CHARACTERISING USER ENGAGEMENT OF PARENTS WITH THE AURORA CHATBOT

Diana Rangel Lopes de Campos Liebetrau

CMPDIA001

Supervised: Professor Melissa Densmore

Co-Supervisor: Francisco Nunes

Submitted in partial fulfilment of the requirements for the degree of Master of Science in Information Technology

October 2023

University of Cape Town
Department of Computer Science

DECLARATION

I, DIANA RANGEL LOPES DE CAMPOS LIEBETRAU, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate …………………………………

Date: …………………………………

ACKNOWLEDGEMENTS

I would like to extend my sincere appreciation to all those who have played a vital role in the successful completion of my thesis. Their support, guidance, and motivation have been priceless during this academic expedition.

First and foremost, I extend my appreciation to my professor and supervisor from the University of Cape Town, Melissa Densmore. Her expert guidance and willingness to lend a listening ear during this research's ups and downs have been pivotal in shaping this work.

I am also very grateful to Francisco Nunes from Fraunhofer Portugal AICOS for his exceptional supervisory role and extensive knowledge, consistent support, and belief in me. His insights and excellent career advice have enriched my understanding.

To my family, a big thank you for their constant support and belief in me. Their love and encouragement have been my strength throughout this challenging journey.

Lastly, I want to extend a furry thanks to my faithful dog companions, Floki, Mini and Chewbacca. Their presence brought comfort and joy throughout the early mornings and late nights spent working on this research, making the journey more bearable.

Thank you all for being an integral part of this journey and helping me turn this endeavour into a reality.

ABSTRACT

Chatbots have the potential to enhance everyday life by providing information, customer support, personal assistance, and more across various sectors, making them versatile tools for modern living. In childcare, they have the potential to provide parents with relevant and acceptable childcare information. This potential arises from the growing disparity between the abundance of childcare information available and parents' ability to access information tailored to their specific needs [5]. Aurora, a rule-based Facebook Messenger chatbot, was developed to support parents and caregivers in caring for their children by providing accessible and comprehensible childcare information, targeted for Portuguese-speaking parents [6].

This research analysed the Aurora chatbot's chatlogs dating back to October 2018 up until September 2021 to scrutinise the interaction dynamics between Aurora and its users. Through this research, the objective is to delineate user engagement patterns, identify topics discussed, highlight topics outside the chatbot's knowledge domain, and assess the dynamics and quality of the conversations. The methodology used to achieve this objective encompassed text pre-processing, engagement metric extraction, topic analysis, content analysis, and sentiment analysis.

The analysis of 1043 Aurora chatlogs indicated that only 718 (69%) users actively interacted with the system. These interactions predominantly occurred during lunchtime and late at night. The data showed that approximately 80% of conversations centred around baby sleep, 13% pertained to breastfeeding, and 7% focused on healthcare topics. While Aurora responded appropriately to in-domain questions, challenges arose when users' questions contained multiple topics, such as questions about the ability to breastfeed while taking certain medicines. User feedback was positive, with an average star rating of 4.37/5 (continuous scale), despite the modest sentiment score of 0.119 in the rating comments.

The research classified users into four groups, based on paid and free subscriptions, each highlighting specific engagement patterns. Users who had the paid subscription showed a 243% increase in interactions and a 162% increase in extended use of the chatbot.

These insights serve as guidance for Aurora's next iteration, highlighting the importance of recognising different user types and refining areas of shortfall. Additionally, this research contributes to expanding the scholarly corpus on how parents interact with chatbots.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

01

INTRODUCTION

AURORA

1 Introduction

A chatbot, often referred to as a conversational agent or dialogue system, is an Artificial Intelligent (AI) program designed to interact with users through human-like conversation [118]. Chatbots are progressively being implemented in several industries and within different contexts. Their capability to provide uninterrupted support and reduce the demand for human resources has generated significant interest in multiple sectors, making them a crucial tool for managing interactions with both customers and other users [22, 63]. Within the healthcare industry, the growing potential of chatbots is evident by their increasing adoption for a variety of functions [52], from addressing user queries and updating health records to disseminating disease-specific information and initiating actions based on user responses [7].

According to the existing literature, the implementation of chatbots in healthcare settings remains relatively limited, even though chatbot solutions have rapidly grown [129]. Likewise, there is a dearth of literature available on parenting and childcare chatbots, revealing a notable gap in the current understanding of this domain [44, 45]. One such childcare chatbot is Aurora, a Facebook Messenger application designed to support Portuguese-speaking parents and caregivers by providing accessible and comprehensible childcare information [6].

Aurora's significance stems from its design as a chatbot tailored to cater specifically to the needs of Portuguese-speaking parents, addressing the linguistic and cultural requirements of this demographic. This demographic is particularly relevant for the research due to their challenges and preferences that parents may encounter when seeking childcare information [122, 140]. Portuguese-speaking parents may possess distinct cultural practices and preferences related to childcare [122], and Aurora's ability to accommodate these specific needs can positively impact their parenting experience.

This research focuses on the second version of Aurora, which was operational from October 2018 to September 2021 and facilitated 1043 unique conversations between users and the Aurora chatbot. Through this research, the objective is to delineate user engagement patterns, identify topics discussed, highlight topics outside the chatbot's knowledge domain, and assess the dynamics and quality of the conversations. The primary contribution of this research lies in the unique insight it provides into how users interact with a childcare-focused chatbot. The practical insights gained from analysing Aurora's chatlogs could potentially fuel advancements in the realm of chatbots for parenting support.

1.1 MOTIVATION

The domain of childcare chatbots presents an area of increasing importance, yet it faces significant challenges due to the scarcity of well-defined guidelines and limited research studies [45]. However, there is a noticeable disparity between the abundance of childcare information and the capacity of parents to access information that caters to their individual requirements [5]. While medical experts, social media, parenting blogs, and baby care apps, like BabyCenter [11] or BabySparks [12], provide childcare information, they do not offer the capability for parents to ask personalised questions about their infants 24/7. In addition, parents often become overwhelmed by the sheer volume of contradicting information [35, 80]. In this context, chatbots can potentially make a significant positive social impact by offering parents personalised information that caters to their specific needs [5]. Although the initial development and deployment of chatbots can be costly, the long-term benefits and scalability make them a valuable tool for supporting financially constrained parents [140].

The development and implementation of AI powered chatbots offer promising solutions to address the current limitations in existing childcare resources and to support parents more effectively [140]. By providing real-time, personalised assistance, chatbots could potentially bridge the gap between the abundance of information and parents' ability to access specific, tailored guidance, which could be lifesaving. However, the adoption of chatbots for parenting support remains an area that requires in-depth exploration and understanding.

Amidst the growing interest in AI chatbots for parenting support, the Aurora chatbot served as a specialised platform for parenting support [6]. By offering personalised support in their native language, Aurora addresses language barriers and enhances accessibility for this specific group of parents, potentially improving their parenting experience.

Notwithstanding the rapid increase and potential of AI-powered chatbots, there is limited literature on the practical application of chatbots in the healthcare sector, and even fewer solutions for parenting and childcare [129]. Exploring the potential impact of Aurora, which operated from October 2018 to September 2021, is essential to inform the development of future chatbot versions. This research seeks to address these gaps by analysing Aurora's chatlogs, providing a unique opportunity to gain valuable insights into user usage patterns, user experience, content preferences and effectiveness, and to better understand the needs of parents who interacted with a parenting chatbot.

The significance of this research lies in its potential positive social impact, particularly for parents living with financial and geographical exclusion [140]. By providing accessible and comprehensible childcare information, chatbots like Aurora can empower parents, reduce stress, and improve the overall well-being of both parents and their infants [140].

This research can contribute valuable knowledge to the broader field of AI-chatbot development for parenting and childcare support. Furthermore, providing valuable knowledge can enhance the development and implementation of future AI-driven childcare resources [129]. Through this exploration of Aurora's chatlogs, the research seeks to identify opportunities for refinement, understand user preferences, and help make informed decisions to enrich the next generation of parenting chatbots. The findings of this research have the potential to lay the groundwork for innovative solutions to better serve parents worldwide, ultimately fostering a nurturing and supportive environment for infants and parents alike.

1.2 Problem Statement

This study is the first systematic evaluation of the Aurora chatbot. As a result, there is a lack of understanding about its efficacy in delivering information to parents and caregivers. Without these insights, it is challenging to assess the impact of Aurora on its user base and make necessary improvements for future iterations. This lack of knowledge also hinders the progress of AI chatbots in the field of childcare support, as they could benefit from the learnings derived from Aurora's operation. Therefore, the problem requires an in-depth exploration of Aurora's chatlogs to understand how this user group utilises AI chatbot technology for childcare support.

1.3 Research Questions

Through the examination of chatlogs generated by Aurora, this study's purpose is to address the following research questions:

> ***RQ1****- What specific user engagement patterns related to query frequency, response satisfaction, and interaction duration can be identified from the chatlogs of parents interacting with the Aurora chatbot?*

To address RQ1, users' conversation content and meta-data were analysed to uncover different user types, thus showing unique engagement characteristics. This exploration may aid in creating personas that embody varying levels of engagement.

***RQ2** - What are the topics in the discussions between parents and the Aurora chatbot, and how do these topics reflect the specific childcare information needs and concerns of parents?*

Answering RQ2 entailed the identification of discussion topics, aiding in understanding the specific childcare needs and concerns of Portuguese-speaking parents. These findings offer valuable insight into the distinct childcare challenges of Portuguese-speaking parents and guide the delineation of new thematic areas for Aurora's subsequent iterations.

1.4 APPROACH

Two research questions guided the investigation into the Aurora chatbot, which employed a multi-tiered analysis approach to examine a comprehensive dataset. Adhering to strict privacy standards, included chatlogs from 1043 users.

The data underwent pre-processing, which transformed and organised the data for effective scrutiny. Subsequently, key metrics were extracted to understand user engagement, user types, usage patterns, conversation quality, satisfaction levels, user sentiment, and themes discussed. Additionally, sentiment analysis provided valuable insights into user sentiment towards the chatbot.

For more specific details and justifications, refer to the Methods section of this paper.

1.5 CONTRIBUTION

This research makes several contributions to the field of AI chatbots for parenting and childcare support. By analysing the previously unassessed chatlogs of Aurora, this research sheds insight into user challenges, engagement patterns, and the chatbot's performance. Through a comprehensive understanding of user interactions, this research aims to enhance the user experience with the next version of Aurora.

Given that there is no standardised approach on how to analyse chatbot conversations, this research contributes to methodologies employed for chatbot analysis [23]. The approach used to analyse and evaluate Aurora's chatlogs introduces a different approach to the field of AI chatbots for parenting support. This research employed a holistic approach to the techniques for assessing chatbot performance by incorporating a blend of qualitative and quantitative methods. These methods not only deepen the understanding of user interactions but also provide a framework for future studies in the domain of AI-driven chatbots.

Furthermore, this research fills a critical gap in the literature by providing in-depth insights into the effectiveness and limitations of AI-driven chatbots tailored to Portuguese-speaking parents. Analysing conversation patterns and topics discussed offers valuable data to optimise the design and content of Aurora and other parenting chatbots.

The empirical findings of this research contribute to the advancement of Aurora's capabilities and hold broader implications for the field of AI-chatbots for parenting support. The knowledge gained from analysing Aurora's chatlogs has the potential to inspire innovative solutions for parenting chatbots.

1.6 THESIS OVERVIEW

This research is organised into seven chapters. Chapter 1 presents an introduction to the research topic and its context. Chapter 2 delves into the background and the technical architecture of Aurora. An overview of the literature on chatbots and digital parenting resources is explored in Chapter 3. Chapter 4 details the methods and techniques employed for data extraction and analysis. The outcomes drawn from the data analysis appear in Chapter 5. Chapter 6 involves discussions of the findings, and finally, Chapter 7 draws the conclusions from the research.

02

THE AURORA CHATBOT

2 THE AURORA CHATBOT

Aurora, a Facebook Messenger chatbot, was created by Cláudia Morgado, who drew inspiration from her background as a community pharmacist and willingness to help parents navigate the challenges of raising children [123]. Cláudia designed and developed the Aurora chatbot using common IT tools [38], as she was not an information technology specialist. The chatbot deployment occurred in October 2018, and it remained operational until September 2021, serving as a specialised resource for Portuguese-speaking parents. Its primary objective was to provide childcare insights, tips, and guidance to assist individuals on their parenting journey. The following section focuses on additional background information about the Aurora chatbot, including its inception, its founder's motivation, and its architecture's technical aspects.

2.1 AURORA MOTIVATION

Cláudia acknowledged that parenting is more challenging today due to the lack of traditional sources of knowledge and support, such as grandparents and siblings, neighbours with babies, and a community of parents [93, 123]. Recognising the limitations of healthcare professionals of not being available 24/7, Cláudia sought to develop a technological solution to assist parents with childcare questions. In response to these challenges and to help new parents, Cláudia developed the chatbot to provide essential parenting information and support.

Cláudia's 12 years of experience working as a sleep and breastfeeding consultant at a clinic, combined with her scientific knowledge, have allowed her to communicate parenting information clearly and comprehensively. She helped parents make informed decisions and embrace evidence-based practices by simplifying complex concepts. Her dedication and passion for supporting families have made her a trusted authority in the parenting community, benefiting countless families in Portugal [123].

Through the Aurora chatbot, Cláudia aspired to empower parents, providing them with the tools and insights necessary to enhance their parenting experience [123]. By providing valuable insights and guidance, Aurora aimed to assist parents in navigating the complexities of modern-day parenting [123]. The intent with Aurora was to be a reliable source of support, offering practical advice and information to help parents make informed choices and foster positive relationships with their children [123].

FIGURE 1 - AURORA DEVELOPMENT AND DEPLOYMENT TIMELINE.

2.2 AURORA TIMELINE AND FEATURES

Aurora's user base encompassed private individuals, corporate employees, and expatriates. The chatbot's reach extended to over 8000 users, most located in Portugal. Although an English version of the chatbot was available, the focus remained on serving users in Portugal. Cláudia engaged in ongoing updates and refinements to ensure the chatbot remained dynamic and responsive to user needs. These updates were informed by analysing user queries and system-generated error messages. The subsequent section details the key features of Aurora.

Chatbot Corpus

Aurora launched in October 2018 with a specialised focus on infant sleep issues and a month later expanded its knowledge base to include breastfeeding topics. Aurora's corpus comprised approximately 2,095 unique responses, covering a range of sub-topics within the main topics. For instance, under infant sleep issues, Aurora provided guidance on bedtime routines, sleep training methods, and tips on how to reduce sleep disturbances. Aurora offered information on breastfeeding positions, latch techniques, and managing common breastfeeding challenges in the breastfeeding domain.

