910.03771

Zamora, J. (2017). I'm Sorry, Dave, I'm Afraid I Can't Do That. (pp. 253-260). https://doi.org/10.1145/3125739.3125766

Zarouali, B., van den Broeck, E., Walrave, M., & Poels, K. (2018). Predicting Consumer Responses to a Chatbot on Facebook. Cyberpsychology, Behavior, and Social Networking. (pp. 491-497). https://doi.org/10.1089/cyber.2017.0518

Zheng, L. N., Albano, C. M., Vora, N. M., Mai, F., & Nickerson, J. V. (2019). The Roles Bots Play in Wikipedia. Proceedings of the ACM on Human-Computer Interaction. (pp. 1-20). https://doi.org/10.1145/3359317

Zierau, N., Engel, C., Söllner, M., & Leimeister, J. M. (2020). Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda. (pp. 99-114). https://doi.org/10.30844/wi_2020_a7-zierau

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist - it's time to make it fair. (pp. 324-326). https://doi.org/10.1038/d41586-018-05707-8

8 Appendix

Transcripts of the video recordings from the think-aloud test. The videos are available in OneDrive or Youtube Playlist.

Tester 1 - Rich

0:06
This is a test of the chatbot prototype that could help Wikipedia editors with frequently asked questions and improve their experience. The study is voluntary and anonymous, you will be asked to answer some general questions complete four quick tasks, and assess 16 short statements about the Chatbot. Please remember to think out loud during your test and tell us if there's any information missing. If the Chatbot does not respond, please refresh the browser.

0:42
Alright, step two, you open an account with Wikipedia years ago, and you want to change your silly password ask the chatbot for help. Alright, so I'm gonna click on let's chat. And see we see: Hello, I'm excited that you're a part of Wikipedia. How can I help you? So let's go with - technical. Okay, usually people ask me about these technical issues. Alright, we're gonna choose how to change my password. And you change your password, the password change. And you can also find a link to this in your preferences. Alright, so I think that's where as far as I wanted me to go. Okay, yep, I asked for the password, change and it directed me to where to go to do so.

1:53
Please rate the task and explain why you gave such a score. I'm gonna give it a seven. I didn't have to kind of, you know, dig too deep in order to find what I was looking for. I also, I know sometimes with these bots, you get into a loop or goes over and over and over again, it doesn't really know what you're looking for. This had no issues with this. How to change my password was one of the options. Technical was kind of the most obvious starting point. And then it clearly told me where to do so. I'm giving it a seven

2:31
You're a blogger and you noticed that someone used your content on Wikipedia without your permission. Use the chat bot to find out what to do about this. Alright, let's see. I can go back up top here. I'm gonna close out. We'll start over. Let's see... your policies and guidelines on this one. And then from here, I think I'm going to choose... what should I do when I see plagiarism? And see if you see an article including inadequately attributed text or images or words copyright violations, please tell us about it. And the topic is of the article then reported at copyright problems. I'm just going to say that was helpful and I'm gonna say I've completed that task

3:51

Please rate the task, again it was very easy. I think the the option felt run to bring two main categories pretty easily to choose from and once I selected one of the main categories it was clearly one of the options so I'm gonna say seven.

4:09

You could see that one of the articles is missing a footnote. Right now how to add it. Let's see... footnote. I'm going to do referencing Let's see. Let's see how do I cite Wikipedia. How do I cite sources. You know what? I don't think maybe that's going to come in - let's go to editing. Yeah. And then how do I edit the page maybe? Here are some common questions regarding references - how do I edit the page. It's great that you want to start your editing journey to edit the whole page at once, click the edit this page tab at the top. That just one section, click Edit, link to the right of section and heading you can learn more about editing here. So I think that's what we want to do we kind of want to edit since a footnote is missing. What I do want to do is to see how I see citing just curious with citing does as well. Yeah, I think I think I want to edit the option since there's a missing footnote. So what I think I would want to do is get into reading or I probably would start watching this editing article. This is a little bit on on that particular. But I think that's where I would look to do since it's missing a footnote, I would think we would need to edit the page. So I'm going to say I've completed it.

6:18

And as far as the ease this one I'm a little bit unsure, I'm not 100% Confident. So I'm going to give this one a... I think I think we went down the right road, but not 100% Sure. So I'm gonna give you the five.

6:33

You're writing an article about a football player, if think that there is a category missing Ask the Chatbot if you if it can help you. Alright, why don't we type? I'm gonna just type missing category. See what happens about that? Here we go. And it does give up a bunch of options regarding category, so for this instance, I think what we want to do is how do we add a category and we click on that. And then it gives you a tutorial on creating new categories. All right, so that was easier. And that's probably what I should have done with the previous one, instead of just looking for the predefined options probably should have liked in a couple of keywords.

7:43

So I completed the task and I'm gonna say that was very easy since I typed in, you know, a couple key words, it took me exactly where I was looking for.

7:55

In the following section, you will be asked to evaluate your experience with a chatbot by reading 16 short statements, please complete this questionnaire by reading each statement carefully. And scoring with the number of the best matches how you feel about the statement.

8:11

The chatbots personality was realistic and engaging. Yeah, I'm gonna say was pretty realistic, I didn't get a feel that I was actually talking to a bot, I could have very easily been talking to person on the other end. So I'm gonna give it five

8:29

Chatbot seemed to robotic, I'm gonna give it a two on this one. The chatbot was walking me during the initial setup. If we go back to the initial, you know that - Hello, the Chatbot I'm excited that you're part of Wikipedia. How can I help you? So yeah, it was very engaging. Didn't just feel like it was what do you want to get right to the point, you know, had that nice introduction. So I'm gonna give it a five.

8:55

The chatbot seemed very unfriendly. I'm gonna say strongly disagree I didn't get that vibe at all. The chatbot explained its scope and purpose. Well, yeah, right from the beginning, it said, you know, again, welcome. It was it clearly mentioned Wikipedia. And it wanted to know how I could help you and even gave you some sub menus, you know, edit, editing, policies, referencing images and technical. So I did. I felt fairly comfortable. I was at the right place to find out information.

9:30

The chatbot gave no indication as to its purpose. Yeah, so it didn't come out in literally say, you know, I understand you have questions regarding Wikipedia. I have the answers. What would you like to know? But it clearly said, you're part of Wikipedia. How can I help you and it gave something you see, you clearly knew. At least I did. You were at a place where you can get questions answered to something you might have for Wikipedia. So I would give it a four out of five.

10:08

The chatbot was easy to navigate Yeah, I had no issues navigating. Sub menus were easy to submit the sub options were easy to find. You didn't get lost you didn't get in an infinite loop. And then you even had the option of actually ask a question so give me a five on that.

10:29

It would be easy to get confused when using chatbot Yeah, I didn't find it confusing or I don't think it would fit. The only thing I'll mention is I don't believe it did this I would have liked to have... after you do the thumbs up thumbs down... Maybe have the sub these editing policies and guidelines referencing images and technical come back up. Or maybe have you know another instead of saying hello I'm the Chatbot you know maybe say I don't think it did say something clever to help you. Is there anything else below or some news in case you need to reference them again? Something like that yeah, I said that's awesome. I'm glad I could help let me know if you have more questions so didn't give that maybe you just mentioned again I'm glad I

could help you let me know if you have any more questions but then say you know blowers some in use as a reference point or something of that nature. It's nitpicking but just something I would have liked to have. But as far as confusing I'm gonna say I'm gonna say warning was not easily obvious not easily confused.

11:56
The chatbot understood me well. Absolutely, every one I already option I select the when I entered a keyword it's exactly what I was looking for.

12:07
Chatbot failed to recognise a lot of my inputs. I strongly disagree with that one I had no issues there.

12:14
Chatbot responses were useful, appropriate and informative. Again, I'm giving a five on that one. I didn't enter anything or select something and it took me a whole different you know, avenue. Everything I selected took me to exactly what I was looking for.

12:32
Chatbot responses were irrelevant. Again, I'm giving a one on that.

12:36
A chatbot coped well with any errors or mistakes I really didn't have any errors or mistakes. So I'm just going to give a three on that.

12:51
The chatbot seemed unable to handle any errors. From what I've seen I didn't have any errors. So I'm going to give a three on this. I didn't really have the opportunity to come across any errors.

13:04
The chatbot was very easy to use. I'm gonna give it a five on it. Chatbot was very complex. We're going one on that it was very easy, very user friendly interface. That's it.

Tester 2 - Ian

0:01
One moment. All right.

0:04
This is a test of the chatbot prototype that can help Wikipedia editors with frequently asked questions and improve their experience. This study is voluntary and anonymous. You will be asked to answer some general questions complete four quick tasks and assess 16 short statements about the chatbot. Please remember to think out loud during your test and tell us if there's any information missing. If the chatbot does not respond, please refresh the browser. Thank you for your participation.