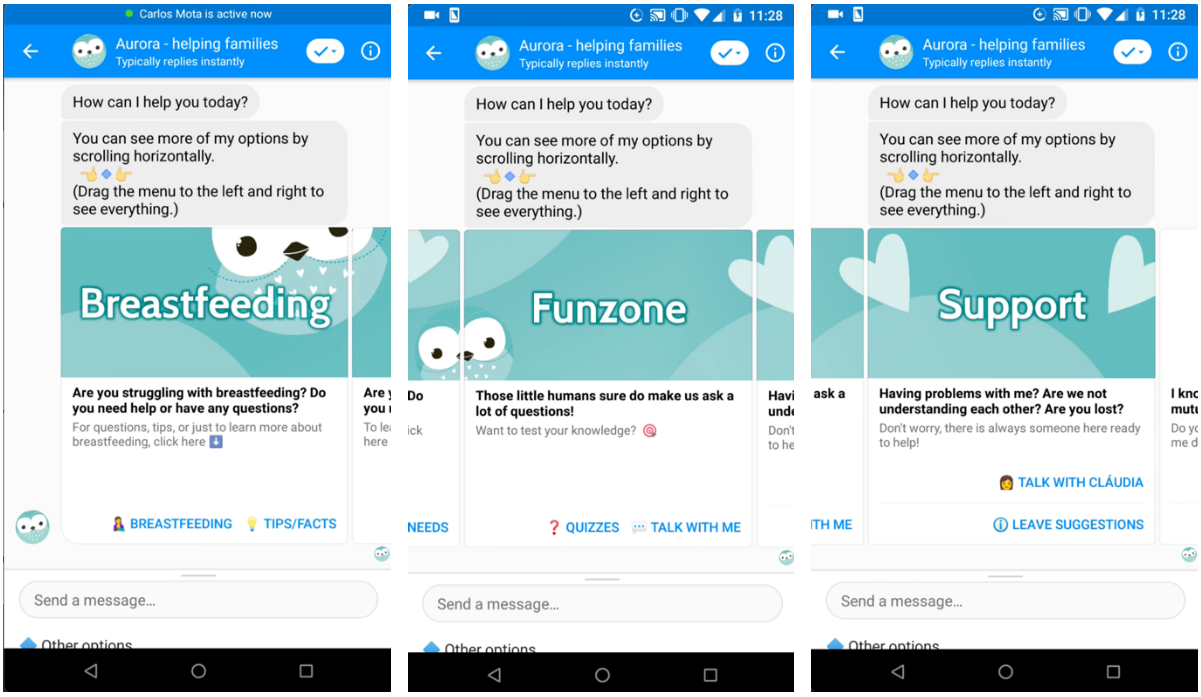

FIGURE 2 - SCREENSHOTS OF AURORA'S FEATURES.

Alongside this domain specialisation and expansion, Aurora also implemented a dual-modal interaction mode, where users could use a menu-based approach and explore different parts of the content, or explicitly write their question in the chat to receive a personalised response[1].

Daily Tips

Aurora provided its users with daily tips aimed at improving infant sleep patterns; this was only accessible through the paid subscription model. Utilising the chatbot's capability to record the timestamp of the last user interaction, Aurora employed an automated system to send the sleep-improvement tips daily for the subsequent 30 days.

Educational Game

Aurora introduced an educational feature named "Funzone" (Funzone: Figure 2), devised as a multiple-choice quiz. The quiz game aimed to debunk common misconceptions about childcare and impart knowledge on a range of topics, including sleep, pregnancy, breastfeeding, and general feeding practices.

1 The project was limited to only the JSON chatlogs, and there was no indication of the menu options or structure available to users within the provided dataset.

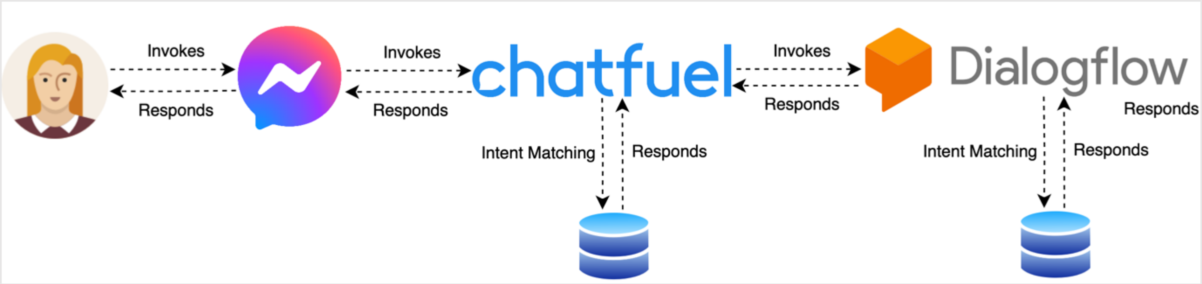

FIGURE 3 – AURORA TECHNICAL ARCHITECTURE, SHOWING THE FLOW OF MESSAGES.

Subscription Models and Personalised Consultation

Aurora offered a paid subscription model that enabled users to access an additional feature—direct communication with its founder, Cláudia. Subscribers could initiate personalised consultations with Cláudia through a typed command ("Cláudia") or by selecting the option from the menu ("Talk with Cláudia" - Figure 2). Simultaneously, Aurora extended its services by introducing corporate subscription models, enabling companies to offer the chatbot's services to their employees, particularly those who were either prospective parents or had young children.

The termination of Aurora in September 2021 was attributed to changes in Facebook's Messenger API (Application Programming Interface)[2]. These alterations encompassed modifications in user authentication processes, adjustments in rate limiting, security protocols, data policy compliance and the deprecation of certain key features integral to Aurora's functionality [37, 139].

2.3 TECHNICAL ARCHITECTURE OF AURORA

Aurora operated as a task-oriented, rule-based chatbot, focusing on discerning keywords relevant to parenting queries [135]. It functioned on a set of predefined rules and patterns tailored to provide information on specific parenting topics. It matched keywords in user messages against its predefined database of parenting topics and provided appropriate responses based on the matches found. The responses used by Aurora were a combination of templated and canned responses. The following provides an overview of Aurora's technical architecture, as illustrated in Figure 3:

1. **Facebook Messenger:** The Facebook Messenger Platform acted as the user interface for interactions between the chatbot and users [38]. Its APIs enabled developers to receive incoming messages and send responses to users [38].

2. **Chatfuel Data Handling:** Chatfuel handled all predefined user responses and menu options, matching them against a predefined database of queries and corresponding responses, functioning as a decision tree. Chatfuel then executed the corresponding response or action.
3. **Dialogflow:** For more complex queries, such as user-typed messages, that require additional processing or that fall outside the scope of the predefined Chatfuel workflows, Chatfuel handed the query over to Dialogflow. In cases where Dialogflow was invoked, usually due to user-typed queries, it defined intents represented in user queries and used these to formulate appropriate responses based on its existing database of responses [31]. Dialogflow would then transmit the response back to Chatfuel.
4. **Response Transmission:** The generated response, sourced from Chatfuel's database or formulated by Dialogflow, gets transmitted back to Facebook Messenger.
5. **User Received Response:** The generated response is displayed to the user on the Facebook Messenger platform, completing the interaction cycle.

This architecture effectively combined Chatfuel's role in managing predefined responses with Dialogflow's capacity to handle a broader range of user-typed interactions. Rule-based chatbots, like Aurora, are most effective in narrow domains and with well-defined queries. However, unlike machine learning-based and natural language-processing chatbots, they have limitations and are not equipped to handle more complex and open-ended questions [135].

---

2 An API defines a set of rules and protocols that govern the construction and interaction with software applications [39] Meta for Developers. 2023. Facebook Pages API. Retrieved September 4, 2023 from https://developers.facebook.com/docs/pages/#.

03

LITERATURE REVIEW

CHATBOTS & PARENTING

3 LITERATURE REVIEW

Chatbots have become increasingly prevalent in our daily lives, revolutionising how we interact with technology and providing valuable assistance in various fields. These virtual conversational agents simulate human-like conversations and assist users in solving problems, providing information, or offering emotional support [29]. Parenting chatbots aim to provide support and guidance to parents by offering personalised advice, answering questions, and sharing relevant resources. However, evaluating the effectiveness of these chatbots is crucial to ensure their quality and usefulness. This section explores chatbots' history, taxonomies, evaluation methods, and user acceptance. Thereafter, a review of parenting in the digital age regarding consumption patterns, themes, case studies, challenges, limitations, and ethical considerations of parenting chatbots.

3.1 HISTORY AND EVOLUTION OF CHATBOTS

The history of chatbots can be traced back to 1950 when a computer scientist called Alan Turing wanted to know, “Can machines think?” [127]. To answer this, Turing challenged computer scientists to create a computer program that would mimic human conversation to the point that it would be indistinguishable from talking to a person [75]. This challenge gave rise to the Turing test, originally known as the imitation game. The Turing Test operated on the principle that if, during text-based interaction, a group of human evaluators cannot discern between responses from a human and those from a computer, and they mistakenly attribute the computer's responses to that of a human, the computer is then deemed to have passed the test [127]. The fundamental concept of the Turing test served as a catalyst for the proliferation and development of chatbots.

In 1966, Joseph Weizenbaum developed a chatbot called ELIZA, designed to interact with psychotherapy patients by responding in natural language [136]. ELIZA responded to users by parsing users' words with scripted responses [136]. Although a simplistic program, ELIZA was successful as users confided to ELIZA and divulged personal stories and secrets [96]. Another noteworthy early chatbot was ALICE (Artificial Linguistic Internet Computer Entity), launched in the late 1990s, which employed a pattern-matching technique to mimic a human conversation [133]. ALICE was a precursor for many modern chatbots and won multiple Loebner Prize awards in the early years due to its impressive conversational abilities [27].

The advancement of chatbot technology has seen notable milestones, transitioning from the foundational stages with ELIZA and ALICE to today's sophisticated capacities. During the 1980s and 1990s, significant advancements in natural language processing (NLP) enabled chatbots to comprehend and generate human language more effectively [47, 74]. Subsequently, in the 2000s, chatbots embraced machine learning (ML) algorithms, allowing them to learn and adapt to user interactions [47, 106].

The 2010s brought about a revolutionary era with the start of deep learning. Chatbots equipped with these advancements gained the ability to process vast datasets, interpret complex user inputs, and generate coherent and contextually relevant outputs [55]. Such advancement in chatbot capabilities became feasible by integrating neural network architectures, specifically recurrent neural networks (RNNs) and transformers [47, 55].

The integration of chatbots in messaging applications has gained significant traction in recent years. A notable milestone was when Facebook (now Meta) launched its chatbot Messenger platform in 2016. In the end of that year, the platform hosted 34,000 chatbots, and now it hosts over 1,000,000 active chatbots covering a wide array of use cases [85].

Recently, transformer models, like OpenAI's GPT (Generative Pre-trained Transformer), have achieved notable progress in NLP [103, 106]. These models possess the capacity to process and generate text resembling human language, achieved by capturing the intricate contextual relationships present within an extensive training data [106]. OpenAI ChatGPT-3, launched in 2022, showcased its proficiency in generating coherent and contextually fitting text across various applications [103]. These breakthroughs have significantly elevated chatbot technology, allowing them to approach human-like conversational abilities and find applications across diverse domains. Despite their ability to generate text that appears human-like, these models often rely on statistical patterns rather than genuine comprehension [14, 110]. This can result in responses that, although coherent, may not always be appropriate or contextually relevant [14, 110].

However, contemporary deep learning approaches, despite their advancements, come with limitations. Studies have shown that most metrics currently used to evaluate AI models, including those used for chatbots, may not adequately reflect a model's performance. This inadequacy is particularly evident in NLP-specific tasks due to language and task-specific complexities [16]. Existing automated metrics, such as BLEU and ROUGE [86], often do not correlate well with human judgments, especially for tasks requiring nuanced understanding, like question generation and free-form answer generation [27, 97]. These metrics focus on n-gram similarity, which may not capture the full extent of a model's performance or its ability to generate contextually appropriate responses [27, 97].

3.2 Adoption of Chatbots in Different Sectors

Chatbots are progressively being implemented across various sectors and for diverse contexts. In customer service, a chatbot is an effective tool for communicating with customers to answer frequently asked questions, resolve common issues, improve customer service, and also help reduce the need and cost of human resources [2, 22, 43, 100, 112].

In education, chatbots have emerged as valuable digital assistants. They play multifaceted roles in educational environments, from facilitating administrative chores to addressing frequent student inquiries [52, 111]. Some advanced chatbots may also assist with personalised learning experiences, tailoring content or resources to individual student needs [111, 137].

The healthcare IT industry has adopted chatbots for assisting in scheduling appointments, disseminating general healthcare information, and conducting initial symptom screening [1, 8, 40, 52, 89, 95, 129]. These technologies are a valuable asset to healthcare owing to their round-the-clock availability, ensuring users can seek answers to their healthcare-related queries anytime [8, 52, 89].

Notwithstanding the rapid increase in healthcare-related chatbot solutions, there is limited literature on implementing chatbots for parenting and childcare [129]. This oversight is significant given that advances in artificial intelligence could make chatbots a viable tool for delivering personalised parenting support [140]. However, chatbots still require improvements, particularly in handling complex user queries [129].

Despite the numerous benefits chatbots offer, it is crucial to balance these against potential problems. For instance, technical issues, such as system overloads or backend processing errors, can also disrupt service [118, 125]. Moreover, there are concerns about privacy and data security, as chatbots handle sensitive user information [46, 57, 91]. Chatbots may misinterpret user queries, leading to inappropriate responses [32, 128]. An example highlighting the risks is the lawsuit against Air Canada's chatbot for failing to address customer complaints effectively, underscoring the need for robust design and comprehensive error management systems [24].

As chatbots find widespread acceptance across multiple sectors, it is important to understand their varied classifications. The multifaceted nature of chatbot applications necessitates a deeper understanding of their taxonomy, as it becomes evident that a one-size-fits-all approach is not applicable.

3.3 CHATBOT TAXONOMY AND TYPES

The landscape of chatbots is complex, with overlapping terminology instead of univocal classifications. Chatbots do not exclusively belong to a singular category; individual chatbots may belong to multiple categories, exhibiting varying functionalities and features. An example is the proposed taxonomy of chatbots by Grudin and Jacques [2019], where they identified chatbots into three categories: a) intelligent assistants with short conversations on a wide range of topics; b) virtual companions that discuss various topics but can carry on with a conversation; and c) task-focused chatbots that aim for brief discussions of concise issues [60].

Følstad *et al.* [2019] suggested a two-dimensional typology that equated to four types of chatbots: locus of control dealing with how a user interacts with a chatbot (chatbot-driven or user-driven) and based on the duration of the conversation (short- or long-term conversations) [50]. A more simplified and commonly used classification is that of Hussain *et al.* [2019], where chatbots are task- or non-task-oriented. The subsequent sections examine some of the different classifications of chatbots and their functionalities.

3.3.1 TASK-ORIENTED AND NON-TASK-ORIENTED CHATBOTS

Chatbots not only vary in their interaction medium but also in their functional objectives. While some are engineered for voice-based or text-based dialogues, leveraging computational linguistics and artificial intelligence, they can be categorised into task-oriented and non-task-oriented chatbots, each with distinct goals and computational algorithms [4, 29].

Task-oriented chatbots have a clear objective: to assist users in accomplishing domain-specific tasks or goals, such as booking accommodations [68]. Its primary focus is to efficiently address users' needs and queries within a well-defined scope. Task-oriented chatbots lack extensive knowledge; instead, they assist users in completing a particular task [82]. The dialogue context of this type of chatbot can take the form of text or voice conversations, thus using different algorithms. While these chatbots often employ pattern matching to understand the context of queries [82], modern task-oriented chatbots incorporate advanced NLP techniques to provide more accurate responses [115]. Text-based chatbot can also be hosted on applications such as Facebook Messenger, WhatsApp, and Telegram [142]. An example of a voice-type chatbot is Amazon’s Alexa, a task-oriented chatbot that aims to provide accurate responses or execute tasks based on users' requests [9].

Conversely, non-task-oriented chatbots aim to simulate interactions in a more open and unstructured conversational style [68]. Developers often build these chatbots to simulate human speech and interaction, serving roles such as companionship, entertainment, or therapeutic support [140]. An example of a non-task-oriented chatbot is Replika, originally conceptualised as a personal AI companion that adapts through user interactions [116]. Replika’s primary function was to facilitate open-ended dialogues with users, focusing on providing emotional support and companionship [116].

While all chatbots qualify as interactive software, many are not design specifically for conversational purposes. Many chatbots merely serve as mechanisms to crawl web pages and alert people rather than engage in meaningful conversations with users [60].

3.3.2 Knowledge Base and Input-Response Mechanisms

This section will explore another axis of chatbot classification relating to a chatbot's knowledge base and input-response mechanisms. Regarding the knowledge base, the classification of chatbots falls into either open-domain or closed-domain categories.

Open-domain chatbots can handle a broad range of topics and facilitate more open-ended conversations [3, 119]. An example of this type of chatbot is Google Assistant, an open-domain chatbot designed to answer questions on various topics [56]. Whereas closed-domain or domain-specific chatbots possess knowledge about a single area of interest and provide solutions for specific, limited scenarios [3, 119]. An example of a closed-domain chatbot is Aurora, which possesses knowledge limited to baby sleep and breastfeeding topics .

The knowledge base classification is closely related to the methods used for processing user input and generating responses. In chatbot development, developers generally employ three principal models for processing user input and generating responses: rule-based, generative-based, and retrieval-based models [3]. The rule-based or scripted model operates on a set of fixed, predefined rules, often structured as a dialogue tree [3]. This type of chatbot utilises regular expressions to match user input with a corresponding set of predetermined human-like responses [3].

Conversely, the generative-based model leverages machine learning and deep learning techniques to facilitate more dynamic interactions [68, 119]. These chatbots generate better responses based on extensive historical data and previous conversations [119].

Retrieval-based chatbot responses are retrieved from a database or set of pre-programmed responses, guided by the user's input [68]. The algorithms for these chatbots can vary in complexity, ranging from straightforward rule-based matching systems to more sophisticated processes that integrate machine learning classifiers [65, 133]. Retrieval-based chatbots are similar to rule-based models; however, they possess the added capability to self-learn and enhance their response selection over time [68].

3.4 CHATBOT TRUST AND USER ACCEPTANCE

Understanding users' perceptions and factors influencing the acceptance of chatbots is essential for improving user experience and optimising the design and functionality of chatbots. This section reviews the existing literature on the diverse elements influencing user readiness to adopt chatbot technology.

The acceptance of a chatbot depends on the expected human effort of using the chatbot, social influences, facilitating conditions, habits, compatibility, and ability to access the system [89]. However, the user's motivation for using a specific technological artefact directly influences the user's acceptance and adoption of a chatbot [114]. Nadarzynski *et al.* [2019] further add that participants' acceptance of chatbots can also stem from a curiosity about new technologies.

Perceived trustworthiness is another factor influencing user acceptance of chatbots. Go and Sundar [2019] noted that various factors, such as a chatbot's anthropomorphic attributes—including its personality, visual characteristics, and ability to handle queries—as well as whether the chatbot's name resembles a human name, dictate a user's level of trust. Findings from numerous chatbot parenting studies corroborate this idea, revealing that users often exhibit lower acceptance levels when they perceive chatbots as lacking the empathy and professionalism typically associated with humans [44, 95, 101]. Additionally, several research studies, including a survey of five healthcare chatbots, concluded that some users remained sceptical about chatbots due to concerns regarding privacy, data security, perceivability, information accuracy, and the ability to address complex queries [40, 95, 101].

Despite various studies indicating a lack of trust in chatbots, some literature revealed that chatbots go beyond being knowledge facilitators. Chatbots also alleviate users` loneliness, particularly for parents of newborn infants [140]. An example is a chatbot developed for mothers in impoverished conditions in India. Its acceptance among users was notable, as they appreciated the convenience of communicating with the chatbot at any time and the non-judgemental interaction, particularly for those reluctant to talk with a doctor [140]. These users felt they could confide to the chatbot about issues that may have been too embarrassing to discuss with others, thus indicating that the chatbot was successful in evoking trustworthy characteristics.

The consensus is that the quality of chatbot responses, empathy and perceived accuracy of the chatbot's information relates to the chatbot's acceptance [89]. Thus, a chatbot's success metrics hinge on high-quality conversations that yield a positive user experience [132]. In addition, privacy factors and substituting face-to-face interaction may also influence users' willingness to use a chatbot [114].

3.5 Chatbot Evaluation

Building upon the characteristics influencing chatbot acceptance, this section will focus on the methodologies commonly employed for chatbot evaluation. While various methodologies are commonly utilised for evaluating chatbots, such as user interviews or self-administered questionnaires, it is important to acknowledge that no approach is entirely devoid of bias [131]. Even techniques that offer objective insights, such as technical metrics and sentiment analysis, have limitations. These methodologies may reduce the influence of users' retrospective views but are not immune to biases embedded in algorithms or the selective focus of specific metrics. Accordingly, there is a scholarly consensus that quantitative indicators alone are insufficient to capture the richness of user experience [5, 127].

A holistic approach encapsulating quantitative and qualitative methodologies is important in studying chatbots. Researchers have extensively employed user-centred studies, focusing on questionnaires and interviews to gain insights into user satisfaction with chatbots [131]. These methodologies, although instrumental, have their caveats. For instance, Agarwal and Wadhwa [2020] argue that evaluations based on chatbot-human interactions are mainly subjective and interwoven with personal judgments, underscoring the need for evaluations to happen in real-time [71]. Echoing similar sentiments is Venkatesh *et al.* [2018], which emphasises that while user experience is a pivotal criterion for chatbot evaluation, an overreliance on user perceptions can overlook objective dialogue patterns [71].

There may be potential pitfalls in relying solely on subjective retrospections, as they might not always accurately represent user experiences [17]. The potential inaccuracy underscores the importance of including more objective measures. Frequently, chatbot analysis utilises a blend of qualitative and quantitative methodologies [5, 127].

3.5.1 Qualitative Chatbot Evaluation Analysis

In chatbot research, access to comprehensive data, including user feedback and behavioural metrics, is often ideal but not always feasible or available. In such situations, the chatlogs of interactions between users and the chatbot are valuable resources for evaluation. In the absence of or to supplement user interviews and questionnaires, methods such as content, thematic and conversational analysis provide valuable insights into chatbot conversations.

Content analysis: This traditional form is a systematic and objective method for describing verbal, written, or visual communication [66]. Researchers use this approach to gauge the core functionalities of chatbots, the range of their responses, and their reliability in disseminating information [25, 66, 141]. The goal is often to understand the content's contextual meaning and interpret it in a way that informs the research questions [66]. Hsieh and Shannon [2005] delineated three methodologies for qualitative content analysis: summative, directed and conventional. These approaches offer researchers a framework for methodological selection based on research questions, goals, and theoretical frameworks.

Thematic analysis: A thematic analysis involves identifying, analysing, and interpreting themes within the qualitative data [21]. This method provides deep insights into chatbot interactions by extracting patterns of meaning, thus uncovering the underlying themes and notions that characterise the topics discussed [21].

Conversational Analysis: By studying the structural flow of conversations, researchers can ascertain where users tend to disengage or encounter bottlenecks, offering a structural view of the interaction [102].

Qualitative techniques offer invaluable frameworks for unpacking the complexities of conversations. They help to discern patterns, understand contextual cues, and identify areas of misunderstanding or system failure. These methods become especially pertinent when other forms of user input, such as interviews and questionnaires, are absent or limited.

3.5.2 QUANTITATIVE CHATBOT EVALUATION ANALYSIS

While qualitative analysis illuminates the nuanced interactions between chatbot and user, quantitative methods provide ways to measure, test and compare phenomena. The Turing test has long been a method used, but Khanna *et al.* [2015] state that it only gauges a machines conversational abilities and not its comprehensive attributes. Venkatesh *et al.* [2018], Verhagen *et al.* [2014] and Abd-Alrazaq *et al.* [2020] suggest employing metrics that shed light on conversation depth, user engagement, and consistency. The following are some of the quantitative methods used for chatlog analysis:

Statistical Patterns: This is a statistical approach to scrutinise user engagement patterns, response durations, accuracy metrics, frequency counts and instances of errors [1, 17]. Such metrics are instrumental in gauging a chatbot's operational efficiency and overall performance.

Usage Trends: By analysing the frequency, duration, and patterns of chatbot use, researchers can identify user preferences, peak usage times, and the popularity of certain features or topics [17, 132].

Sentiment analysis: This assessment tool enables a comprehensive understanding of user emotions and perceptions reflected in the text of messages [43]. This approach incorporates both qualitative insights into users' feelings and contexts, and quantitative metrics, classifying sentiments into measurable categories [43, 69, 72, 119, 124]. The result identifies chatbot strengths and weaknesses, guiding refinements to improve user satisfaction [43, 124]. By integrating sentiment-based evaluations with other methods, a richer and more holistic perspective on chatbot performance emerges, striking a harmonious balance between objective technical evaluations and the depth of subjective user experiences [43, 130].

In the absence of a standardised methodology for chatbot research and evaluation, researchers often combine various methods to provide an in-depth and holistic assessment of chatbots [23].

3.6 PARENTING IN THE DIGITAL AGE

The digital landscape today presents parents with a blend of advantages and challenges. Within this domain, social media parenting groups, childcare blogs and parenting influencers provide advice and serve as platforms for exchanging experiences [18, 58]. Parenting chatbots, along with specialised websites like BabyCenter [11] and BabySparks [12], offer guidance and support to parents as they journey through the intricate landscape of childcare. Such applications offer a range of tools and platforms, such as specialised parenting forums, mobile apps equipped with advice and developmental trackers, in-depth online parenting courses, educational activities, and informative e-books [11-13, 61].

The accessibility of information means parents are never truly alone in their parenting journey [81, 140]. Support, advice, or a sympathetic ear is just a click away at any hour. The internet also offers diverse perspectives, allowing parents to gain insights from global communities, broadening their horizons and tools [18]. With advancements in AI and ML, parents have access to instantaneous, personalised learning and support tools that can adapt to individual needs [68].

However, the vastness of the internet leads to information overload, making it difficult for parents to distinguish between accurate and relevant advice [35, 80]. Additionally, the curated lives portrayed on platforms such as Instagram can set unrealistic standards, leading parents into a cycle of comparison and social pressure [34, 81]. Moreover, the digital/information age brings with it privacy concerns [114]. While sharing a child's first steps or birthday online can seem innocent, it raises questions about consent and potential future implications for the child [10, 64].

3.7 CONSUMPTION PATTERNS OF DIGITAL RESOURCES FOR PARENTING

There has been a shift in how parents access, consume and interact with childcare-related information. Today's digital landscape is rich with various tools and platforms to support parents. In light of this expansive array of digital tools and platforms, it is interesting to examine how parents engage with and consume these resources.

Parents often consult digital platforms for childcare, particularly pertaining to healthcare and education [13, 61, 78, 105]. These resources have seamlessly integrated into their daily activities, from perusing parenting blogs, social media and Google to weekly catch-ups with parenting vloggers [13, 78, 105]. A study by Baker *et al.* [2017] revealed that parents, mainly non-working mothers, dedicated more than 10 hours a week to searching for parenting information.

As for the mode of access, the trend leans heavily towards mobile-first consumption and desktop next [105]. Additionally, users consume information through active and passive approaches, leading to variations in consumption patterns. Some parents adopt a proactive approach, actively searching for specific information or solutions, while others might stumble upon information passively, often driven by algorithmic feeds or recommendations from other users [61]. Research also indicates that mothers use online resources more frequently than fathers [13]. Given the varied consumption behaviours of parents, it is pertinent to explore what drives these interactions with digital tools.

Several factors drive the way parents interact with digital resources. A child's developmental stage heavily influences the nature of online queries parents might have [41, 78]. However, as parents acquire experience or the child's developmental needs evolve, they seek more specific and targeted information [13, 61]. For example, the early days of parenthood with new-born's could lead parents to seek advice on feeding or sleeping habits [41, 78, 129]. Parents also turn to digital resources when faced with new challenges, such as an illness [61]. Often, parents gravitate toward online parenting support groups during significant developmental milestones, such as pregnancy or the child's infancy [13, 53]. As parents navigate through the various stages of parenting, their initial interests might shift, giving precedence to newer topics [13, 53]. As the child transitions into adolescence, the concerns might shift towards ensuring digital safety or comprehending and addressing the nuances of peer pressure [41, 78].

The digital consumption habits of parents are not static, and one significant influencer is the age and technological acumen of the parent [41]. Literature indicates that younger parents seem more engaged with diverse online resources, while older parents might be more prudent, often gravitating towards familiar and trusted platforms [41]. While digital natives effortlessly navigate through various digital platforms, those who are less tech-savvy might stick to a few trusted sources or rely on peer recommendations [42, 78].

Parents' cultural backdrop and economic standing can also affect their digital consumption [13, 140]. From an economic perspective, the literature indicates that parents in more affluent brackets demonstrate a higher tendency for information-seeking activities than those with limited financial resources [41, 78, 140].

Between the vast scope of digital resources and the dynamics of parental consumption, trustworthiness and determining reliability are important factors to examine. Trust becomes a central concern in the extensive digital landscape, especially for parents seeking reliable guidance. Peer-driven platforms, including forums, seminars, and social media groups, have become valuable reservoirs of trusted information [90, 93]. When fellow parents share or endorse resources, it carries considerable weight and can influence which resources are explored and trusted [13, 61, 78]. Parents trust professional resources and resources endorsed by paediatricians, medical experts, or renowned parenting experts [13, 61, 78]. Regarding the perceived reliability of information, parents value more seminars tailored to them and individualised programs [13].

3.8 Chatbots for Parenting Support

Chatbots have gained prominence as parenting support tools, reflecting technology's deepening integration into our daily routines [29]. One of the most intriguing intersections of technology and parenting is the rise of chatbots designed to offer parents support, advice, and resources. This section explores the functionality and examples of chatbots and the role, benefits, and potential challenges of chatbots in parenting support.

3.8.1 Functionality and Scope

Within the parenting domain, chatbots demonstrate a variety of complexities and functionalities, encompassing informative (task-based) and interactive (non-task-based) chatbots.

Informative chatbots offer immediate answers to prevalent parenting questions, from feeding schedules to developmental milestones and healthcare concerns [3]. These chatbots derive their responses from a fixed database and often exhibit limited interactivity, functioning mainly as a one-way communication channel from the chatbot to the user [3]. Such chatbots guide parents to pertinent articles, videos, or other media in response to their questions [121]. Given their deterministic behaviour, informational chatbots follow predetermined conversational pathways and lack the capacity for complex dialogue [3].

On the other hand, **interactive/conversational chatbots** simulate dialogues to assist parents in addressing challenges related to behavioural issues and parenting decisions and provide emotional support [45]. These chatbots usually incorporate machine learning algorithms to enhance conversational capabilities and improve responses over time [45].

While the functionalities of chatbots can vary greatly, understanding their specific applications in parenting can shed light on their potential impact, especially for parents who may benefit most from their services.

3.8.2 PARENTING CHATBOTS

Chatbots have been predominantly used for commercial purposes as task-oriented chatbots. There is an opportunity for chatbots to have a significant positive social impact, particularly for parents living with financial constraints [140]. Parenting chatbots cover various parenting-related topics, including child development, nutrition, general healthcare, sleep habits, behavioural problems, and educational activities, and provide emotional support for parents [29, 129, 140]. Although chatbots have expanded their coverage of childcare topics, they still exhibit limitations regarding knowledge depth and their ability to handle complex queries [129]. Current literature reveals a limited scope in the content coverage of parenting chatbots, with a predominant focus on breastfeeding and prenatal care [129, 140]. The following are some examples of parenting chatbots:

Yuki: This non-task-based chatbot with the objective to educate mothers and Accredited Social Health Activists (ASHA) in underserved Indian communities [140]. Its focus was exclusively on breastfeeding, inherently restricted its long-term utility. During the course of user evaluations, there were instances where mothers sought information that extended beyond the chatbot's predefined knowledge base, encompassing broader topics relates to personal and child healthcare [140]. The chatbot exhibited limitations in parsing local colloquialisms and typing errors [140]. Despite these constraints, the user reception remained predominantly positive as users felt comfortable confining with the chatbot without fear of judgment [140].