0:30
You open an account with Wikipedia years ago, and you want to change your silly password. Ask the chatbot for help. So let's see - let's chat. Looks like it has prearranged things. But maybe let's see if I could do it directly - change password, see if it accepts that. Change your password via password change. So I would believe that that's the way that I would find it. And it works. So I'd say that I completed the task.

0:58
Please rate the task and explain why you gave such a score. It was very easy. The first thing that, without even having to kind of look through what's being offered already as quicklinks. The fact that it was able to recognise that I wanted to change my password and provided me a direct link to it rather than just stating it, makes it, it did exactly what I wanted it to do.

1:22
You are a blogger and you notice that someone used your content on Wikipedia without permission, use the chatbot to find out what to do about it. So... this I might look. I might type in like just, you know, content permission, I might use that. But I might also first want to look at policies and guidelines and just see what they have to say about that. Factual error, plagiarism. So, I could have found, possibly find it the other way. And then I see here, there's a talk page and the ability to report it. Let's see, including attributions copyright violations, etc. So I found it, although let me just type in plagiarism. See, see if it finds the same thing that way as well. Yeah. Okay. So it does, there's multiple ways of doing that. Once again, very easy, I was able to think of how I might type in instead of using natural language, just trying to use important keywords. And that worked. As well as using the reference bubbles up at the top buttons up at the top.

2:41
You can see one of the articles is missing a footnote, find out how to add it. So let's see, let's try this time, let's go with something simple as again, add footnote. If it understands that, if you want to know how to add a citation, choose the most helpful

option. So... I'm not sure let's try article here. I don't know referencing for beginners. Let's try that and see what it does. So it takes me to the How To Guide. And then let's see editing. Let's see if it does something different there as well. How do I edit a page? Let's go to how do I edit a page. And then here's more information about editing. So it seems that there's multiple ways of solving this problem, depending on the kind of level of detail that I'm looking for. Let's try missing footnote and see if it does anything different when I say when I put it, put it together like that. Want to know how to add an option. Okay. And then let's look at a code snippet too, and then it shows me like information for how to add things. So it looks like that there are multiple ways of being able to solve this problem. I would give this a six out of a seven, I think it's still very easy. It's just that it seems that there's it's kind of like a footnote is built into a much larger structure of learning how to edit pages, learning how to add things to pages, delete things, so on and so forth. So it's not really that it's, I mean, there's probably a more direct way that you could like if it was specific about adding a footnote if it provided that. Whereas it seems to be more in general editing.

4:25

You're writing an article about a football player and you think that there is a category missing for Uruguayan football players. Ask the chat bot, if it can help you. Let's see. Category missing category. That's probably where I would try it. So let's see. Missing category. This again might be something that I might look up at the top let's see how do I add a category. And then also what are categories... so it provides me this information in multiple ways here too is so that I can build up. I do like this that it offered kind of new multi path directions in how I would how I would find the solution here. Categories, renamed redirected. What are categories? How do I add an article to a specific category, whereas this is, gives me a video tutorial about how to add a missing category may try that to missing categories as well. Okay, so it takes me the same place. So I think if and I would still give them, I would still give that a very easy despite that it didn't. I mean, it gave me this kind of situation where it gives me multiple options about understanding categories.

5:46

In the following section, you will be asked to evaluate your experience with a chatbot by reading 16 short statements, please complete this questionnaire by reading each statement carefully and then scoring with a number that best matches how you feel about the statement.

5:59

The chatbot personality was realistic, engaging. It's pretty neutral, just as I you know, well, maybe adding a little bit of flavour with emojis, things like that. But I would say I don't know what you mean by realistic? Is it realistic in a sense is that act like a chatbot? Or realistic in the sense that it acts like a human? So I give that a four? I would say it's engaging enough to be... I think of the scenarios that I encountered. Very quickly, could I get to the next stage of either knowing what I need to do like with being able to report plagiarism, or finding the location of the tutorial or

information to solve the problem that I'm currently having. And if it's successful doing that, which which it was, I would say that it's engaging in doing its job. So I would give it like a 4.5 here. I think if you were to humanise it. Well, I don't know, it'd be be it 4.5.

7:11
Seems too robotic? No, I think it operated fine. It wasn't confusing. It wasn't stilted. It wasn't stiff.

7:18
The chatbot was welcoming during the initial setup, you could say that, it's a generic greeting. And the same with like, you know, hello on wiki. I mean, it's fine.

7:31
Chatbot seemed very friendly, completely disagree.

7:34
It explained its scope and purpose. Well, well, I mean, yeah, I think right here, the very first thing that you see is enough. And then the second stage is kind of getting into some of the details and providing you some kind of like, initial areas that you could look into. I think it explains it perfectly.

7:52
It gave no indication as to its purpose. No right away, it did.

7:58
Easy to navigate? Extremely.

7:59
It would be easy to get confused when using this chap on? I don't think so. And I think that it probably would even I don't know, let's let's just type in some random thing and see what it does there. Yeah, even even when it doesn’t understand it's very direct about that.

8:16
The chatbot understood me Well, I would say that it, to one degree or another, even including the non-understanding, it was able to understand me in the sense that it stated what it thought or I don't know, what you want to call it thought would be the correct answer. And it provided as such. And then also even when it didn't know what I was stating, it asked me to rephrase it.

8:44
Failed to recognise a lot of my inputs? No, each each test that I made, as I went through it without even having to use natural language. The article is missing a category. Let's see if it tags it well that way. How do I add an article? Yeah, so how to

add an article to a category? Wow. Okay. So that's, well, no, no, I guess that's probably my fault, because the it should be the categories missing an article. So it's actually correct here. I just kind of mentally did that backwards. Someone stole my blog post. See if it finds that too. Why is someone changing, deleted, feedback, writing negative, what prevents someone from contributing false. So... this doesn't, it is missing a little bit of things, but I think that's more advanced. I just wanted to get to that point. That's I mean, it did recognise a lot of my inputs.

9:50
Chatbot responses were useful, appropriate, informative? Definitely agree.

9:54
Chatbot responses were irrelevant. No, and if with these two tests, this was my fault. And it correctly looked at it that way. And then with someone stole my blog post, it didn't understand. It's looking more about it from an editing point rather than a plagiarism point. So it's not perfect. I wouldn't call this irrelevant... somewhat irrelevant.

10:19
Chatbot coped well with any errors or mistakes. Yeah, I think maybe you could add a it's something else category when you have stuff like this. So that it, you know, gives you the opportunity to kind of maybe even like, rephrase, rather than not understanding and rephrase.

10:36
Unable to handle any errors? I disagree.

10:40
Very easy to use strongly agree.

10:43
Was very complex? Disagree.

10:45
Thanks. That's all thank you for participating. So yeah, I mean, literally. It did everything that I essentially expected it to be able to do, though, I do suspect that when using natural language, rather than just you know, key terms. Let's see. Let's try stolen. Stolen. See if it understands that. No, what is an orphan so it's, it's not getting the stolen part here for like plagiarism. But that's, I mean, there probably are things like that, that it just have to expect that certain keywords might not provide the information that the person is looking for. And then that would be the kind of only suggestion that I have here is that it's possible that to add maybe a button that says you know, no, I mean something else or I'm looking for something different.

Tester 3 - Jessie

0:04
So this is a test of the chatbot prototype that could help Wikipedia editors with frequently asked questions and improve their experience. The study is voluntary and anonymous, you'll be asked to answer some short general questions, complete four quick tasks and assess 16 short statements about the chatbot. Please remember to think out loud during your tests and tell us if there's any information missing. If the chat bot does not respond, please refresh the browser.

0:39
You opened an account on Wikipedia years ago and want to change your silly password ask the chat bot for help.

0:58
Okay, so now it was very easy. I kind of guessed that it would be technical and it quickly gave me the option that I needed to change my password.

1:14
You're a blogger and notice that someone used your content jn Wikipedia without permission. Use a chatbot to find out what to do about it.

1:33
How to get rid of that? Oh, here we go. You want to say someone use your content on Wikipedia without permission?

2:16
Think it's that? Yeah, I think it's that copyright problems. Ah, that was fairly easy. I just wasn't quite as sure.

2:27
You can see that one of the articles is missing a footnote, find out how to add it.

2:44
So this is telling me I'm assuming that would help you with footnotes? Or maybe it's referencing? No, I think it would be that bit. How to edit a page I think I think that's right. Just not so sure about whether the editing will be about the footnote.

3:06
You're writing an article for football player about a football player and you think there is a category missing for Uruguayan football players ask the chatbot if it can help you.

3:20

With the editing How do I create a new page? Yeah, I think it would be in their head to create an article.

3:41
Again, I'm not entirely sure if that's the right bit that I'm looking for. But that would be where I'd be tempted to go.

3:47
The following section you will be asked to evaluate your experience with a chatbot by rating 16 short statements. Please complete this questionnaire by reading each statement carefully and scoring with a number that best matches how you feel about the statement.

4:02
Yeah, it was fine it was easy it's very clear and easy.

4:08
No it's fine. I expect that.