MamaBot: Created to provide assistance and guidance to mothers with young children and pregnant women [129]. The objective of MamaBot was to present its users with two macro-category services: prompt and informative recommendations for emergencies and response to maternal scenarios, encompassing lifestyle and preventive healthcare advice [129]. MamaBot was designed to address a spectrum of topics and support prolonged engagement [129].

Entenberg *et al.* [2021] Chatbot: The purpose of this chatbot was to provide education to Argentinian parents regarding the management of reported that the chatbot was effective in delivering useful information and imparting valuable skills [44]. Nevertheless, users expressed that the chatbot needed additional parent training skills to address more complex issues and to refine its communication style [44].

Wong *et al.* [2021] Chatbot: In contrast to previous examples, this chatbot did not offer parental guidance, it functioned as a confidante for new mothers, positively impacting their mental well-being [138]. The chatbot collected conversational data from parents of babies below six months of age to compare parents' experiences with pre-term and full-term babies in aspects like sleep, stress, and infant feeding [138].

The existing literature suggests a growing trend of chatbot tools becoming increasingly accessible, particularly to individuals with the necessary resources. While some are narrow in scope, focusing on specific parenting and childcare topics, others encompass a broader range of topics, especially if designed for long-term use. Even with certain limitations, these chatbot explorations contribute to refining the user experience, broadening content, and enhancing user query responses.

3.8.3 BENEFITS AND CRITICISMS OF CHATBOTS IN PARENTING SUPPORT

Chatbots have become a tool for parents, offering numerous advantages to address their unique needs. Parents frequently encounter challenges at unpredictable hours, particularly with newborns [52, 138]. Chatbots can step in to offer uninterrupted, 24/7 support, eliminating the conventional wait times and unavailability associated with human interactions [52].

Moreover, chatbots ensure parents receive consistent and reliable information based on their precise programming and reference databases [123]. A significant advantage is the provision of anonymity; chatbots create a non-judgmental space, allowing parents to ask sensitive or private questions without apprehension [140]. Additionally, with technological advancements, many chatbots now possess the capability to personalise responses [5]. Some chatbots have the ability to adapt to individual user behaviours and preferences, offering tailored advice that considers previous interactions [119].

Despite the advantages of chatbots in assisting parents, there are some inherent challenges and concerns associated with their use. Chatbots, although informative, lack the human touch; they lack the empathy, understanding, and delicate feedback that one would expect from human interaction [44, 54, 95, 101, 114, 140]. In addition, as chatbots lack the capacity to contextualise users' emotions, they may also fail to detect emergency situations [77].

Furthermore, the chatbot's information expertise and reliability come into question if it is not updated [54, 100], as distributing outdated or misinformation is particularly worrisome in matters concerning childcare. Another concern is the potential over-reliance on chatbots, as some parents may place excessive trust in chatbot recommendations, potentially overlooking the invaluable insights of professionals [77].

The blend of chatbots with parenting support represents a transformative shift towards more accessible and instant help for parents. Hiniker *et al.* [2016] highlight that the balance between technology and the human element remains pivotal as digital tools become further ingrained in the parenting paradigm. As technologies evolve and adapt, it is essential to acknowledge and address ethical concerns associated with parenting chatbots.

3.9 Ethical and Privacy Considerations of Parenting Chatbots

As societal norms change, so do the concerns of parents. The prevalence of digital technology and social media has brought misinformation and safety concerns [83]. The ethical ramifications of deploying chatbots in sensitive domains, such as parenting, can be profound [140]. A study by Vaidyam *et al.* [2019] and Coghlan *et al.* [2023] raised notable concerns about mental healthcare chatbots, highlighting limitations and occasional inappropriateness in the generated responses. These observations point to broader issues such as the risk of misinformation, the possibility of an over-reliance on technology, and the transition from human interactions to machine-guided communications are pressing concerns under this topic [35, 80]. According to Floridi and Cowls [2019], when designing chatbots in sensitive domains, such as parenting or mental healthcare, they should include ethical principles of respect for autonomy, explicability, non-maleficence, justice, and beneficence.

In addition, it is also important to protect user privacy, especially since chatlogs may contain sensitive information [46, 57, 91]. Data storage, transfer, and analysis must comply with privacy standards, such as the General Data Protection Regulation (GDPR) in Europe [46] and the POPIA (Protection of Personal Information Act) in South Africa [57].

To ensure ethical data handling practices during the analysis of chatbot interactions, obtaining informed consent is crucial [46, 57, 59, 92, 129]. Users must be informed about the research's purpose, their participatory role, and their rights [46, 57, 91]. Of equal importance is de-identification to ensure user anonymity. All personal identifiers, such as names or addresses, should be removed or anonymised from chatlogs before the analysis process [46, 57].

04

METHODS

DATA, TOOLS & METRICS

4 METHODS

Following the literature review on chatbot applications and their ramifications in parenting, this chapter describes the methods employed for the analysis of the Aurora chatbot. The analysis revolved around three main objectives: defining the nature of the conversations, assessing their quality, and identifying common and unaddressed topics. To achieve these research goals, various methods and techniques, including NLP, were employed to extract key metrics such as user engagement and sentiment, conversation dynamics, chatbot satisfaction, and performance. This chapter delineates the data and methods for processing and evaluating the chatlogs.

4.1 DATASET DESCRIPTION AND PRIVACY

The provided data consisted of already pseudonymised chatlogs presented in the form of JSON files[3]. It facilitated easy data usage, reading, parsing, and generation, ensuring efficient handling of the chatlogs throughout the analysis process [36]. A total of 1043 files were available, representing 1043 distinct user conversations with the chatbot. To protect the user's identity and ensure anonymity, the name field within the JSON file was substituted with the pseudonym "P[number of participant]". This change ensured the privacy and confidentiality of the individual's actual name. To further protect user identity, any mentions of the user's first name within the JSON file was replaced with "P[number of participant]". This adjustment accounted for automatic mentions generated by the chatbot and covered any instances where Cláudia mentioned the user's name in a conversation.

In certain instances, users mentioned a child's first name within the conversation. Although there was no identifiable connection or link [57, 83] to a specific individual, the researcher still employed manual anonymization of the child's name. This approach preserved the authenticity of the conversation while effectively addressing any potential privacy concerns associated with personal identification.

[3] JSON is a lightweight data-interchange format designed to be user-friendly for humans and machines [36] Douglas Crockford. 2012. JSON. Retrieved August 2, 2023 from https://www.rfc-editor.org/rfc/rfc7159.

The JSON files contained several attributes, with the more relevant attributes being the *sender_name*, *timestamp_ms*, and *content*. The *sender_name* attribute distinguishes between messages sent by the chatbot and those sent by the user. The *timestamp_ms* attribute provided the timestamp for each message, while the *content* attribute encompassed the text of the exchanged messages between the user and the chatbot. An important aspect to highlight is that users could either type their messages or choose them from a set of predefined menu options.

The content of the chatlogs was in Portuguese. The quotes appearing in this thesis were translated by the researcher from Portuguese to English.

Schema

The structure of these chat logs is methodically organised into two main components, namely participants and messages.

The "participants" section of the JSON file lists the entities involved in the conversation. Each participant is represented by a name attribute. The participants are the user and the Aurora chatbot.

The "messages" section contains an array of message objects, each representing a single message in the conversation. Each message object includes several attributes:

- The \`sender_name\` indicates who sent the message, either the user or the Aurora chatbot.
- The \`timestamp_ms\` records the time of the message in milliseconds since the Unix epoch.
- The \`content\` field holds the actual text content of the message.
- The \`type\` attribute categorises the message type, such as "Share" or "Generic".
- There are boolean flags like \`is_unsent\` and \`is_taken_down\` that indicate the message's status.
- The \`bumped_message_metadata\` attribute contains metadata related to the message if it has been "bumped," including the original message content and its status.
- If a message contains photos, it includes a \`photos\` array, where each photo object has a \`uri\` for the photo's location and a \`creation_timestamp\` for when the photo was created.

4.2 TOOLS

For this research, the primary tool chosen to organise and analyse the dataset was the Python programming language. Python is an interpreted, multi-paradigm programming language with built-in data types and structures to facilitate dynamic typing and binding [30, 109]. Python was used to parse the dataset and extract the necessary metrics for the research analysis.

Jupyter Notebooks [113] served as the platform for conducting the analysis. Jupyter Notebooks, often simply called 'notebooks', have the ability to interface with Python as it is a tool that operates in the browser and is freely available as open-source software. [113]. This platform supports research processes, including code, data, and visual representations, aiding in documenting the research findings. The utilisation of Jupyter Notebooks helped systematise the analysis and enhance the reproducibility of the research.

Using Python, the data from JSON files were processed and consolidated into a user-friendly CSV file format. This conversion involved parsing the content of each JSON file and extracting only the data from the relevant attributes.

Two key modules, namely Pandas and Matplotlib, were utilised extensively throughout the research. Pandas, a widely used Python library for data analysis, empowers Python to handle data in a tabular format [88]. To visualise the findings, Matplotlib was employed as it is often used for creating static, animated and interactive visualisations in Python [67].

The two primary data structures in Pandas are DataFrame and Series. The DataFrame is similar to a two-dimensional table comprising columns of uniform data type, whereas a Series is a one-dimensional indexed array of a specific data type [28]. DataFrames were the most suitable for analysing chatlogs, given their common use in analysing actual data. It was particularly effective for handling data in a tabular layout, such as in CSV format [88]. Utilising Pandas made loading, processing, and analysing the chatlogs simpler by using queries reminiscent of SQL (Structured Query Language).

4.3 DATA PRE-PROCESSING

To facilitate effective text analysis and ensure accurate results in subsequent text analysis tasks, the dataset underwent pre-processing to enhance readability and analysis. Text pre-processing denotes transforming unprocessed text data into a uniform and clean format [62]. There were limitations on solely relying on pre-processing techniques, as they would result in the loss of context and contextual meaning for some of the qualitative analyses. Consequently, the research used the raw dataset for some analysis and a pre-processed set. The following overview details the pre-processing methods applied:

- **Lowercasing** text data is a simple and effective technique often used in text mining and NLP [132]. When lowercasing is absent, words appearing in different cases are considered different entities [132]. Given that the dataset was not very large, lowercasing assisted with the consistency of the anticipated output. The text data was lowercased using the built-in string methods in Python.
- **Removal of special characters,** such as emojis and emoticons, avoided causing complications during sentiment analysis. The method employed Python's **re** (regular expression) [72] library to perform regular expression matching and remove special characters.
- **Numerical to textual conversion** process involved transforming numerical values into their corresponding textual representations.
- **Stop words** that have little semantic significance, such as “is” and “the”, were removed [72]. The Natural Language Toolkit (NLTK) [99] library, which includes a predefined list of Portuguese stop words [110, 111], carried out this task [15].
- **URLs and HTML tags** were excluded from the dataset using regular expressions with the help of the **re** library [72]. These regular expressions pinpointed and removed patterns associated with URLs and HTML tags, ensuring a cleaner analysis without interference from these elements.
- The dataset was subject to **lemmatisation**, an NLP technique that reduces words to their base or root form [107]. This process removes the word's inflexions and variations, such as tenses and plural forms, thus reverting the word to its canonical form. Lemmatisation enhances text analysis, such as for sentiment analysis [107]. The **Spacy** library in Python carried out lemmatisation using the Portuguese language model package **pt_core_news_sm** [87]. This language model, designed to process Portuguese text, encompasses features like part-of-speech tagging, tokenization, named entity recognition, and lemmatisation [87].

The pre-processing steps were essential for tasks such as sentiment analysis, topic modelling, and textual analysis. These techniques enabled the research to extract meaningful insights, identify emotional sentiments, uncover underlying topics, and categorise text effectively, contributing to a more robust and insightful analysis of the chatlogs.

4.3.1 USER ENGAGEMENT METRICS

The research employed several metrics to comprehend and categorise user interactions with the chatbot. Utilising various methods for chatlog analysis and following a methodology similar to that of Maroengsit *et al.* [2019], to extract the following metrics:

- **Total User Count** offered a numerical representation of unique users who interacted with the chatbot.
- The **Bounce Rate** measured the percentage of users who disengaged from the chat without a question posed or information gathered from the chatbot. The initial data analysis established that 321 (31%) users did not solicit childcare information from the chatbot. The messages these users transmitted appeared as a test of the chatbot's capabilities or mere playful interaction. Bounced users also included users who sent less than five messages to the chatbot. This research excluded these users because of their minimal engagement with the chatbot, and the level of interaction did not lead to users gaining any childcare tips.
- **Interaction Modality** served as an instrument for discerning the methods of interaction between the users and the system, identifying whether users opted for menu-prompted options or users who input their queries manually. The research classified messages with fewer than three words as menu selections and those with more words as typed responses because the JSON files did not explicitly differentiate between the two.
- The **Retention** metric enabled the evaluation of the frequency of user engagement with the chatbot over a certain period and the corresponding drop-off rate.
- An **ANOVA** (Analysis of Variance) test was conducted to determine if there was a significant difference in the number of messages exchanged with the chatbot across different days of the week [98].

4.3.2 DIALOGUE METRICS

Understanding the intricacies of dialogue between users and chatbots is essential for enhancing the interactive capabilities of these AI-driven platforms. Leveraging the multifaceted methods proposed by Jiang *et al.* [2022], the research extracted the following metrics:

- The **Chat Volume** provided a statistical representation of the number of messages users sent to the chatbot. Included in this were descriptive statistics like the median and standard deviation of these messages, attained using the **.describe()** Pandas method.
- **Session Count** measured the time intervals between consecutive messages from each user to identify unique sessions. It was established that an interval exceeding 30 minutes marked the start of a new session.
- **Tenure** is the number of days that lapsed from a user's first interaction with the chatbot to their last interaction. This metric served as an indicator of user retention and long-term engagement with the chatbot.
- **Unique days** refer to the number of days a user has sent at least one message to the chatbot. In this metric, multiple messages sent on the same day are counted as one "unique day," while messages sent on different days increase the "unique days" count.
- **Dialogue Paths** established the distribution of conversations based on conversation initiation. These metrics distinguished conversations initiated by users from those initiated by the chatbot. This metric aided in identifying the primary drivers of conversations.
- The **Interaction Timing** metric ascertained the specific times and days of the week when users sent messages to the chatbot. The information derived from this metric was used in plotting a user interaction distribution table, as further explained in section 4.4.1.
- **Error Frequency** captured instances where the chatbot failed to understand or respond to user queries. The research quantified this metric by counting the chatbot's error response messages.
- **Response Accuracy**: The methodology introduced by Procter *et al.* [2018] provided a distinct coding theme, which became the foundation for classifying chatbot responses during user interactions. This approach enabled the measurement of the frequency with which the chatbot accurately met user queries and informational requirements. The methodology excluded menu-based user selections, focusing solely on user typed questions. Subsequently, chatbot responses underwent the following classification:
    - *Partially addressed response* characterised situations where the chatbot provided a general or incomplete answer, only addressing part of the user's question.

- *Single-topic complete response* encapsulated the instances where the chatbot delivered more accurate and directly relevant replies, effectively addressing the user's queries.
- *Referral responses* denoted the instances when Cláudia, the healthcare professional, stepped in to provide personalised responses. This response occurred when users utilised a specific chatbot feature (typing "Cláudia") that enabled requests for her participation in the conversation.