4:12
Yeah. don't actually know about arrows couldn't send em don't know about that either. That's it.

Tester 4 - Jacob

0:03
Hello there, hello.

0:06
Okay, so... this is a test of chatbot prototype that could help Wikipedia editors with frequently asked questions to improve the experience. The study is voluntary and anonymous, you'll be asked to ask some general questions complete four quick tasks, and assess 16 short statements about the chat pot. Okay, so please remember to think out loud, join us has to tell us if there's any information missing. If the chatbot does not respond, please refresh the browser. Thank you for your participation. The reason why I'm speaking fast is because there's 27 steps. So yeah, so I don't know how, you know, I don't know if it'd be steps can be a minute, if it is take 27 minutes. So that's why I'm going fast with reading out the instructions.

0:45
Okay, so you opened an account on Wikipedia years ago and wants to change your silly password. Ask the chatbot for help.

0:52
Okay, so we're going to ask the chatbot for help. So let's chat. So help, I guess. So I'm a wiki Chatbot. I can help you anything Wikipedia, just type in your question. So what do I type in change passwords, change password. You can change your password by the password change, you can find a link and this is in your preferences. So we have to log in and then change the password. So it looks good to me so far. That looks good. So I could complete the task. Okay. I think it was okay. They gave us the link to login. Well, you have to log in to do the link, but you know, I think it was fine overall. And yeah, I think it was fine. So I'm gonna give it a six. So far, it looks very easy to do.

1:47
Okay, so you're a blogger, and you noticed someone used your content on Wikipedia without permission. Okay, so now we're going to do... to help someone used my content without permission, okay. What, okay, so they said, what, what should I do if I find a copyright violation on Wikipedia? Well, it's my article deleted, what prevents someone? So we're gonna go this one is the first option. Okay, so we take this very seriously. So we can go to licencing or copyright problems. So it's good that they've linked it as well. So I think that's good to talk to us about licencing here, and it talks to us about copyright problems here. I don't know if this one helps, though. But licencing I think helps maybe, I'm not sure. I'm not sure if they helped me out here. So I've completed this task. I'm gonna give it a four. I'm not sure if they helped me out here. They didn't talk to me about any reporting. You know, report needs people.

2:49

Okay, so you can see once that one of the articles is missing a footnote, how do I... Okay, how do I add a footnote to an article? So answer this question. Okay, so if you're new to Wikipedia, this article is helpful. So it links us to referencing for beginners as we can see here. So it talks to us how we could do that. Okay, that's pretty good. Okay, that's pretty good to me. Okay, I have completed this task. Task was very easy to complete so far. Okay.

3:22

So, again, so you are writing an article about football player and there is a category for it, you're going missing for your game football players ask the tap what you can if you can help you help. There is a category missing on missing on a page for Uruguayan footballers? How do I think it's how do I add a category? article to a category? Okay, so it tells us here so it gives us even the sort of jargon, how to do it. So edit the article and add a category at the bottom of the article, for example. So that's interesting. So it tells us how to add it. So that that helps out so I have completed the task. Okay, and I think that's very easy to complete.

4:12

Okay, so in the following section, you'll be asked to evaluate your experience with a chatbot by raising 16 short statements. Please complete this questionnaire by reading each statement carefully and scoring with the number that best matches how you feel with the statement.

4:25

Okay. Okay, so the chatbots personality was realistic and engaging. Yes, it was it was quite realistic and engaging.

4:36

Okay, so the chatbot seems to robotic? At times, yeah, it did. It did seem quite robotic. So give it a free there.

4:44

Okay, so the chatbot was welcoming during the initial setup. Yeah, it was so go to the source app right now. Yeah, well, this is Hello. So that was nice. So I'll give that strongly agree.

4:55

Okay, so the chat box in very unfriendly? No, not really. I disagree with that. There. It was quite friendly.

5:01

The chatbot explained its scope and purpose. Well, yeah, it did explain the scope of purpose quite well.

5:08
The chatbot gave no indication as to its purpose? Disagree with that the chatbot did give indication as to why it exists. And I'm pretty happy about that. So you know, if that makes sense, like, you know, yeah.

5:23
Okay, so the chatbot was easy to navigate. Yes, it was, it was very easy to navigate over all it gave us, you know, different areas, etc. So it's really useful.

5:33
It would be easy to get confused when using a chat bot. No, disagree, that is quite easy to get to find useful information from the chatbot itself and the information it provided.

5:43
So the Chatbot understood very well. Yes, yes, it did. He understood very well, I believe. Maybe, maybe, maybe not. 100%. But I think it could be improved in some aspects. But yeah, I think overall, it's pretty good.

5:55
Okay, so the Chatbot failed to recognise any of my inputs.

5:59
The chatbot didn't fail. Now, I disagree with that. I think mostly it didn't fail. I mean, maybe the copyright violation, I think was fine, but they didn't tell us how to oppose it, which is annoying. But other than that, it didn't fail.

6:16
Anyway. chatbot responses were useful, appropriate and informative. Yeah, I strongly agree to that. Yep.

6:22
Okay. Chatbot responses were irrelevant. Disagree. Strongly Disagree. That's pretty good. Overall.

6:28
The chatbot coped well with any errors and mistakes? Coped very well with mistakes. Yes. It did. I think I don't think I made a mistake. But I think it did go. Well, usually, you know, when I did say everything.

6:44
The chatbot seems unable to handle any errors. Disagree. Okay. disagree that there was no I was no errors in the chatbot.

6:57

Okay, so the chatbot was very easy to use. I strongly agree with that.

7:00
The chabot was very complex. Yeah, it was quite complex. Yes, it was. As you know, it wasn't complex. It was quite nice. It was quite not complex. It's crazy. Just ask it a question, gives you the answer. So, you know, an all bunch of answers, which is really good.

7:16
Okay, so thank you for participating. So have a nice day. Thank you very much.

7:19
I thought it should take longer than four. But yeah, very easy to go through. The chatbot is very useful.

7:26
So I'm going to ask how do I help you? How do I report copyright? So I'm gonna ask that. Did you mean, what should I do if I find a copyright violation? When they go here? We can also go to this again. So hello, report something that's not working correctly. Okay. It says goes back down there. How do I cite sources? Okay. Article. So they can even show civility to a code snippet. That's pretty useful. It's code snippets. So the code snippet really helps.

7:58
No, wait, no, hold on. That's not good. They were supposed to show us the citation in the code snippet, but it couldn't do that. Hold on so that's interesting. Okay, there we go to now it works so it's a bit of an error here, just an FYI.

8:17
Anyway, thank you very much, and have a nice day.

Tester 5 - Tokoni

0:03
This is a test of a chatbot prototype that will help Wikipedia editors with frequently asked questions and improve their experience. The study is voluntary and anonymous, you will be asked to answer some general questions, complete four quick tasks and assess sixteen short statements about the chatbot. Please remember to think out loud during your test and tell us if there's any information missing. If the chatbot does not respond, please refresh the browser.

0:28
Okay. You opened an account on Wikipedia and you want to change your silly password ask the chatbot for help.

0:40
Okay, so I've opened it I think it's going to be technical stuff. How do I change my password? Alright, so it's it's was able to give me the info so I have completed the task.

0:54
Please read the task and explain why you do such a score. Very easy because as soon as you sort of come in you immediately think you know - how can I help you and you have a go this is probably technical, and how do I change my password is the second thing so it's extremely easy.

1:09
You are a blogger and notice that someone uses your content on Wikipedia without permission. Use the chatbot to find out what to do. So here is my content without my permission.

1:21
I think it's a referencing issue. N,o it's not. I'm editing I guess. Cited my work without permission. Copyright violation - I think that should be it. Okay, would be the copyright holder. Okay, yeah, so I've been able to. I'm gonna give it a five because this was a situation where you couldn't just like tap on a few buttons on your question eventually came up, you had to manually like write it in. But overall it was pretty straightforward. I got the answer I needed as soon as I typed it in.

2:16
You can see that one of the articles is missing a footnote, find out how to add it.

2:22
Let me just try to add a footnote. Article, yes. I imagine so says referencing for... Yeah, I think I've completed the task. It was talking about referencing for beginners so I think that should be like that should be added under.

2:45
This has I'm gonna give it a six because I only need it was only two steps. And as soon as I said footnote I got referred to something that helps with all referencing so I think that was pretty easy.

2:57
You're writing an article about a football play and think there has been a category missing for Uruguayan football players. Ask the chatbot if we can help you.

3:07
So I think I'm just like very much looking for the keywords. So I'm going to say missing category. Right, no, how do I add a category? Yeah. Creating new categories. Okay, so yeah, I found it. I've been able to complete the task. Very easy. I'm sorry. And the reason is easy is that just by adding the keyword, two keywords, I was able to find categories. Pun unintended, that sort of showed me what I was looking for. So it was pretty easy.

3:51
In the following section you will be asked to evaluate your experience with chatbot by reading 16 short statements please complete the questionnaire by reading each statement carefully and scoring with the number that best matches how you feel about the statement.

4:02
Okay, the chatbot personality is realistic and engaging. I'm going to say realistic. Yes, it's not like, it's actually it is quite engaging in that, it responds very well too. When you put in keywords and ask questions. I'm actually going to give it a five. It was quite realistic and that it got to the point and give you what you're looking for.

4:22
The chatbot seems too robotic. I'm gonna give it a three It is literally a robot AI thing so I don't expect it to tell me jokes but it seems like the right amount of robotic and that it gave you what you needed but he didn't feel uncanny if that makes sense.

4:38
The chatbot was welcomed during the initial setup. Yeah, I got a nice welcome message at the beginning which was nice.

4:45
Chatbot seemed unfriendly. No strongly disagree, chatbot seems pretty, pretty helpful. I didn't I didn't get ever ever get the vibe that the chatbox was unfriendly, chatbot, my bad.

4:57

The chatbot explained its scope and purpose. Well, I'm gonna give it a four, it explained what it does, and we'll how we can help you. It didn't go very in depth, but at the same time, I don't expect the chatbot to give me several paragraphs of explaining. I think it's very much oh, here's the basics. You'll figure it out as you use it, which is how I expect the chatbot to work in the first place.

5:20
The chat bot give no indication as to its purpose. No, the chat bots are pretty early in the process what its purpose was.

5:28
The chatbot was easy, easy to navigate. Yes, strongly agree. Like I've used chatbots with a number of sites. I think this is one of the easiest ones. You're not constantly doing guesswork. And... I figured after just a few tries of, if you put, you know, certain keywords, you're going to get what you need. So I say this is actually one of the best I've used so far.

5:47
It will be easy to get confused when using the chatbot. I would disagree with that. I think that the chatbot... No, I don't think you're going to get confused. I think most of the time you'll find what you're looking for. So I'm just gonna, I'm gonna say strongly disagree.

6:04
The chatbot understood me well. Yes, usually within two tries, whether it was putting in a keyword and selecting something within two actions, I was typically able to find what I was looking for.

6:16
The chatbot failed to recognise a lot of my inputs strongly disagree, there was only one input that I had to make any adjustment to. Sorry about that.

6:28
Chatbot responses were useful, appropriate uninformative. Five out of five. Pretty much everything that was given to me was useful and appropriate for what I was looking for.

6:39
Chatbot responses were irrelevant. Strongly disagree. I think they saved me a lot of running around and trial and error trying to find the information I was looking for.

6:49
The chatbot coped well with any errors or mistakes. Um, I don't think many errors or mistakes were made during this testing session. So I can't properly judge that. So

I'm going to give it a three but I think the chatbot, if mistakes were made, we'll probably cope with the wall since it seems to have done everything well, so far.

7:11
The chatbot seemed unable to handle any errors. Again, not a lot of errors occurred. What I'm probably going to say disagree. I feel like the chatbot would handle errors pretty well.

7:24
The chatbot was easy to use. I strongly agree.

7:29
The chatbot was very complex. I'm gonna give it a two.

7:34
The chatbot was fairly simple, in my opinion. And I think that's what you want a chatbot to do. I think it gave very, very much - Oh, you're looking for this and refer to this for the information. So I think it was pretty simple.

7:47
Thank you

Tester 6 - Tomeo

0:06
This is a test of the chatbot prototype that could help. That could help. Wikipedia editors with frequently asked questions and improve their experience. This is voluntary and anonymous, you will be asked some general questions, complete for quick tasks and assess 16 short statements about the chatbot. Remember to think out loud, and if there's anything you would like, if there's any information missing, if this chatbot doesn't respond, refresh the browser, okay?

0:41
You opened an account with Wikipedia years ago, and you want to change your silly password ask the chatbot for help. Okay, let's click, let's chat. And... I'm going to type. Simple. So you can change your password via this link. You can also find a link in this in your preferences. Okay. So that's where I could change my password. Personally, I thought that was very simple and very intuitive. So I've completed this by just typing password, and they figured I wanted to change my password. Please break this task, I thought it was very easy. I didn't even have to type change password or change my current password, I just typed the password that automatically intuitively figured out that I wanted to change my password and linked me to the proper link.

1:39
You're a blogger and you noticed someone used your content without permission, ask the chatbot to find out what to do about it. Use content without permission. Do you mean what should I do with if I find copyright violation? What prevents someone from contributing false information? So let's go ahead and type what should I do if I find copyright violation. So that's similar to using my content on Wikipedia. So we try hard to keep copyright violations out. But we don't always succeed if you're a copyright holder licencing or otherwise go to copyright problems and report this instance, in question. Thank you.

2:28
Alright, so that helps me in regards to copyright information. I'm not sure if I have copyright in terms of my content, but I'm going to assume I do. There is copyright, I can go to copyright problems. And this is information that allows me to that directs me to copyright problems involving text on Wikipedia. So this is very good information. There's an entire Wikipedia page on it. So personally, I found that to be very valuable and very simple. So I would give it a five. And the main reason is because it didn't specifically pertain to my use of typing just use content without permission. So something is intuitive that I thought would have given me a bit more direct results. But I did really like that they have multiple options to make sure that I get the right information based on what I was asking. So that was still relatively easy.

3:25

You see that one of the articles is missing a footnote, find out how to add, add. I'm going to type add a footnote, if you want to know how to add a citation, you can find choose the most helpful option for you, I'd probably click on Article... referencing for beginners. So this is very useful and allows me to get this information. And there we go. It has the footnote information right here. So it has the referencing for beginners was exactly what I needed. And I thought that was very easy. You gave me multiple different options that are most helpful based on my preferred method of use.

4:16
You're reading an article about a football player and you think that there's category missing for Uruguayan football players ask the chatbot if we can help you with the type of category missing, so how do I add a category? So how do I add a category? You could read more about creating new categories or use this video. So looks like I could categorise things and creating new category pages it even directs me to straight to this specific subsection that allows me to read on that pertains to my questions specifically. So personally, I found that to be very easy, you even also gave different methods for me to learn how to complete the task. So I thought this was extremely easy.

5:07
In the following session, you'll be asked to evaluate your experience with a chatbot by reading 16 short statements, complete the questionnaire by reading each tatement carefully and scoring with a number that best matches how you feel about the statement.

5:21
The chatbot personality was realistic and engaging. I'm going to say yes, personally, I thought the chatbot was very helpful. And I thought it was very engaging gave me multiple solutions, and was quite realistic.

5:36
The chatbots seem too robotic, I'm gonna disagree with that. I personally didn't think it was very robotic, it felt a bit natural, it felt like I was talking to a person actually, to some extent.

5:49
The chatbot was welcoming. I'm gonna agree to that it was they were excited. There were emoticons or emojis. And they were enthusiastic into in the initial setup.

6:02
Seemed very unfriendly. I'm going to strongly disagree with that. I personally thought this felt quite friendly. Everything was everything was very friendly and welcoming.

6:20

Explained it purpose, explained the scopes and purpose well. I would definitely so say so it explains everything. And I didn't have any questions and anything like that.

6:30
It didn't give any it didn't gave no indication as to its purpose. I'm gonna disagree with that. I mean, I guess it never really gave an indication of its purpose. But I personally thought it was very intuitive. But here it says wiki is here to help you edit the content. So you can below start below to get see your chat. So I don't think it ever stated its purpose. But personally, I didn't think it needed to.

7:02
It was easy to navigate. Definitely, if you type anything, it's usually quite intuitive. And it gave me the answers relating to my pertaining to my particular questions and concerns.

7:15
It would get, it would be easy to get confused when using the chatbot. I would not say so I think they gave multiple answers. And they gave multiple solutions to pretending to your specific situation.

7:26
The chatbot understood me well, I'm going to agree to that everything that I typed was... made a lot of sense. And it personally I felt as though it did understood my concerns well.

7:38
It failed to recognise a lot of my inputs, I'm going to disagree to the I think it recognised everything of my inputs. And if it never had, if it didn't have a specific understanding, they would give me multiple solutions to perhaps what I was asking about. So I didn't, I did not think that it failed to recognise my inputs.

8:01
Chatbot responses were useful, appropriate and informative. And we strongly agree to that I personally felt that all of the answers were both great, and I personally felt that it was all appropriate and very informative.

8:14
Chatbot responses were irrelevant, I did not think the responses were irrelevant. I personally thought that everything related to my pertaining question

8:23
It copes well with any errors and mistakes. I'm going to agree to that or anytime that they didn't have any specific information, they would give me suggested answers and suggested questions based on that, and gave me my information accordingly. So personally, I thought it coped very well with my mistakes and errors.

8:47

It seemed unable to handle any errors. I'm gonna disagree with that. Personally, I thought it was able to handle my errors because I just typed in key words, and they were able to cope with my errors to direct me to a specific questions that may be that may be more relevant and they have answers for.