4.3.3 TEXT MINING: TOPIC METRIC

Adopting the text mining analytical strategies proposed by Jiang *et al.* [2022], this research sought to uncover the popular and recurrent topics users queried with the chatbot. Topic metrics focused on analysing the topics most frequently discussed during user interactions with the chatbot. This process involved searching for words associated with existing and new topics. The following is an overview of the methodology used:

1. **Data Exploration:** This analysis listed all words from the pre-processed dataset with their respective frequencies.
2. **Topic Identification:** From the list of words, relevant terms were manually selected and then organised into three specific topic categories: Sleep, Breastfeeding, and Healthcare. This analytical approach also facilitated the discovery of emergent themes, which were subject to manual analysis for contextual interpretation.
3. **Word-Topic Mapping:** The subsequent step involved associating words with their respective topics. A Python dictionary served this purpose, aligning each word with its relevant topic.
4. **Topic Assignment:** The process included coding a function to allocate topics to each message within the DataFrame. This function split each message into distinct words, checked if each word was present in the word-topic dictionary, and allocated it to the corresponding topic list. This list was then deduplicated and sorted to provide a unique and orderly set of topics for each message.
5. **Topic Analysis:** The distribution of topics across all messages was analysed. The process counted the frequency of each topic and implemented an additional procedure for messages encompassing several topics. This analysis offered insights into the dataset's major and minor themes.
6. **Manual Verification:** A subset of messages containing intriguing or less common words underwent manual inspection. The contextual review served to cross-validate findings and to clarify the nuances of the chatbot's interaction strategies.

7. **Multiple Topics Analysis:** Messages encompassing several topics underwent detailed scrutiny. Calculating the frequency of each combination yielded an understanding of the most to least prevalent topic pairings.

The research also sought to identify areas where the chatbot lacked the capacity to provide satisfactory responses. The **Knowledge Gap Analysis** for the chatbot documented the topics where the chatbot responded to users' queries with errors. This analysis facilitated the identification of potential subject areas for expanding the chatbot's knowledge domain.

4.4 USER ENGAGEMENT

Informed by the work of Booth *et al.* [2023], this research set out to understand the distinct user groups by examining user interaction patterns with the chatbot. The research involved analysing interaction distribution and frequency and classifying users according to their distinctive interaction characteristics.

4.4.1 USER DISTRIBUTION

The values obtained from the Interaction Timing metric aided in analysing the distribution of user messages, enabling a visual representation of the specific days and times when users sent messages [17, 132]. Employing the Seaborn library [134], the **seaborn.heatmap** function enabled the creation of a heatmap. This visualisation tool graphically depicted the distribution of all user interactions across different times and days.

4.4.2 USER SATISFACTION

The study conducted an analysis of star ratings to assess user sentiments towards chatbots. The chatbot prompted users to *rate the value of the information they received* through a star rating system, ranging from 1 to 5 stars. This rating system allowed users to provide feedback and assess their overall satisfaction with the chatbot's performance. Rate the value of the info they received

To gain deeper insights into this type of feedback, a cluster-type analysis was conducted based on the frequency of star ratings provided by users. Specifically, the research focused on three groups: users who rated two times, three times, and four times. The research employed descriptive statistics to quantify user sentiment at various levels of user engagement. A line graph illustrated the frequency and average of ratings. This visualisation aided in discerning user satisfaction trends over time, specifically observing if user ratings enhanced following successive evaluations.

4.5 SENTIMENT ANALYSIS

Sentiment analysis aims to determine users' positive, negative, or neutral sentiments about the chatbot. In alignment with the methodologies employed by El-Ansari and Beni-Hssane [2023] and Sawant *et al.* [2021], which analysed only user-generated feedback, this research concentrated exclusively on the written feedback accompanying users' star ratings, as this provided insight into how users perceived their experience with the chatbot. Pre-set or menu-based responses are not ideal for sentiment analysis because they restrict user expression, introduce bias, and may not capture the depth and complexity of user sentiments. The subsequent section delineates the approach taken to capture and interpret the sentiment embedded in users' messages.

4.5.1 ENGLISH-BASED SENTIMENT ANALYSIS

Due to the limited availability of sentiment analysis tools for Portuguese text, the research adopted a translation-based methodology detailed in Tavares *et al.* [2021], translating user conversations from Portuguese to English [124]. The need for this translation-based approach arose as most sentiment analysis tools cater predominantly for English text.

The Python **translate** [126] library translated the dataset from Portuguese to English. It involved a function facilitating the translation process by utilising a translator object [126]. The data, once translated, was stored in a new column labelled *translated_content.* The limitation of using the **translate** library was that it imposed a daily limit of 15,000 characters, with each query not exceeding 500 characters [79].

To overcome these limitations, a code segment iterated through each *translated_content* row to identify and populate missing *translated_content*. The code first checked if the translated content was empty while the original content existed. If this condition was met, the code searched for previous rows with matching content and non-empty translated content. This search utilised Boolean indexing and a bitwise AND operator. Upon matching the rows, the code selected the most recent row and copied the translated content to the empty *translated_content* row.

This process helped fill in the missing translated content by leveraging existing translations for identical content. It ensured that the translation efforts were maximised and reduced the reliance on the translate library's limitations, enabling a more comprehensive translation coverage for the text data in the DataFrame.

For the rows in the *translated_content* that exceeded 500 characters, manual translation was performed by the researcher, thus ensuring accurate and meaningful translations for the longer content.

4.5.2 VADER

Valance Aware Dictionary and sEntimnet Reasoner (VADER) was used to obtain the sentiment scores of the translated text. VADER is a lexicon and rule-driven tool for sentiment analysis tailored to manage text originating from social media [69]. It employs quantitative methods to calculate sentiment measures [69]. Given that user feedback in this study was also conversational and informal, VADER's capabilities were well-suited for analysing this type of text. The sentiment scores were calculated on the translated text using the VADER sentiment analysis tool from the NLTK library. The following was the process undertaken to establish the polarity scores for each translated user message:

1. The research initiated an instance of the **SentimentIntensityAnalyzer** for the sentiment analysis process.
2. A function called ***get_sentiment*** was defined to calculate sentiment polarity scores for each row in the *translated_content* column. It checked if the input was a string and, if so, utilised the **SentimentIntensityAnalyzer** to obtain sentiment scores. The system stored these scores in a column named *sentiment*.
3. The system generated a compound sentiment score for each row, ranging from 1 (most positive sentiment) to -1 (most negative sentiment). The system assigned a default sentiment score of 0.0 for inputs that were not strings.

4.5.3 SENTIMENT ANALYSIS PROCESS

After calculating the sentiment scores, a comprehensive analysis followed to determine the overall sentiment and distribution. To illustrate these findings effectively and highlight any patterns, outliers, or inconsistencies among the methods, graphical representations were constructed using Matplotlib and Seaborn libraries.

Various methods were employed to analyse the sentiment scores derived from the chatlogs. These methods encompassed descriptive and statistical techniques, each serving a specific purpose in unveiling valuable insights. The research applied the **.describe()** statistical method to the sentiment scores. This method provided a statistical summary, encompassing the count of sentiment score instances, mean and median values, and standard deviation to assess score dispersion and quantify positive, negative, and neutral feedback instances.

The research utilised a histogram for the visualisation of the distribution of sentiment scores. This visualisation provided insights into whether sentiment distribution was skewed towards positive, negative, or neutral sentiments.

The methods mentioned earlier facilitated a meaningful interpretation of the data. The forthcoming findings section will present the results of this analysis, offering insights into the interactions observed within the chatlogs.

05

FINDINGS

AURORA USAGE ANALYSIS

5 FINDINGS

The following sections present the findings of user metrics and dynamics, scrutinising interaction modalities, response rates, and user categorisations. The analysis also examined the nature and quality of conversations, addressing engagement times, chatbot response accuracy, and user feedback. Finally, the analysis presents the topics of conversation.

5.1 USER METRICS AND DYNAMICS

There were 1043 unique users who used the chatbot from 15 October 2018 to 30 September 2021. The analysis excluded a subset of users, labelled as "bounced users," comprising 321 users or 31% of the total. Bounced users had less than 6 message exchanges with the chatbot and did not request or receive parenting information. These users sent odd messages and seemed to be testing the chatbot's capabilities or using it for fun. This reduction left the research with 718 (69%) users with 24350 interactions with the chatbot. The subsequent sections present the findings gathered from the various metrics, as depicted in Table 1.

User interactions with the chatbot occurred through user-written questions or menu selections, with the latter being the preferred input method, accounting for 87% of interactions. The preference for the menu selection method likely stems from its convenience in accessing or exploring information. When users asked the chatbot a question, the chatbot responded 82% of the time about the related topic, demonstrating a high response rate. However, there were instances where the chatbot failed to understand the question, accounting for 18% of the interactions. Given that only the chatlogs were available for analysis, it was not possible to confirm the exact reasons behind the occurrences of the errors. However, common causes include technical issues, backend processing errors, user interruptions, and system overload [118, 125]. Among the written messages, 46% of user queries resulted in an error message from the chatbot. These error messages were typically characterised by their extensive length, encompassing multiple sentences, or their vagueness, lacking any recognisable keywords that the chatbot could identify, such as the query by P926:

P926: When is the best time to express? – user message sent at 04:32 on December 2018.

Errors also emerged when users employed slang or terminology specific to regional dialects of Portuguese. For instance, discrepancies arose with the term “breast”, where the European Portuguese "mama" and the Brazilian and Angolan Portuguese "chucha" led to misunderstandings.

The response rate referred to the proportion of chatbot prompts that elicited user replies, serving as a metric for evaluating the degree of user engagement with the chatbot. Users responded to 46% (Table 1) of the chatbot's prompts, whether within an ongoing conversation or to begin a new one. This low response rate suggests that not all prompts generated active user engagement. Several factors might contribute to this observation, such as users discontinuing their interaction with the chatbot, acquiring the necessary knowledge and no longer needing further assistance, or the prompts' timing not aligning with users' convenience to respond.

TABLE 1 - CONVERSATION METRICS OF USERS AND CHATBOT INTERACTIONS

Users	
Bounced users (removed)	321 (31%)
Active users	718 (69%)
Interaction Style	
Menu-based Interactions	87%
Written Messages Interactions	13%
Message Response Rate	
Users' response to chatbot prompts	46%
Chatbot correct responses	82%
Chatbot error responses	18%
Usage Interaction	
Number of interactions	34 (± 31)
Number of Sessions	6 (±6)
Tenure	82 days (± 127)
Unique days (user messages only)	5 (±4)

The analysis examined the average number of user interactions with the chatbot, the number of sessions, user tenure (counting the days from the first to the last use of the chatbot) and unique days. The term *unique days* refers to calendar days on which a user interacted with the chatbot. Rather than counting each interaction as a separate event, this metric aggregates all interactions occurring on the same calendar day into a single unique one.

The analysis showed that users, on average, interacted with the chatbot 34 times, with a standard deviation of 31. The wide spread of values around the mean suggests that user interactions varied significantly between users. On average, each user participated in 6 sessions with the chatbot, with moderate variability, as indicated by the standard deviation of 6. The narrower spread of values around the mean indicates relatively less variability than the number of interactions.

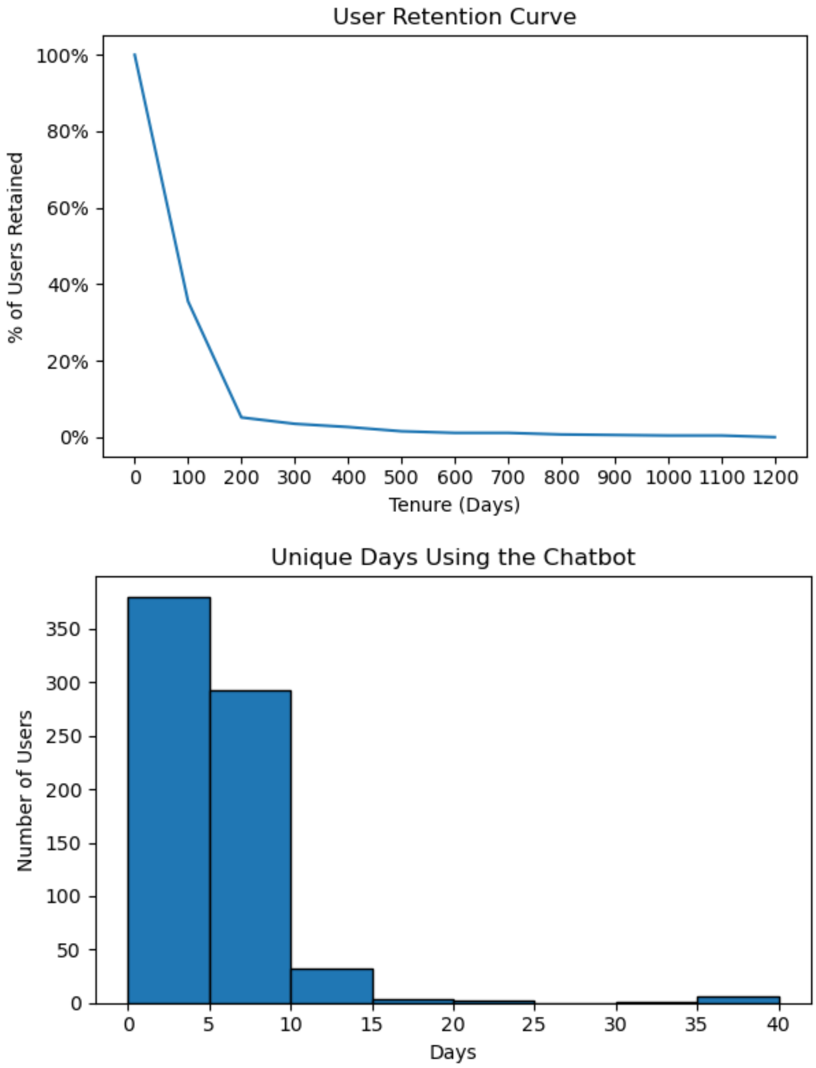

Figure 4 - User retention and unique days use of the chatbot, excluding bounced users.

User tenure was 82 days, with 127 days as the standard deviation, indicating a considerable variation in user engagement. As per Figure 4, 61.42% of users used the chatbot for 82 days or less. This finding suggests that user engagement was sporadic, with the most ardent user engaging with the chatbot for 1157 days. Therefore, users engaged with the chatbot on multiple occasions throughout their tenure. However, the average number of unique days users utilised the chatbot was only 5 days, with a standard deviation of 4 days. While user tenure showed significant variability, the number of unique days of use is more consistent among users.

The analysis revealed a substantial difference between user tenure and the number of unique days of use. Examining these two variables aimed to differentiate between users who were consistently active over a long period and those whose interactions were more sporadic.

5.1.2 USER TYPES

On analysing the conversation logs, we categorised four distinct user groups (see Table 2) based on their subscription preferences:

- Free Content Users: Users (678 users) who only accessed the free content without opting for any subscription or engaging with the healthcare professional during the 4 day free trail.
- Free Health Engagement: Users (27 users) who accessed the free content and chose to engage with the healthcare professional during the 4 day free trial period.
- 28-day Subscription: Users (9 users) who subscribed for 28 days, benefiting from a comprehensive program offering tips on feeding, sleep, and direct communication with the healthcare professional.
- 7-day Subscription: Users (4 users) who opted for a 7-day subscription plan, which granted them access to additional sleep information and the ability to interact with the healthcare professional for personalised advice.