9:08

So I thought it was very easy to use. For sure. I just needed to type in any key words and they would figure out and fill out the sentence for me so it was extremely easy to use.

9:16

I don't want to say it was very complex to use. I thought it's a complex system, but I don't think it was complex to use.

9:24

And Perfect. Thank you for your time. Hope this was useful.

Tester 7 - Steven

0:28
So I'm back to ask the chatbot for some help in... change my password on Wikipedia.

0:41
So, so I've got this come up wiki bar, I'm here to help you with Wikipedia editing.

0:53
Hello, I'm Wiki Chatbot. I'm excited that you are part of Wikipedia. How can I help you? So it's given me five options, editing policies and guidelines, referencing, images technical. So I think it will be technical. And that's skim them three options, how do I change my username or delete my account? How do I change my password? And the third one is, is it possible to download the contents of Wikipedia. So I want how do I change my password so if I click on that, so you can change your password via password change you can also find a link to this in your preferences. So that's good so that that solves that that was quite easy. So I just then go into password change and change your password and then login so that was all straightforward so very easy, very straightforward. The options were good they were the options were clear and they related to the task at hand and like I said before, but it's just very very straightforward and simple today

2:25
So I'm going to do the chat again so this time...

2:44
We're going to ask a question about someone using my content on Wikipedia without permission. So we're going to use the chat bot to see what I can do about that.

3:01
Let's chat again... so I'm just gonna say in this time permissions

3:55
So first of all, I just I think it would be easier if I could like start over.

4:03
He carries on with my old chat so so what I've done I've typed in

4:12
I need help with permissions and then I've got some options now just going to get back to those options at the top.

4:38
policies and guidelines. referencing no

5:16
so I don't feel that this is this is covered so I think this is quite difficult I'm having I'm having trouble with this one to be honest

5:43
so I'm gonna just type it in again somebody you know used content on Wikipedia without permission.

6:14
Okay, so this time, what should I do to find a copyright violation on Wikipedia?

6:26
Okay, so I can go to licencing or copyright problems and report the instance in question. So I could take it from there. So I have done that, that was a little bit more tricky that one. So, I've completed that task. Really, a bit trickier, you have to sort of phrase the question, it wasn't, the answer wasn't in the sort of options at the top. So it was a case then of framing your question correctly. So, it was a little bit he definitely more difficult. So, you'd have to spend a little bit more time on that, but got there eventually.

7:11
Say, for this one, I'm modelling the articles missing footnote. So I'm gonna type in how can I add a footnote? So I'll put that I've read the question, and how can I add a footnote and this is come back. If you want to know how to add a citation, choose the most hopeful options for you. And then you've got a choice. You can have an article, a video tutorial, or a code snippet. So I'd probably then want to go to Article referencing for beginning. So that's, that's done. That was quite, that was straightforward. That one. so straightforward. I've got the question sort of right and came up with it with an answer sort of straight away, I think as well at the top. Probably, if I went on to the referencing, it would be under there. But I was quite happy to sort of write out a question after the last task, and that seemed to work very, very well.

8:45
So I'm writing an article about a football player. And I think that there is a category missing for Uruguayan football players

8:59
Want to add that so I'm just gonna ask a question, can I add a category

9:12
So you can read more about creating new categories if you prefer, here's a short video. So again, that seems pretty straightforward. There's a tutorial and there's an article as well. So again, I think it's all about getting your your question right, making sure that it's written in good enough language that the chattbotst gonnau nderstand it,

it's not going to get the wrong end of the stick. So again, that that was seem pretty straightforward. So again, six very easy, making sure I get get the question and the information I want making sure I use simple enough language for the chat botnot to get in chat, but not to get confused

10:07
So the chat bots personality was realistic and engaging, I'd say yes, definitely. It's very, very helpful.

10:17
The chatbot seem too robotic. No, I don't think so. I think it was, you know, it's very, like I said, very helpful. Like I was talking to some real person.

10:28
Chatbot was welcoming during the initial setup. Yeah, very good, you know, make me feel welcome me to using it.

10:36
The chatbot seemed very unfriendly, definitely not... managed to get the answers that I wanted. And I'm quite sensitive. If somebody has been unfriendly or something, you know, like that. And I didn't feel like that at all.

10:51
The chatbot explained its scope and purpose. Well, yeah, definitely. Again, I thought, you know, he knew what it was doing knew what it was there to do. So I put four for that one.

11:05
The chatbot gave no indication as its purpose to disagree, it explained what it was there to do. What it could help with. So I put one.

11:16
The chatbot was easy to navigate. Again, I thought it was very good. You know, there was that one question that the I think it was the second question I did. I had a bit of difficulty with but once I got around that and sort of got used to the language and words I should be using, it was absolutely fine.

11:34
It would be easy to get confused when using this chatbot? I don't think so. I think if I picked it up quite quite easily.

11:42
The chatbot understood me well. Yeah. I think you know, it's keep the keep the language simple is important. Don't overcomplicate it. So I've put four for that one.

11:55
The chatbot failed to recognise a lot of my inputs. No, I disagree with that. It got things mostly rights.

12:02
The chatbot responses were useful, appropriate and informative. Yeah, all very good. And in most cases, he gave you more than one option, as well to you so that that was good.

12:15
The chatbot responses were irrelevant. I'll totally disagree. They all most of them, I think was just that again, that third that second question that I had trouble with, but that was my fault. So I've put I've put one for that strongly disagree.

12:29
The chatbot coped well with my errors or mistakes, yes. And then sort of offered alternatives and that was good.

12:37
The chatbot seemed unable to handle any errors. I didn't really encounter that. So I'm going to say I strongly disagree. The chatbot was very easy to use. Yes, very easy to use some detail but five for that one.

12:52
Chatbot was very complicated, complex. I disagree with that.

12:58
And above all, I found it very, very easy to use and I would be happy to use it in the future.

Tester 8 - Thara

0:04
Hey wiki.

0:05
All right. Hello wiki is here to help you with editing and contributing to Wikipedia. Click below to start a chat. Alright, let's see what the first task is.

0:14
This test is a chatbot prototype that could help Wikipedia editors with frequently asked questions and improve their experience. All right, the study is voluntary and anonymous. You will be asked to answer some general questions, complete four quick tasks and assess 16 short statements about the chatbot, please remember to think out loud during your test and tell us if there's any information missing. If the chatbot does not respond, please refresh the browser.

0:45
Alright. So let's see what the next step is.

0:48
You opened an account on Wikipedia years ago and you want to change your silly password. Ask the chatbot for help. Alright, so let's see.

0:57
How do I speak to this chat bot? And I'm going to click on let's chat. Wiki bot. Alright.

1:05
Hello, I'm wiki chatbot. I'm excited that you are a part of Wikipedia. How could I help you?

1:10
So I created a account on Wikipedia years ago, and I need to change my password. That right?

1:43
Did you mean how do I change my password?

1:45
Or was I supposed to click on something?

1:51

You can change your password via password change. You can also find a link to this in your preferences. Alright, so I did ask for help. Looks like I completed that task.

2:08
So let's see what's next. Please read the task and explain why you give such a score. Overall, the task was very easy. All I had to do on the homepage was click on let's chat and then the wiki bot chatbot appeared. And I entered my question. And the bot responded, how do I change my password. And I clicked on that and it sent me a link to change my password. So that was very easy.

2:14
You are a blogger and you notice that someone used your content on Wikipedia without permission. Use the chatbot to find out what to do what to do about it.

2:59
Alright, let's see. Someone use my content on Wikipedia without my permission. But could I do about it? Are you and then it gave me a couple of things that I could select what should I do if I find a copyright violation on Wikipedia? What prevents someone from contributing false information? Someone keeps writing negative things about me?

3:36
What should I do if I find a copyright violation? That sounds like the right one, right? We take this very seriously. We try hard to keep copyright violations out of Wikipedia. But we don't always succeed. If you're the copyright holder go to licencing otherwise could have copyright problems and report the innocence in question. Thank you. All right. And so I completed that task. I could either go to one of two links, so I'll see what the next step is.

4:14
This was also very easy. I'll give that a seven. And all I had to do was ask my question. And it gave me a series of related topics that I could click on. And I was able to click on the first topic and it gave me the exact answer that I was looking for. So great.

4:39
See what's next. You can see that one of the articles is missing a footnote, find out how to add it.

4:48
All right. So how do I add a footnote from my Wikipedia page, is that right? To my article on my Wikipedia page. Did you mean, how do I add an article to a category? How do I start? Or can I start an article about myself? How do I? How do I cite sources? That sounds like the right one. If you want to know how to add a citation,

choose the most helpful option to you. Article. If you're a new Wikipedia, if you're new to Wikipedia, this article could be helpful referencing for beginners.

5:47
Right, let's see what this is. Let's see is this here that numbered footnotes like this? All right, excellent. They're generally added either directly following the fact that they support or at the end of the sentence that they well while editing a page, yep, so it shows me how to add a footnote. So that's great. I have completed the task.