Users who engaged with the healthcare professional during the free content period sent an average of 53 messages per user, which was higher than the rest of the free content users who sent an average of 30 messages per user. The free content users interacted approximately 50% less and engaged approximately 86% less compared to the 7-day plan users, and approximately 77% less and 88% less compared to the 28-day plan users.

TABLE 2 - USER GROUPS THAT ENGAGED WITH THE AURORA CHATBOT AND THE HEALTHCARE PROFESSIONAL.

Features	7-day Subscription	28-day Subscription	Free Health Engagement	Free Content Users
Users (%)	4 (0.6%)	9 (1.25%)	27 (3.8%)	678 (94.35%)
Tenure	204 (±81)	301 (±349)	71 (±91)	78 (±120)
Total Interactions	103 (±20)	207 (±75)	53 (±35)	30 (±20)
Number of Sessions	27 (±6)	62 (±14)	15 (±3)	18 (±2)

The 28-day plan users had the highest average session count with 62 sessions per user, followed by the 7-days plan users with 27 sessions per user. Both subscription groups had higher session counts (approximately over 244\% more) compared to users who engaged with the healthcare professional (15 sessions per user) during the free trail and the rest of the free content users (18 sessions per user). However, users who engaged with the healthcare professional for free and the rest of the free content users exhibited relatively consistent session frequencies, with lower standard deviations compared to the subscription groups.

Users who subscribed to the 28-day plan had the longest average tenure, with an average of 300 days, followed by users subscribed to the 7-day plan, with an average of 204 days. In contrast, users who engaged with the healthcare professional during the free trial period and the rest of the free content users had shorter average tenures of 71 days and 78 days, respectively. Interestingly, users who engaged with the healthcare professional for free exhibited a shorter range of tenure, as indicated by the lower standard deviation, suggesting variability in the duration of engagement among this group.

The 7-days plan users sent an average of 103 messages per user, while 28-days plan users sent an average of 207 messages per user, reflecting a substantial increase of approximately 100\% in message volume between the two subscription groups. This discrepancy in message volume suggests that the duration of the subscription plan may influence the intensity of user interaction, with longer subscription periods allowing for more prolonged engagement and interaction with the chatbot.

Additionally, an examination of the conversations among users who engaged with the healthcare professional revealed that their main concerns revolved around specific baby sleep problems, with fewer questions about breastfeeding, and general baby healthcare. Users were able to explain their concerns in greater detail, enabling the healthcare professional to gather additional information about the user's problem. Unlike the chatbot's automated responses, which where constrained by pre-programmed responses, the healthcare professional possessed the flexibility to probe further and ask clarifying questions to better grasp the nuances of the user's problem, and in turn provide a personalised response. Some users also expressed curiosity about the identity of the healthcare professional, indicating a desire for reassurance and credibility in the advice provided. The message exchange between users and the healthcare professional often assumed a conversational tone, allowing for a more fluid and interactive dialogue.

However, it is important to acknowledge that users sometimes encountered delays in receiving responses, as the healthcare professional was not always immediately available to attend to queries. Despite this limitation, the hybrid-interaction enabled the healthcare professional the opportunity to follow up with users to assess the effectiveness of recommendations provided and provide further assistance to the user.

5.2 Nature and Quality of Conversations

In the following sections, the findings presented are for user engagement times, chatbot accuracy response, user feedback for star rating, and sentiment of comments. This multifaceted analysis aimed to identify emergent patterns and trends in user interactions, evaluate the chatbot's efficacy in dispensing accurate and relevant information, and scrutinise quantitative and qualitative user satisfaction indicators.

	12am	1am	2am	3am	4am	5am	6am	7am	8am	9am	10am	11am	12pm	1pm	2pm	3pm	4pm	5pm	6pm	7pm	8pm	9pm	10pm	11pm
Monday	112	114	19	46	16	32	9	57	193	115	279	253	124	221	281	226	129	56	131	99	143	208	241	124
Tuesday	182	40	35	13	27	36	4	39	82	108	151	168	314	334	255	190	219	151	103	167	135	197	314	205
Wednesday	108	48	38	49	23	6	51	13	111	103	213	363	261	217	378	258	161	115	203	135	137	310	453	215
Thursday	154	46	8	28	28	22	51	80	156	144	227	300	430	208	251	327	292	220	189	191	147	387	387	164
Friday	140	151	12	18	6	22	61	42	118	95	237	188	207	188	233	187	184	133	157	130	97	165	258	136
Saturday	158	104	26	12	36	28	15	13	71	68	451	171	151	137	250	202	186	198	178	129	191	180	372	320
Sunday	137	66	60	101	62	16	16	44	153	85	135	178	147	128	232	181	153	110	103	172	102	196	380	178

Figure 5 - Time of messages sent to Aurora relating to sleep.

5.2.1 USER USAGE TIMES

Interactions were recorded and organised by time and day on a table, as depicted in Figure 5, to understand when users engaged with the chatbot. Notably, users displayed their highest engagement levels between 10:00 and 15:00 and from 21:00 to 00:00, predominantly from Monday to Friday. Thursdays stood out as the day with the most frequent messages transmitted to the chatbot. Figure 5 demonstrates a decline in the overall volume of messages sent to the chatbot as the weekend approaches, with Sunday registering the lowest number of messages. Despite these peaks, the p-value (0.4738) is much higher than 0.05, which suggests that there is no statistically significant difference in the number of messages between the different days of the week.

However, data suggested that parents engaged with the chatbot during critical parenting moments (see Figure 5). For instance, messages such as the one from user P380, transmitted at 00:00, frequently focused on queries associated with sleep challenges. As substantiated by existing literature, these interaction patterns aligned with standard infant sleep schedules [104]. It is customary for babies to have a shorter mid-morning nap around 10:00 and a more extended nap lasting from 12:00 to 15:00.

> *P380: I cannot get my baby to sleep before 23h30-00h30. Even if he goes to bed at about 22h00, he will not fall asleep before 23h30-00h30. - user message sent at 00:00 on November 2019.*

5.2.2 CHATBOT ACCURACY RESPONSE

As previously mentioned, the chatbot had a correct response rate of 82%, which includes both menu-selections and user-typed messages. This response rate referred to instances where users received chatbot responses without encountering error messages. To further analyse the chatbot's performance of how appropriately it answered user-typed messages, a subset of 1709 messages from 454 users was examined (Table 3).

TABLE 3 - CHATBOT ACCURACY ANALYSIS

Users and Messages	
Users	454
Total written Messages	1709
Chatbot Message Responses	
Partially Addressed response	64% (1094 messages)
Single-topic complete response	20% (342 messages)
Referral responses	16% (273 messages)

There were instances similar to P1026, where users posed complex questions that involved multiple topics that included sleep-and-feeding, sleep-and-health, or sleep-and-feeding-and-health issues questions. In these instances, the chatbot would provide a *partial* response focusing only on one topic, either sleep or feeding, disregarding the other topics mentioned or the full context of the question. This response accounted for 64% of user messages containing multi-topic queries.

> *P1026: My baby cannot sleep for 3 hours straight, and it is most likely because of colic, but I am not sure, and I have been to the doctor because of it but I have not had much help and I have been waking up constantly* - user message sent at 16:00 on December 2018.

> *Aurora: P1026, the very first tip, and one that impacts the entire sleep program.* 😊 *Quick and abrupt changes rarely bring the expected result. On the contrary, they tend to destabilise and often bring results contrary to those initially expected.* 😉

In the instance represented by user message P1026, the query featured multiple interrelated topics—specifically, sleep disturbances and potential colic. Despite the multifaceted nature of this query, the chatbot confined its response to the single topic of sleep, thereby failing to address the complexity inherent in the user's question. Aurora rarely asked users to simplify their questions to enhance comprehension. The lack of user guidance on how to pose questions may have contributed to the chatbot's limitations in effectively processing complex or multi-faceted queries.

When users asked questions that only related to sleep or feeding, the chatbot demonstrated its ability to provide that topic's answer. Such *single-topic responses* accounted for approximately 20% of the chatbot's replies. These responses included scenarios where the chatbot requested additional details about the baby, such as age, to deliver customised and age-appropriate recommendations.

Single-topic responses also encompassed instances where the chatbot displayed a peculiar behaviour of confirming the topic and then deferring the delivery of pertinent information. In such cases, the chatbot would respond to the user's query by indicating that tips would be provided in the forthcoming days, thus delaying an appropriate response. Subsequently, it would prompt the user to provide a star rating and feedback.

When the chatbot failed to address a user's query, it either generated an error message or presented an option for the user to connect with the healthcare professional, Cláudia. These responses were labelled as *referral responses*. This feature was only available to paid users and was activated when users typed "Cláudia" in their messages. Cláudia would then review the users' messages and provide them with a more personalised response. In total, 13 (1.8%) users exchanged messages with Cláudia, yielding an average of 24 messages exchanged per user.

5.2.3 ANALYSIS OF USER FEEDBACK: STAR RATING

The chatbot asked *how users would rate the value of the information they received*. The dataset for this analysis segment included 1302-star ratings, contributed by 402 users. The primary objective of this analysis was to discern the frequency at which users provided their ratings and to examine the implications of these frequencies on the average ratings received.

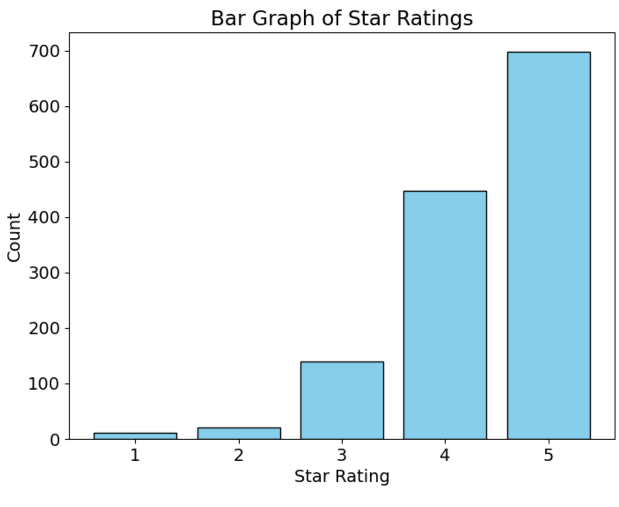

FIGURE 6 - BAR GRAPH ILLUSTRATING THE DISTRIBUTION OF ALL STAR RATINGS PROVIDED BY USERS FOR CHATBOT EVALUATION.

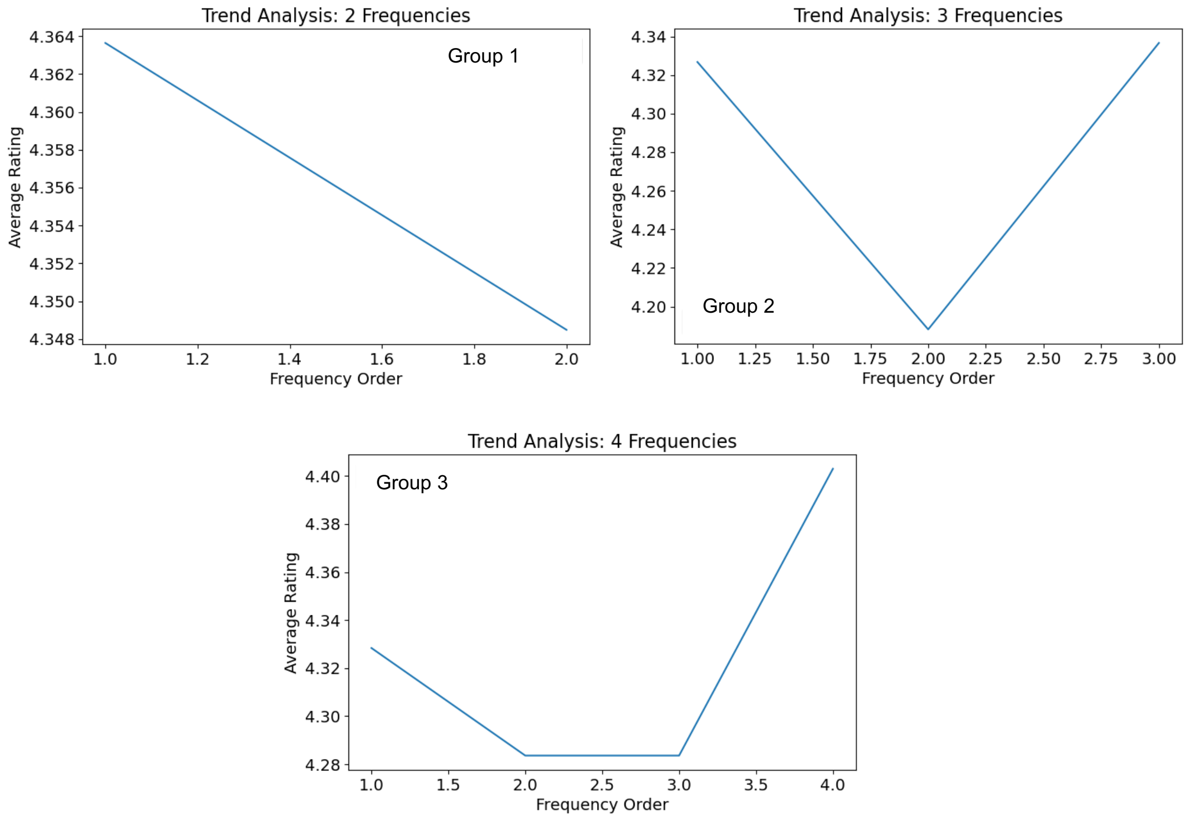

FIGURE 7 - THREE LINE PLOTS ILLUSTRATING THE DISTRIBUTION OF AVERAGE STAR RATINGS FOR USERS WHO EVALUATED THE CHATBOT 2, 3, AND 4 TIMES, RESPECTIVELY, INDICATING CHANGES IN USER SENTIMENT WITH INCREASED INTERACTIONS.

The findings illustrated in Figure 6 show that 53% of users rated the chatbot with 5 stars, indicating a high level of satisfaction. Several users (34.33%) gave a 4-star rating, signifying positive feedback. However, a relatively smaller percentage of users rated the chatbot with 1, 2, or 3 stars, accounting for 0.85%, 1.54%, and 10.75%, respectively. Overall, the feedback indicates that a substantial proportion of users' experience with the chatbot was positive, with the majority articulating their utmost satisfaction by assigning a 5-star rating.

The analysis indicated a variance in the frequency of user ratings. Of the total ratings, 20% were from users who rated the chatbot only once. Conversely, 12% of users issued two ratings, suggesting more interaction with the chatbot. A further 25% of users contributed three ratings, while 33% gave four ratings. Remarkably, 10% of users demonstrated heightened interest by rating their experience five times or more.

As shown in Figure 7, an analysis was conducted by plotting the ratings on a line graph based on frequency, particularly of users who voted more than once. In Group 1, which consists of users who rated the chatbot on two occasions, the average rating (4.356 stars) was relatively stable despite a slight dip, indicating consistency in user satisfaction. In Group 2, comprising users who submitted ratings three times, the average rating initially dipped (4.2 stars) slightly but subsequently increased (4.34 stars), suggesting a positive response to user feedback. In Group 3, encompassing users who rated the chatbot four times, where users provided the highest ratings, the average rating experienced a notable upward trend, indicating enhanced user satisfaction as they engaged more frequently.

The analysis revealed slight variations in average ratings across different user groups and their rating frequencies. These ratings increase suggests that more frequent engagement equates to heightened user satisfaction. However, these observed variations among different user groups are relatively minor and are closely related to overall star rating. The cumulative average star rating of 4.37 stars signifies a generally favourable perception of the chatbot among this user base.