6:25
Please rate the task and explain why you give such a score. Um, that was also relatively easy. But this time, I had to click on two sub topics so it asked my question and usually I've been getting a list of topics where I could just click on one of them and it would give me my answer but it looks like I had to click on another topic which was article to be more specific about my question and it gave me the answer that I needed but overall it was very easy

7:05
You are missing an article about a football player and you think that there is a category missing for Uruguayan football players ask the chatbot if it can help you.

7:28
Category missing for Uruguayan football players. How do I add a new category to my Wikipedia page or article. How do I add a category. You can read more about creating new categories here right so creating category pages after you have determined an appropriate category name and notes parent category you're ready to create the new category. Alright, and it shows me how to do that with the example here so that's great great so looks like I have completed this task that was very easy. All I had to do is ask the question as usual. And I was given a link to see where I could learn how to create new categories and the webpage was very informative.

8:59
In the following section you'll be asked to evaluate your experience with a chatbot by reading 16 short statements. Please complete this questionnaire by reading each statement carefully and scoring the number that best matches how you feel about the statement.

9:16
Okay

9:22
The chatbot's personality was realistic and engaging. Um... realistic? It was definitely realistic for a chatbot. And it asked specific questions in order to get me the specific answers that I needed. So yeah, I would give that about a five.

9:59
The chat box seemed too robotic. I would give it about like a four, I definitely know that I'm talking to a bot. And when the bot says like, Did you mean and then gives me like a list of topics. I know it's definitely a bot, but it's not too bad. It's definitely a very intelligent bot.

10:24
The chatbot was welcoming during the initial setup. Um, let's see. Hi, I'm wiki chatbot. I am excited that you are a part of Wikipedia. How can I help you? Yeah, I would say it's pretty welcoming. Yeah, it gave me an introduction. And I knew exactly what I was supposed to do after I read the introduction. So that's great.

11:00
The chatbot seemed very unfriendly. Um, no, the chatbot was definitely very helpful. Yeah, like, when I asked about copyright violations, it was like, We take this very seriously. It's not something a bot would say. It's very humanising. So that's great. So I'll give that about a five.

11:30
The chatbot explained its scope and purpose well. The scope and purpose of the bot? I'm excited that you are a part of Wikipedia. How can I help you? Wiki is here to help you with editing and Wikipedia. Click below to start a chat. I guess so. I mean, it didn't really explain it itself. But it does say what the scope of the chatbot is and its purpose. And it's not. It's not anything too complicated.

12:27
The chatbox, the chatbot gave no indication as to its purpose. But I will definitely disagree. Yep, written in the introduction. It was like, I'm excited that you're a part of Wikipedia. How could I help you? So it was there to help. So I would give that about a one.

12:52
The chatbot was easy to navigate. Yeah, I'd give out about like a five. All I had to do is ask questions and I got my answer.

13:02
It would be easy to get confused when using the chatbot. Um, I would give that like a three. I am familiar with like caveman speak. So that that's like typing like keywords for things that I'm using, that I'm looking for. And I know not to talk to a bot like I would a human being. And I don't expect it to understand me. If I were to talk to it as a human being. So I don't exactly know what the entire idea of this is. But I'm assuming you guys are trying to make the bot seem as human as possible. If that's the case, I would Yeah, just give that a three because it is a bot.

14:17

The chatbot understood me well. Yeah, definitely agree. I asked questions related to Wikipedia, and it definitely had information about it.

14:32
The chatbot failed to recognise a lot of my inputs. Strongly disagree. I asked questions that were very specific and straight to the point and it recognised my inputs

14:51
Chatbot responses were useful, appropriate and informative. Yeah, I would definitely say so. It provided links and information and it showed empathy.

15:15
I disagree, it wasn't irrelevant. It was very relevant.

15:24
The chatbot coped well with any errors or mistakes. I didn't experience any errors or mistakes. But if I would consider maybe like that one question that I asked about the footnote to my Wikipedia page. It asked additional questions to see if they could get me to the right answer. So I would say that it did cope well. So that's fine.

16:03
The chatbot seemed unable to handle any errors, or I'll give that a one. Because I didn't experience any errors.

16:15
The chatbot was very easy to use. Yes, it was.

16:25
The chat botwas very complex. I don't think so. Yeah, nope.

16:38
That's all thank you for participating. Thank you.

Tester 9 - Elisabeth

0:11
Hello. Wiki is here to help you with editing and contributing to Wikipedia. Click below to start a chat.

0:22
Okay.

0:27
This is a test of a chatbot prototype that could help Wikipedia editors with frequently asked questions and improve their experience. The study is voluntary and anonymous, you will be asked to answer some general questions. Complete your complete four quick tasks an assess 16 short statements about the chatbot. Please remember to think out loud during your test and tell us if there's any information missing. If the chatbot does not respond, please refresh the browser.

1:16
You opened an account on Wikipedia years ago and you want to change your silly password. Ask the chatbot for help.

1:29
Okay. See. Ask the chatbot.

1:40
Okay, I just clicked on let's chat. A screen popped up. And it says I'm here to help you with Wikipedia editing. Hello, I'm wiki chatbot I'm excited that you are a part of Wikipedia. How can I help you now they have on 1,2,3,4,5 little boxes to click on: editing policy and guidelines, referencing, images, and technical. I clicked on the type to change my password. I'm typing now. You can change your password via password change you can also find to link to this in your preferences. Okay gave me an answer. Okay. Now I'm going to the next task.

3:01
It is asking me why do I give a rating, high rating for the task? Because that when I type my answer my question my question it responded quickly. I was very easy to do this task about changing my password.

3:27
You are a blogger and you noticed that someone used your content on Wikipedia without permission. Use the chat bot to find out what to do about it.

3:42
Okay. You are a blogger and you knows that someone used your content on Wikipedia without permission. I'm about to ask the chat but the question. I'm typing it

now. When my information is used on Wikipedia. My content was used. Okay, now I clicked on one of the boxes provided as as what prevents someone from contributing false information. Wikipedia is content control mechanisms are reactive rather than preventative. Anyone can go to almost any page and change the information to make it false or misleading although the majority of edits attempt to improve that at Wikipedia. Vandalism is frequent you can help us with vandalism by editing the articles. I picked a high score because they answered the question in a timely manner.

6:12

The next step. You can see that one of the articles is missing a footnote find out how to add it. An article is missing a word just typed a question see an article is missing a footnote a missing. If you want to know how to add a citation, choose the most helpful option for you. They gave me three options article video tutorial, code snippet. Okay, so I click on Video Tutorial. They just got back to me with a video you can watch this video to learn how to add citations on Wikipedia. I click number seven because it is easy to ask the question and they get back to me right away

8:10

You are writing an article about a football player and you think that there is a category missing for Uruguayan football players as the chatbot if it can help you.

8:36

I'm typing. I think there is a are missing for... You can. I'm typing to get the answer.

9:46

I see answer. I'm typing help with existing category. Okay, it gave me options.

10:14

How do I add a category? Can categories be renamed moved or redirected? How do I delete a category? How do I add an article to a category? What should I do when I see plagiarism? Those are the options given to me

10:43

I pick number seven because it provided answers to the question I had about missing category

11:00

In the following section, you will be asked to evaluate your experience with a chatbot by reading 16 short statements. Please complete this questionnaire by reading each statement carefully and scoring with a number that best matches how you feel about the statement.

11:21

Okay. The chatbot's personality was realistic and engaging. I will say five.

11:35
The Chatbot seem too robotic. No, it didn't. I'm gonna give you a five No, sorry. Strongly Disagree. Didn't seem robotic.

12:02
The chatbot was welcoming during the initial setup. Yes. Five. Totally agree.

12:12
The chatbot seemed very unfriendly. Strongly Disagree. One.

12:19
The chatbot explained its scope and purpose. Well, five strongly agree.

12:29
The chatbot gave no indication as to its purpose. One, strongly disagree.

12:41
The chatbot was easy to navigate. Five, strongly agree.

12:48
It would be easy to get confused when using the chatbot. One, strongly disagree.

12:56
The chatbot understood me well. Five. Totally agree.

13:04
The chatbot failed to recognise a lot of my inputs. I'm going to say disagree. I mean, if you type in some something a little bit wrong, a might come up with something different. But it did work out when I typed what I was looking for so shrine disagree.

13:36
A chatbot responses were useful, appropriate and informative. Strongly Agree.

13:44
Chatbot responses were irrelevant. Strongly Disagree. One.

13:50
The chatbot coped well with any errors or mistakes, strongly agree.

14:03
The chat bot seemed unable to handle any errors, strongly disagree. One.

14:10
The chat bot was very easy to use. Five, strongly agree.

Tester 10 - Israel

0:01

Okay, okay, here we are. Okay. We're going to check the task to see what's going on or what we need to do. So step number one, that is 27 steps. Whilst this is a test of chatbot prototype that could help Wikipedia article with the frequently asked question and improve their experience. The study is voluntary and anonymous, you will be asked to answer some general question complete four quick task and assess 16 short statement about a chatbot. Please remember to think out loud during your test and tell us is there is any information missing. If the chatbot does not, does not respond, please refresh the browser. Thank you for your participation.

1:00

Okay. Let's go guys. Let's go. I'm going to click on the next step, which is step number two, and says you open an account on Wikipedia years ago, and you want to change your silly password. Ask the chatbot, ask the chatbot for help.

1:26

Okay, let's see if I can say that. Okay, they're here. Watching the, the webpage. And I'm seeing the webpage and says hello. Wiki is here to help you with editing and contributing to Wikipedia. Click below to start chat. Okay, I will click the button let's chat. And I have to change my silly password from years ago. So how can I do that? Let's I'm thinking if I need to type or I need to do anything else guys. Sorry to be in this way. But I'm wondering what can I do. You opened an account on Wiki years ago, change password, how to change, ask th chatbot for help, okay, I would do that. So how can I help you chat says I will say I want to change my password. Okay, what I'm looking here in the chatbot is something good because he's not on the top on the bottom right corner is in the middle. So I cannot be distracted for another things. And now we'll press the button send and chatbot is doing something. And as for me, you can change your password by password change. And this password change word is a link. You can also find a link to this in your preference. Okay. Okay, two options, the chatbox gave me two options. So I have completed the task. Yes.

3:34

Okay, please rate the task and explain why you give such a score. Over this task was, overall this task was very difficult. Okay. Overall, this task was very difficult, very easy. Guys, for me, this task was very easy. At the beginning, I was wondering why what can I type or write but when I did just to type in, I want to change my password, chatbot gains to say is two ways to do it. So for me, it's amazing.

4:35

So step number four, says you are a blogger, you notice that someone use your content on Wikipedia without permission use the chat box chat bot sorry to find out what to do about it. So someone uses content on WIkipedia. Okay. So I will ask the

chatbox for help to see my, I don't know, maybe permissions or what can I do? Okay, your content on Wikipedia without permission. Okay. How can I? I don't know what what can I? What I what I? What can I look for really? Because if someone took my... used my content without without permission. I don't know what what what to type of, right? So. Okay. I would say to the chatbox wiki bought help me to change the permission of my content. I don't know if I'm okay with that. The cat bogus is showing me like five chances to do this this one like Did you mean in the chatbot is asking me? Did you mean? What should I do? If I find a copyright violation on Wikipedia? The other one says Why is someone changing my article? How can I can I how do I change my password? What prevents someone from contributing false information? Categories renaming? Okay, the the one for me that makes sense is the first one. What should I do if I find a copyright violation? So I will click on the bottom. What should I do? And chatbox show me, message me, we take this very seriously, we try hard to keep copyright violation out of Wikipedia. But we don't we don't always succeed. If you're a copyright holder, go to licence otherwise, go to copyright problems and report the incident in questions. Thank you. Okay chatbox gave me an answer for what can I do so I have completed task.

7:40

And I'm going to the next task. Please read the task explain why you give such a score and our work. Okay. I will say this one was hard for me. I said number two was very difficult. Why? Because I didn't understand or in my mind one was something spinning or was this idea of Wikipedia is free, is free for everyone to see it. How? But it's just my understanding of Wikipedia sorry guys, but I was thinking that Wikipedia is free. Why? If someone used my content will be like a violation but i don't know i don't know guys. But after all of that, after all my explanation I can I can tell you this that this task was hard for me. So I'm going to the next one.

8:51

You can see that one of article is missing a footnote find out how to... you can see that one of the article is missing a footnote find out how to add it. Okay. Okay, again, I will use the the chatbox to ask. And what I'm going to what I'm going to do now is to asking ask the chatbot what can I how can I how can I add a footnote? And I will do it like just typing How can I add a footnote? How can I footnote and chatbot is looking something for me. It says if you want to know Oh, how to add a citation, choose the most helpful, helpful option for you. Article, video tutorial, code snippets. Great. So I am asking for a footnote I will use the question that you gave me, or the sentence that you gave me to write in the chatbot because I didn't understand how the chat bot relates this article maybe in the article part, that is a footnote explanation for so maybe. Okay, I will say that. So, if you are new to Wikipedia, this article could be helpful referencing for beginners. Okay, that's good, because is the one that I'm looking for. So, task number six, you can set you can see that one of the article is missing a footnote. Okay, I have completed the task guys. Sorry, I am lost at the moment.

11:08

So I am going to click the button next. Because I, I already got my idea, my answer, answer. So please rate the task and explain why you give such a score. Okay, that task was confused, not hard. It was confused. For that reason I gave this task a four. Because I, I do not have the this mindset to a person who is a beginner in Wikipedia. So I would say a number three, is a little bit difficult for me. So that's it. That's it for that reason I did it. I'm going to next task. You are writing in an article about football player and you think that there is a category missing for Uruguay football player ask the chatbot if it can help you.

12:19

Okay, I'm getting used to the to the question that I have to make for get good answers. So I will ask to the chatbot. How can I... beat, category missing, yeah? How can I add the category? I will said to the chatbot. And let's see what chatbot says.

12:56

You can read more about the categories if you prefer there is a short video. Good, good. So I had I had my first my answer. Okay, I have completed the task. And the chatbot gave me the answer.

13:13

So please read the task and explain why you're... No, no, now I'm getting used to the chatbot and I can tell you what it was easy. I will give. I'm going to give. I'm giving the rate six to this question because it was easy. But just because I'm getting used to use the chatbox and let's go for the next step.

13:45

Step number 10. Is the following section you will be asked to evaluate your experience without chatbot by writing 16 short statements Please complete this questionnaire by reading each statement carefully and scoring the number that best matches how you feel about the statement.

14:08

Okay, good. Let's go. The chatbox personality was a realistic and engaging. Guys I I have to say... if I didn't know what I'm looking for, in what is the second name of something that I'm looking for. And will be will be not not engaging not realistic. But with the question that you gave me. I can tell you that I am agree, I'm agree with I am agree with the chatbox realistic and engaging. I will say that I strongly agree number. I will rate with number two. I'm going to the next one.

15:08

So the chatbox seem to be robotic? Not at all really? I will say number three is in between of agreement. Yeah. Because... gave me options, a chatbot gave me options, the chatbot suggest me things that a person who will suggest me. So I will say, strongly disagree number two.

15:44
And let's, let's go for the next one, which is the chatbot was will come in during the initial setup. Definitely, as strongly, strongly agree with that. And also, like I said before, I like this chatbot because it's in the middle. It's not in the corner. It's for me in the middle.

16:10
So, um, let's go for the next. Question 14 says, tchatbox seems very unfriendly. Strongly Disagree. It looks it looks on friendly. Yeah. So why? I didn't have the sense that it's not a person. Yeah. Like I said before, gave me options. Gave me a sack, what I am looking for and also gave me advices to do something to do anything.

16:54
Let's go for the next one. And says, the chatbox is playing its scope and propose well, purpose well. Definitely, definitely. According to your question, and you will be answered. And that's for me. I'm okay. With with the scope.

17:16
And let's go for the next one. Number 16 says the chatbot gave me no indication as to its purpose. Strongly Disagree, guys. I always found the one I'm looking for doing. Doing my search. So I'm strongly agree what you said.

17:42
Number 17. A chatbox. The chatbox was very easy to navigate. Definitely. Total victor. Yeah. What else do you need on the chat box? The input field, the send bottom, just that. That's good.

18:09
Let's see number 18. It will be easy to get confused when using the Chatbot. No, no, definitely no. What was confusing for me was the question that I need to type or write for the chatbox. Yeah, just that. But by the way, chatbot gave me options. I'm okay. So I'm gonna strongly disagree with that statement that you bring me or brought me. And let's go for the next one.

18:48
Number 19 says the chatbot understood me well. Definitely yes. But not at all. But I would say number two, I am strongly disagree. Not, sorry. I am a strongly agree with the chatbox under understanding to me. And number four is my rate and it's a strongly agree. Let's go for the next one.

19:20
Number 17 says the chatbox failed to recognise a lot of my input. Strongly Disagree number one. Every time that I type something Chatbox came with something related. So I'm okay with how come of the chatbox.

19:41
Number 22 says chatbox response were useful, appropriate and informative. I strongly agree about it. Every question that I shout to the wiki bot was, well, answer. Definitely. For me, it happened for me.

20:12
Let's see what is in the step number 22. Chatbox response were irrelevant? Nope. I strongly disagree with that. I'm always I had this the sense of a live in person behind the chat. Because always, I had, well instructor answer.

20:43
So let's go for numbers 23 which is the chatbox cope well with any errors or mistakes? Definitely. Yeah. I, I, I, I type, I miss type category word. But the chatbox show me in a good way. So strongly disagree. Not sorry, strongly agree.