5.2.4 USER FEEDBACK: SENTIMENT ANALYSIS

As previously mentioned, the sentiment analysis focused solely on user-written feedback accompanying star ratings, which gauged positive, negative, or neutral sentiments. This research excluded pre-set and menu-based responses since they did not convey user sentiment. Out of the user population, only 7% of users (comprising 74 written feedback entries) responded with comments about their experience with the chatbot.

The mean sentiment score was positive, at 0.119. This score suggested that, on average, feedback about the Aurora chatbot was mildly positive, although the median was exactly 0. The mean score indicated that half of the scores were precisely neutral, and the other half were either positive or negative, but overall balanced out. The standard deviation is approximately 0.36, suggesting that there was quite a wide variation in the sentiment scores.

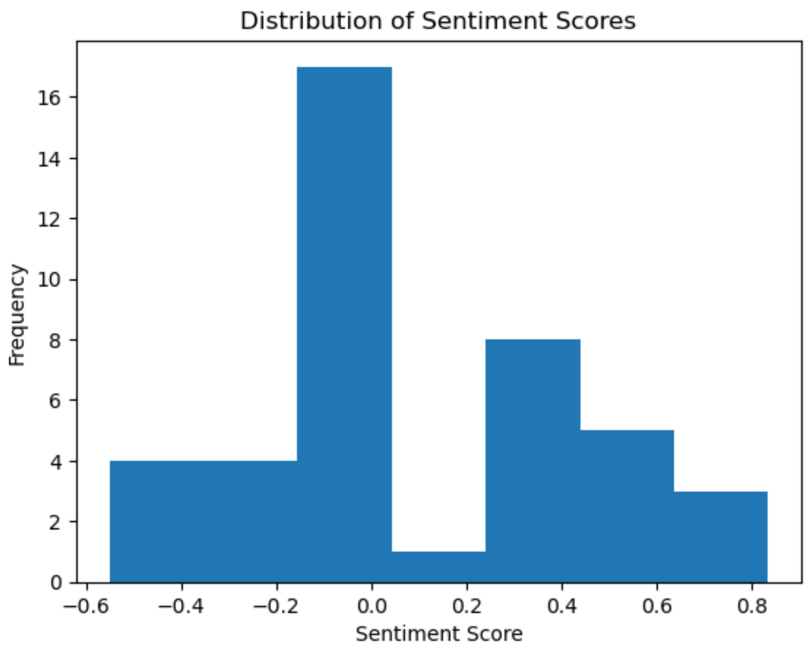

FIGURE 8 - DISTRIBUTION OF SENTIMENT SCORES FROM USER FEEDBACK.

Regarding the distribution of sentiment, there were more positive feedback messages (17 instances) than negative ones (11 instances), with a significant number of neutral messages (14 instances). While the feedback was generally more positive than negative, many users did not express a strong sentiment. However, as depicted in Figure 8, some feedback fell into the mildly negative category. This finding reflected user comments where they expressed a liking for the chatbot while expressing dissatisfaction with a particular aspect, e.g., paying to use the chatbot. Similarly, other users, such as P254 and P997, appreciated the tips provided by the chatbot but felt that the information was not novel or unique.

P254: I wanted to find something new, something I hadn't read yet or didn't know. – user message sent at 18:20 on May 2019.

P997: It's not that I don’t like it. Only the tips that are given are not new ones, I already use them, so I haven't added anything new. – user message sent at 01:50 on December 2018.

5.3 Topics of Conversations

The topics of conversation with Aurora were diverse and included baby sleep problems, breastfeeding and baby feeding, and healthcare-related discussions. Sleep problems accounted for 80% of the conversations, followed by 11% related to breastfeeding and baby feeding in general. Healthcare-related topics, encompassing the utilisation of medical products, constituted the remaining 9% of the conversations. In subsequent sections, exploration of these topics in greater depth aims to yield a comprehensive understanding of user preferences and needs.

5.3.1 Conversations about Sleep

The chatlogs included numerous messages on sleep issues that impacted the mother and the child. Predominantly, the issues revolved around the mother's sleep deprivation, emotional strain, impact on well-being, decision-making regarding sleep training, changing sleep patterns, baby sleep schedules, causes of night-time disruptions, and co-sleeping with parents. Users could use menu options to gain more information about baby sleep, such as in Figure 9. Other users opted for typing out their questions:

> *P75: My little one is almost 7 months old and used to sleep super well…now she wakes up hourly at night. I don't know if it is a growth spurt, or do I need to have more patience or change something? – user message sent at 23:06 on April 2020.*

Similar to P75, many users struggled to understand why their baby woke up several times at night. Numerous users sent messages describing a common scenario where their infants would wake up multiple times throughout the night and could only return to sleep after being fed. These users sought guidance and support to gain a deeper understanding of their babies' sleep patterns and to facilitate a more restful and uninterrupted night's sleep.

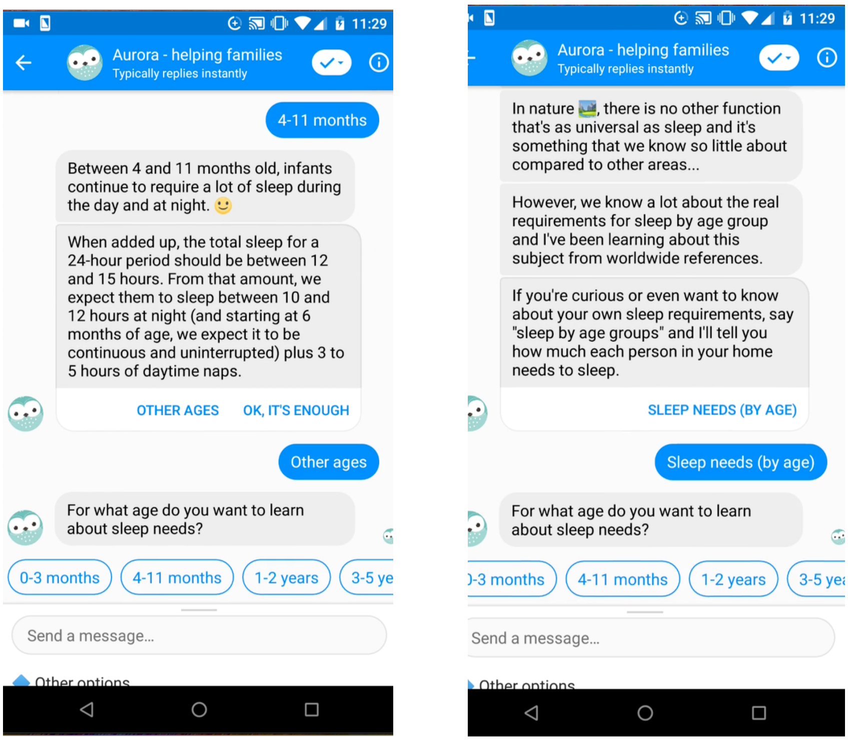

FIGURE 9 - AURORA'S CONVERSATION WITH A USER ABOUT SLEEP.

Some of the user's sleep-related messages also incorporated references to feeding and health concerns. These references likely indicated that users perceived a connection between sleep problems and underlying issues related to feeding or health. Approximately 13% of the sleep-related messages included terms associated with feeding and healthcare topics. Among these messages, 11% focused on sleep-and-feeding-related problems, and 2% discussed sleep-and-healthcare concerns, implying that users considered feeding or healthcare factors to be potential causes or contributors to their baby's sleep problems.

A similarly re-occurring query was that of P864, where their baby would only fall asleep after, or while being breastfed. This trend pointed to a dependency on breastfeeding or bottled milk as a method for infants to fall asleep, highlighting a strong association between feeding and sleep. The chatbot responded by asking for more information about the baby, such as the baby's age, to provide a personalised solution for the user.

> *P864: ... she resists bedtime often… only falls asleep after having her milk bottle…*
> *- user message sent at 15:16 on November 2018.*

Within the messages concerning baby sleep issues, a subset of parents attributed sleep difficulties to potential health factors, such as teething and colic. P793 and P494 sought specific advice on how to improve their babies' sleep considering these conditions. P793's baby was grappling with colic, while P494's baby was experiencing a teething phase.

> *P793: I would like my baby to sleep more consecutive hours and not have colic. - user message sent at 20:39 on March 2019.*

> *P494: My baby is teething and wakes up many times crying. - user message sent at 10:03 on February 2018.*

Some of these users had returned to the chatbot with positive updates, sharing that their babies' sleep had shown improvement after implementing the suggested strategies. On the other hand, some parents who gave feedback expressed the need for additional tips and advice, as they had not yet seen the desired improvements in their babies' sleep patterns. This type of feedback results from Aurora’s corpus being static and providing limited or repetitive tips, especially for users who only accessed the free subscription service. This static nature of the corpus resulted in repetitive responses and a lack of dynamic content generation based on varied user queries. The problem might also stem from how queries are being asked and how the chatbot constructs its answers, which showcases the content limitation inherent in the system.

5.3.2 CONVERSATIONS ABOUT BREASTFEEDING AND FEEDING

Feeding practices for babies appeared in multiple user queries. Questions regarding feeding times, determining if the baby is full, and exploring various breastfeeding positions were commonly raised. Many mothers expressed concerns about potentially having low milk production and sought advice on how to identify low milk production and how to increase their breast milk supply. Parents were also worried about their babies not receiving sufficient milk or not feeding for adequate durations.

Discussions surrounding breast-related issues and care were prevalent. Numerous mothers, including P584, actively sought advice and guidance regarding breast care. The main topic of concern in this domain included mastitis, which is characterised by symptoms such as hardened breasts or difficulties with milk flow. In addition, some users expressed that they experienced pain during breastfeeding and sought advice on alleviating it.

> *P584: I had cracks in my nipples in the first few days, it is already better, but they are still not good. It continues to hurt me in the first few seconds of breastfeeding. What can I do to make it stop hurting? - user message sent at 12:25 on 6 December 2018.*

Other mothers sought guidance on discouraging their babies from using the breast as a pacifier to fall asleep and eliminating night-time feedings. Additionally, there were queries about gradually weaning the baby off breastfeeding and introducing bottle feeding into the routine.

5.3.3 CONVERSATIONS ABOUT HEALTHCARE

The chatbot also received queries from users regarding a range of other baby-related topics, such as teething, acne, reflux, colic, bowel movements, respiratory issues, allergies, fever, and medications. These diverse inquiries shed light on parents' concerns and questions about their babies' healthcare and well-being.

Among the various user queries, a recurring theme was parents seeking assistance in recognising symptoms and guidance on treating illnesses affecting their babies. An example of such an instance is P729, who approached the chatbot with a specific concern regarding their baby's colic. They expressed a need for information on effectively alleviating and managing the discomfort associated with the colic. Unfortunately, the chatbot could not respond adequately as the question was outside its knowledge domain.

> *P729: How do you treat colic? What medication can I use on my baby? – user message sent at 23:29 on December 2018*
>
> *Aurora: I think I still do not know enough about this topic to be able to help you. Do you have another question to ask? Use simple sentences or just the key phrase. – chatbot message sent at 23:29 on December 2018*

Similar to P729 and P101, several users inquired about the safety of taking medication while breastfeeding, with allergy medication and immune boosters frequently cited for both the baby and mother. These queries highlighted the desire for clarity and guidance regarding the compatibility of certain medications with breastfeeding and appropriateness for babies.

> *Participant 101: While breastfeeding, can I take some medication for allergies? - user message sent at 14:23 on 9 December 2019.*
>
> *Aurora: ❓❓❓ - Chatbot message sent at 14:23 on 9 December 2019*

In some instances, exemplified by interactions with users P729 and P101, the chatbot signalled its inability to comprehend the user's query. The chatbot occasionally employed emojis, such as a thinking face or question marks, which failed to provide substantive or clarifying information. This usage of emojis did not contribute to a meaningful exchange and could potentially have introduced ambiguity in the communication. Moreover, there were several instances in which the chatbot delayed its responses and indicated that it would furnish more information in the following days.

5.4 SUMMARY

In summary, the research findings offered a detailed look at how users interacted with the Aurora chatbot. These findings cover preferred interaction modes, response rates, conversation quality and topics discussed. These metrics critically evaluated the chatbot's functional scope and informational accuracy. User feedback, captured through star ratings and sentiment analysis, further identified the chatbot's strengths and areas for improvement. Textual analysis unveiled the most prevalent topics in discussions and areas falling outside the chatbot's domain.

The following section will interpret the collected data, assess its broader implications for the field of chatbots in parenting, and position it in the context of established academic literature.

06

DISCUSSION

6 DISCUSSION

The primary objective of this research was to delve into the user interactions of parents with the Aurora chatbot, understanding the predominant topics and engagement patterns. Several findings were uncovered, which are closely aligned with previous studies and offer novel insights.

6.1 BEYOND THE CORPUS

Aurora, designed initially to disseminate information on baby sleep and breastfeeding, frequently encountered baby sleep-related issues as the dominant subject in user interactions. This finding resonates with Mindell *et al.* [2006], who highlighted the prevalence of baby sleep-related queries in parenting forums. Similarly, Honaker and Meltzer [2014] emphasised that baby night-time wakings remain a prominent worry for parents, especially considering that 25-50% of infants over six months of age still face this challenge [90]. The consistent emergence of sleep as a focal topic with the chatbot, which is similar to other parenting chatbot studies [129, 140], was unsurprising. However, the findings revealed that parents often articulated their concerns interconnectedly, discussing sleep disturbances in conjunction with other issues such as healthcare concerns or breastfeeding dependencies for sleep. This complexity underscored a limitation in the design of Aurora’s response generation system, which often focused on addressing a single topic from a multifaceted query. This design choice resulted in responses that might not fully satisfy parents seeking comprehensive advice on interconnected issues. Therefore, the issue lies not in the domain-specificity of Aurora, but in the chatbot's response generation mechanism, which needs improvement to handle multifaceted parenting questions more effectively.

This complexity in query formulation frequently appeared in questions posed by users, revealing a discernible gap between user expectations and the chatbot's output. Users approached Aurora, anticipating a nuanced, context-aware conversation, particularly given the complex and interconnected nature of parenting issues like sleep disturbances, breastfeeding, and general child healthcare. Of particular concern was Aurora's method of handling unfamiliar queries. Aurora often returned error messages, emoji placeholders, promised more tips in future interactions or narrowly focused responses that failed to account for the broader context presented by the user. This approach is less than ideal, especially in a domain as time-sensitive as parenting, where concerns require immediate clarity.

The limitations observed in Aurora's responses stem from its design as a domain-specific and rule-based chatbot [135]. Such a design may confine its ability to grasp user questions and sustain a fluid dialogue. As highlighted in Luger and Sellen [2016], chatbots often struggle with meeting user expectations, especially when their design is restrictive. that This observation aligns with Janssen *et al.* [2021], who identified content deficiency and comprehension as a reason for chatbot failure and user discontinuation.

The limitations of Aurora's corpus and its resulting partial responsiveness to user queries resonate with broader HCI issues concerning the design and functionality of chatbots, particularly in parenting support. As Brandtzaeg and Følstad [2017] point out, the efficacy of a chatbot is highly dependent on its ability to handle a diverse set of queries, a factor crucial for user satisfaction. In line with this, Aurora's high error response rate manifests a significant gap in question processing and corpus depth.

Cláudia's attempt to organically expand Aurora's corpus had dual implications. While providing some degree of adaptability, it led to a suboptimal situation where only 64% of user-typed queries received partial answers. This limitation stemmed from the chatbot's dependence on Cláudia for content creation and monitoring for gaps in the corpus. Further complicating matters, Aurora failed to effectively educate users on how to converse with the chatbot, a deficiency evident in the minimal usage of error messages urging users to simplify their queries or explain how to structure their questions. Moreover, 16% of users-written queries received redirection to consult Cláudia directly, which suggests that users favoured personalised responses. This scenario underscored the need for robust human supervision and constant updates to the chatbot's corpus, for it to remain effective and relevant. This observation aligns with Coghlan *et al.* [2023] assertion that chatbots require continuous human supervision to operate successfully. Despite advancements in AI and NLP, chatbots still fall short of replicating the depth of expertise and nuanced understanding demonstrated by human professionals.