21:14
Yeah. Next one, which is the chatbox seems unable to handle any error. I strongly disagree is just why is for that reason, I rates this explanation of category misspelling words. So I am agree with the error handling of chatbox.

21:43
Next one, which is chat box, the chat box was very easy to use. I strongly, strongly agree. There is nothing to say about it. Next, the chatbox why I said there is nothing to say it's because I am glad to use it is just definitely good.

22:08
Me. Well, my, the chatbot was very complex. I strongly disagree with that, number one. Nah, nah, nah, why do you ask that when you have a chatbot, a simple simple chatbox, with all that you need to ask and get answer from them. I'm not agree with that. I mean, I strongly disagree with what you said. And but it's, it's okay to ask. definitely. But you are doing so well with this chatbot. Thanks.

22:48
So the next the last one says that. That's all thank you for your participating. And that's all for me. Thank you so much. That's good chatbot.

Tester 11 - Latasha

0:02
This is a test of the chatbot prototype that could help with Wikipedia editors with Frequently Asked Questions improve their experiences. The study is voluntary and anonymous, you'll be asked to answer some general questions complete four quick task and assess 16 short statements about the chat bot. Please remember to think out loud during your tests and tells us there's any information missing. If the chatbot doesn't, does not respond, please refresh the browser. Thank you for your participation.

0:26
So you open an account with Wikipedia years ago, and you want to change your silly password. Ask the chatbot for help. Okay, so looking at this here, so hello. Wiki is here to help you with editing and contributing to Wikipedia. So click below to start a chat. So I guess I click on those chat. So if it's a matter of wanting to change my password. So it says hello, I'm wiki chatbot, I'm excited that you're a part of Wikipedia, how can I help you? So there's editing, policies and guidelines, referencing, images, and technical. So I think it was a password. For me, I feel like I couldn't necessarily type in password just because those are here. But you know what, I would just put in password and see if it actually updates, or provides me anything else. Okay, so it says you can change your password by password change, you can also find a link to this in your preferences. Okay, so go to password change. And then it would, I guess, relate me to this information here. Okay. So I have completed the task.

1:30
So please rate the task and explain why you get such a score. So overall, this task was one very difficult to five very easy. So I think for the most part it is fairly easy, but because they're like limited with there's like with respect to editing policies and guidelines, referencing images and technical, I wasn't really sure if I can put in some more thing that didn't fall within, or I didn't think fell within there. So I think if they actually had an other or even just kind of a statement that says, you know, is there any or you know, where you can relay your information on your own, then I think is more transparent about that. So that's why I'll give it a six.

2:04
Okay, so you're a blogger, and you just know that someone used your contents on Wikipedia without your permission, so use the chat bot to find out what to do about it. Okay, so I guess this is with respect to referencing so I will click on referencing and son Harrison common questions regarding references of how do I source sources? And how do I cite Wikipedia. So it doesn't really fall in within that. Just let's go to editing... How to edit a page and how to create a new page and where can I get feedback? Policies and guidelines... What should I do when I find a factual error in Wikipedia? What should I do when I see plagiarism? And can I see an article about

myself and my company? So I guess this is a situation of plagiarism. So click on that so if you see an article including inadequate contribute text or images or worst copyright violations, please tell us about it in the top page of that article, and then report to the copyright problem. So it was this answer helpful. Okay, so I guess I found it out eventually in policies and procedures. So I guess I completed a task.

3:11

Please, rate this task and explain why it gets a score. And overall this task was... So again, I wasn't entirely certain as to where I would find it. But you know, so I did kind of a trial and error but eventually I did find the information as to where it could be found. And then you know, I didn't find that is helpful I can give you a rating. So overall, I would say that it was fairly easy to say I find finding it eventually.

3:34

Can you see that one of the articles is missing a footnote so I'll find out how to add it so that would again feel that it would be referencing. So... I guess, let me just type in a footnote, so if you want to know how to add a citation choose the most helpful option for you. So yes, go to article. I guess we can go to referencing for beginners and see how how that works. And then maybe you kind of give you information for footnotes and stuff like that... but I'm just not entirely certain because you forgot to do it in an article but I think it's probably you will be able to say you're referencing it for beginner so be found within there as to how to say you're adding it so complete this task.

4:43

Please read the task explain why case reading. So overall, So overall, I think this reading that this was fairly okay to say you're finding it. I think I don't mind necessarily being redirected to another page. But I feel like I prefer to kind of get a little bit more like a summary as to what I can do within the chatbot. And then if I needed further information, I can actually go to the article about that.

5:11

So you're writing an article about a football player, and you think that there is a category missing for Uruguayan football players. ask a chatbot, if it can help you. Okay, um, so, um, I guess it's really editing. So how do I edit a page. So that's great you want to start your editing journey. So to edit the whole page at once, click the edit this page at the top to just one session for the edit link to the right of the section heading. And you can learn more about editing here or just watch this video. So this is kind of more what I expect to say I'm seeing instead of actually kind of being referenced to another page, just kind of getting synthesis to what you can see you're doing from within there. So I have completed the task. So please rate the task and explain why it gave such a score. So overall, I think this one was very easy, I felt it was a lot more intuitive as to what is they are doing for this one.

6:09

In the following sections you will be asked to evaluate experience with a chatbot rereading 16 short statements. Please complete this questionnaire by reading each statement carefully and scoring with a number that best matches how you feel about the statement.

6:20
So the Chatbot personality was realistic and engaging. So I didn't think that it was really that engaging, I think I've seen some that are a little bit more realistic, I think this one was more robotic. So I'll give it a neutral setting

6:33
The chatbot seemed to robotic. So I say strongly agree, I didn't feel it had much of a personality.

6:39
The chatbot was welcoming during the initial setup. So you know, the initial setup was just like a regular hello and wiki Chatbot. And we say that you're part of that Wikipedia, how can I help you? So I think it was relatively welcoming. It would have been nice if it was able to identify the user and stuff like that. But for the most part, it was fairly, it was okay.

7:00
The chatbot seemed very unfriendly. So I wouldn't say it was fairly unfriendly, I think it was it was neutral. I didn't think it was overly welcome overly friendly, but it wasn't unfriendly.

7:11
The chatbot explained his scope and purpose as well. So I felt like for some things, it referred you to kind of like a longer pace to go through and some things they actually, you know, identify what you can do, I would expect it within the chatbot to kind of get a lot more information and context. So I don't necessarily have to review a full page as to how to get to everything.

7:29
So the chatbot gave no indication as to its purpose. So I don't think so. I think for the most part, let you know where where you can find resources and stuff like that, it could be a little bit more improved. But for the most part, It gave some idea of its purpose.

7:46
The chatbot was easy to navigate. So I would say yes, you know, the links work to say you're redirecting to where you need to go to, you have kind of like the pre select, or maybe the frequently asked questions that you can say you're clicking out from the front. So yes, it was easy to knit. So you're navigating,

8:01
It would be easy to get confused when using the tablet. I also strongly disagree, I don't feel like I had any, you know, confusion as to what to do or where to go next. I think for the most part, it was very intuitive.

8:13
The chatbot understood me well. So I use very simple language when I was typing things. And so I felt that, you know, it understood me fairly well as to what I was going through, I just kind of had to make sure I was selecting the right area, especially for the pre selected frequently asked questions.

8:29
The chatbot failed to recognise a lot of my inputs. So as I strongly disagree, I think I didn't necessarily kind of go with too much complicated sentences and stuff like that. So for the inputs that I did indicate, it kind of knew what I was looking for.

8:44
The chatbot responses were useful, appropriate and informative. So it's a neutral setting. Some of them were very informative, especially the last one, when it actually showed me showed me a video. Some of them just kind of redirected me to a website. So you know, referencing for beginners, I didn't really feel like that's that useful, I would want to have a little bit more insight onto that. And then for the password one either I can link on it to changing my password or kind of giving me an error to say I'm going to so for the most part it was, for the most part, it was fairly appropriate, useful in some cases no. But that's why I'm giving you three.

9:23
Chatbot responses were irrelevant. So I don't think any of them were irrelevant. I think really it was just me maybe selecting the wrong area. But for the most part they were all fairly relevant to what I was looking for.

9:34
The chatbot coped well with my any errors or mistakes. I didn't really come across any errors or mistakes, so I'll have to give it a neutral seven because I I can't really give you a proper estimation of that because I don't feel like it actually came into errors or mistakes.

9:55
The chatbot seemed unable to handle any errors. Again, I'll have to give it a neutral setting because I didn't really feel like it came into any situations where I had to handle an error.

10:04
And the Chatbot was very easy to use. So yes, I think it was easy to understand that I can use a kind of frequently asked questions. If that wasn't the case, I felt it wasn't

clear that I can kind of type in something that felt without that, like without that didn't fall within this FAQ section. So that's what I'll do with a four.

10:21
And a chatbot was very complex. So I don't think it was very complex. I think it was very simplistic and how it was laid out. And that's it.

10:29
So that's also thank you for your participation. Okay. You're welcome.