In addition, the chatbot demonstrated limitations in accommodating linguistic variations [26], failing to capture the full scope and intricacy of user queries. This shortfall is particularly relevant given the diverse origins of Portuguese speakers, who may come from different countries with slight variations in the use of the language. Coleman *et al.* [2023] observation about the distinct needs of rural and semi-urban settings resonates with the findings, highlighting the importance of a user-centred design approach to consider geographical and cultural variations. Shawar and Atwell [2007] propose employing NLP techniques to adapt a chatbot to the unique linguistic characteristics of different user demographics, leveraging a user-provided training corpus.

Furthermore, the findings reveal an ethical imperative: chatbots like Aurora should not enable conversational capabilities without the supporting architecture of robust language models, especially in a domain as sensitive as parenting [49]. A limited and static corpus constrains the chatbot to provide partial or incomplete responses, leading to unmet user expectations and diminishing user engagement and satisfaction. This problem emphasises the need for integrating large language models to better handle the variety and complexity of queries that users, in this case, parents, are likely to pose [73]. However, technology is not a panacea; even the most sophisticated language models have their limitations. The real challenge lies in balancing technological capabilities with ethical considerations to achieve a system that provides accurate and appropriate support. While not a definitive solution, the thoughtful integration of advanced language models offers a pathway for incremental improvements, potentially enhancing the user experience without overpromising results.

The disconnect in user expectations underscores not only the limitations of the existing chatbot architecture but also reveals critical insights into user behaviour and needs. It reiterates the necessity for future chatbot designs to move beyond merely responding to isolated queries and towards a more holistic, context-aware interaction model [19].

6.2 FOSTERING PARENT ENGAGEMENT WITH CHATBOTS

The research identified distinct groups of user engagement with the Aurora chatbot, revealing usage patterns aligned with findings from other healthcare domains. For example, Booth *et al.* [2023] found similar trends in mental healthcare chatbots, noting that user engagement depended on diverse needs and preferences of users. O'Brien and Toms [2008] further support this idea of heterogeneous user interactions, emphasising that users interact differently with digital tools based on their unique requirements and contexts.

Given the varied user engagement patterns observed in Aurora, special considerations are essential when approaching HCI design for parenting support systems like Aurora. Drawing from the research of Ford [2012] and Murkoff [2014], as infants develop, the nature of parental challenges changes. The first few months post-birth are often marked by urgent queries and support-seeking behaviours, indicating an elevated need for information and a higher level of engagement with digital aids like chatbots. The challenges presented by early childhood development necessitate that HCI design for parenting chatbots consider the evolving needs of parents. This period-specific interaction intensity corresponds with HCI theories on adaptive interfaces, such as those discussed by Findlater and McGrenere [2004], who argue that systems should adapt to the changing needs of users.

Given the observed phase-specific intensity of user interactions and the theoretical framework on adaptive interfaces, there is a compelling rationale for developing features in parenting chatbots that can adapt their content and interaction styles based on the child's age and developmental milestones. For example, the chatbot might prompt the parent for updates on the child's developmental milestones to enable personalised advice. This feature would then shift discussions from newborn sleep patterns to advice on toddler sleep training. The dynamism inherent in the parenting journey necessitates digital aids capable of evolving alongside the parent. Such adaptability holds significance in the context of parenting, given that the need for information and support changes as children mature and reach new developmental milestones [51].

Given that Aurora’s corpus only covers some elements of the early stages of a baby’s life, the expectation of perpetual user interaction is not applicable. This particular focus makes the decline in user interaction an especially relevant metric for evaluating its success, as described by O'Brien and Toms [2008] engagement model. According to O'Brien and Toms [2008], the engagement cycle involves initial engagement, sustained engagement, and subsequent disengagement, which may be either positive—indicating meeting the user’s needs—or negative, reflecting user frustration or lack of usefulness. The model suggests that the ultimate metric for success is not sustained engagement but rather the effective resolution of parental queries [102]. Therefore, a decline in user interaction could signify that Aurora has successfully met users' educational and supportive needs, thus facilitating autonomous parenting.

This discussion underscores the broader HCI implications for designing chatbots in the parenting domain, where user engagement is not static but a dynamic element influenced by various contextual factors. Therefore, to foster increased and meaningful parent engagement, HCI researchers and designers must look toward creating systems as nuanced and adaptable as the lives they aim to assist.

6.3 TIMELY AND PERSONALISED ADVICE

The research noted a link between the timing of user messages and typical baby sleeping schedules, a finding consistent with the research by Silva *et al.* [2013] of Portuguese parenting trends. Particularly noteworthy is the surge in messages during late evening hours, coinciding with cultural practices in Portugal, where children frequently go to bed around 21:00 or even later [122]. Another explanation for this trend could be that when babies sleep, parents have a window of opportunity to interact with the chatbot, seeking advice or support. The finding indicates that users actively consulted Aurora during critical moments, like during bedtime routines or breastfeeding challenges.

This immediacy in seeking chatbot assistance is consistent with research by Brandtzæg *et al.* [2021], highlighting how chatbots serve as instant support systems when individuals face pressing concerns or seek quick information. The observations emphasise the benefit of round-the-clock accessibility that chatbot applications offer in the context of parenting. As highlighted in the literature, a fundamental expectation from a chatbot is its constant availability [20, 22, 63]. Consequently, Aurora's practice of delaying information undermines its core purpose and hampers its effectiveness in fulfilling users' needs. When a parent seeks information, a chatbot's response must be prompt and precise, catering to the user's immediate needs. In addition, the research found that parents were seeking personalised information tailored to their specific childcare challenges. This observation was substantiated when parents commented that the information dispensed by the chatbot was neither new nor trivial, indicating a desire for more contextually relevant guidance.

These findings align with HCI principles advocating for a user-centric approach in the design and functionality of interactive systems. The use of co-design methodologies, as suggested by Sanders and Stappers [2008], could provide a collaborative platform for incorporating these user preferences. Co-design would enable an iterative feedback loop, allowing real-time adjustments based on parent feedback, thereby enhancing the chatbot's ability to offer personalised and context-sensitive information. In addition, the involvement of parents in the co-design process could establish ethical safeguards and better address the multifaceted nature of parenting challenges, thus potentially elevating user satisfaction.

Additionally, the chatbot could implement real-time sentiment analysis features to address users' negative sentiments. Upon detecting negative sentiment in a user's message, the chatbot could alter its response strategy, provide additional resources, or recommend a connection with a human healthcare expert. Alternatively, incorporating machine learning or artificial intelligence mechanisms can empower Aurora to adapt and learn from user interactions, broadening its operational domain [68].

This research has contributed valuable insights into the user interactions and preferences with the Aurora chatbot, offering important guidance for developing more effective and personalised parenting chatbot applications. Through ongoing refinement and the extension of the chatbot's functionalities, Aurora possesses the potential to evolve into a valuable resource for parents in search of assistance and information for addressing a wide range of parenting-related issues.

07

CONCLUSION

7 CONCLUSION

This research's objective was to comprehend how parents engage with and interpret the information provided by the Aurora chatbot, as well as to identify the dominant topics of discussion in the context of childcare. The research has affirmed that baby sleep-related concerns dominate parental inquiries, a finding consistent with existing literature. However, this consistency also reveals Aurora's inability to effectively handle queries that involve multiple, interconnected topics, like sleep disturbances, breastfeeding, and general child healthcare. Aurora tended to address only one topic at a time, which resulted in incomplete responses when users ask complex questions.

Aurora's rule-based, domain-specific design poses limitations in meeting user expectations for context-aware responses. This inadequacy contributes to user dissatisfaction and aligns with broader HCI issues, emphasising the need for more versatile chatbot architectures. Despite attempts to organically expand the chatbot's corpus, Aurora often failed to provide complete and satisfactory answers to user queries, which indicated the need for a more adaptive, user-centric approach. The research highlighted the significance of timely and personalised advice, as user interactions often occur during critical moments, such as bedtime routines or breastfeeding challenges. Aurora's delayed responses negate its core purpose and compromise its effectiveness in offering immediate, context-specific support to parents.

These findings endorse the necessity for future iterations of the Aurora chatbot to adopt a more holistic, context-aware model. Understanding users' intricate concerns and expectations through co-designing could enhance the chatbot's utility. This research also provided insights into the specific childcare concerns of Portuguese-speaking parents, thus offering a culturally contextual understanding that may inform the development of more targeted chatbot functionalities.

While the research offered valuable insights, it is not devoid of limitations. The research focused primarily on Portuguese-speaking users and did not investigate the potential effects of linguistic or cultural diversity within this group. Furthermore, the chatbot's architecture itself, being rule-based and domain-specific, posed a limitation in grasping the full complexity of user queries. This specific design and focus on the early stages of a child's life make its architecture potentially inapplicable to chatbots with different frameworks or objectives. Given these limitations, these findings may not be applicable to chatbots across diverse cultural, linguistic, and technological settings. The restrictions in the study's scope thus necessitate caution in extrapolating its conclusions to a broader array of chatbots worldwide.

Future research could broaden the scope by incorporating user interviews and advanced AI mechanisms or exploring chatbots in other niche parenting domains. Furthermore, future research could include detailed logs of menu selections. This data would enable a more thorough investigation into how the design and positioning of menu options impact user queries and engagement, providing deeper insights into user behaviour and preferences. The findings advocate redesigning the Aurora chatbot, emphasising the need for a more context-sensitive and multi-dimensional response system. Such an approach can better serve parents' nuanced and often interconnected concerns.

This research contributes to the growing work on HCI in the parenting domain. By highlighting the advantages and limitations of a chatbot platform like Aurora, the research provided a comprehensive view of the multifaceted challenges and opportunities in utilising chatbots for parenting support in the digital age. Furthermore, the research goes beyond the mere identification of issues by emphasising continuous improvement and adopting a user-centric design philosophy for prospective advancements in this domain. This multi-dimensional contribution serves as a reference point for academicians and practitioners as a roadmap for future exploratory and evaluative studies in the intersection of HCI and parenting.

REFERENCES

ACM Citation Style and Reference Formats

REFERENCES

[1] Alaa Abd-Alrazaq, Zeineb Safi, Mohannad Alajlani, *et al.* 2020. Technical metrics used to evaluate health care chatbots: scoping review. *Journal of medical Internet research 2020*, 22, Article 6:e18301, (February 2020), 15 pages. https://doi.org/10.2196%2F18301

[2] Martin Adam, Michael Wessel and Alexander Benlian. 2021. AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, 31, Article 2, (March 2021), 427-445. https://doi.org/10.1007/s12525-020-00414-7

[3] Eleni Adamopoulou and Lefteris Moussiades. 2020. An Overview of Chatbot Technology. In *Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations,* June 5-7, 2020, Neos Marmaras. Springer Nature: Cham, Switzerland, 373-383. https://doi.org/10.1007/978-3-030-49186-4_31

[4] Eleni Adamopoulou and Lefteris Moussiades. 2020. Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2, Article 100006, (November 2020), 18 pages. https://doi.org/10.1016/j.mlwa.2020.100006

[5] Ritu Agarwal and Mani Wadhwa. 2020. Review of state-of-the-art design techniques for chatbots. *SN Computer Science*, 1, Article 246, (July 2020), 13 pages. https://doi.org/10.1007/s42979-020-00255-3

[6] Aurora Tech AI. Aurora For Families. Retrieved May 15, 2023 from http://auroratechai.com/pt/familia/

[7] Manal Almalki and Fahad Azeez. 2020. Health chatbots for fighting COVID-19: A scoping review. *Acta Informatica Medica*, 28, (December 2020), 241-247. http://dx.doi.org/10.5455/aim.2020.28.241-247

[8] Flora Amato, Stefano Marrone, Vincenzo Moscato, *et al.* 2017. Chatbots Meet eHealth: Automatizing Healthcare. In *Proceedings of the Workshop on Artificial Intelligence with Application in Health*, November 14, 2017, Bari, Italy, 40-49. https://ceur-ws.org/Vol-1982/paper6.pdf

[9] Amazon. 2023. What is Alexa? Retrieved July 11, 2023 from https://developer.amazon.com/en-US/alexa

[10] Brooke Auxier, Monica Anderson, Anderson Perrin and Erica Turner. 2020. Parents' attitudes – And experiences – Related to digital technology. Retrieved July 27, 2023 from https://www.pewresearch.org/internet/2020/07/28/parents-attitudes-and-experiences-related-to-digital-technology/

[11] BabyCenter. 2023. BabyCentre free app: My Pregnancy & Baby Today. Retrieved May 15, 2023 from https://www.babycentre.co.uk/mobile-apps

[12] BabySparks. 2021. What is BabySparks. Retrieved May 15, 2023 from https://babysparks.com/about-us/

[13] Sabine Baker, Matthew R. Sanders and Alina Morawska. 2017. Who uses online parenting support? A cross-sectional survey exploring Australian parents' internet use for parenting. *Journal of Child and Family Studies*, 26, Article 3, (November 2017), 916-927. https://doi.org/10.1007/s10826-016-0608-1

[14] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell. 2021. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* 2021, Virtual Event, Canada. Association for Computing Machinery, 610–623. 10.1145/3442188.3445922

[15] Steven Bird, Ewan Klein and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. (1st. ed.). O'Reilly Media, Inc., Sebastopol, CA.

[16] Kathrin Blagec, Georg Dorffner, Milad Moradi and Matthias Samwald. 2020. A critical analysis of metrics used for measuring progress in artificial intelligence. *eprint arXiv:2008.02577*, (Aug 2020). 10.48550/arXiv.2008.02577

[17] Frederick Booth, Courtney Potts, Raymond Bond*, et al.* 2023. A Mental Health and Well-Being Chatbot: User Event Log Analysis. *JMIR Mhealth Uhealth 2023*, 11, Article e43052, (July 2023), 17 pages. http://dx.doi.org/10.2196/43052

[18] Ellen Brady and Suzanne Guerin. 2010. "Not the romantic, all happy, coochy coo experience": A qualitative analysis of interactions on an Irish parenting website. *Family relations*, 59, Article 1, (March 2010), 14-27. http://dx.doi.org/10.1111/j.1741-3729.2009.00582.x

[19] Petter B. Brandtzaeg and Asbjørn Følstad. 2017. Why People Use Chatbots. In *the 4th International Conference on Internet Science*, November 22-24, 2017, Thessaloniki, Greece. Springer International Publishing, 377-392. http://dx.doi.org/10.1007/978-3-319-70284-1_30

[20] Petter B. Brandtzæg, Marita B. Skjuve, Kim K. Dysthe and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots: Social Support in Chatbots. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 08-13, 2021, Yokohama, Japan. ACM, New York, NY, 1-13. http://dx.doi.org/10.1145/3411764.3445318

[21] Virginia Braun and Victoria Clarke. 2008. Using thematic analysis in psychology. *Qualitative research in psychology*, 3, Article 2, (July 2008), 77-101. https://doi.org/10.1191/1478088706qp063oa

[22] David Cancel and Dave Gerhardt. 2019. *Conversational marketing: How the world's fastest growing companies use chatbots to generate leads 24/7/365 (and how you can too)*. John Wiley & Sons, Inc., Hoboken, New Jersey, NY.

[23] Jacky Casas, Marc-Olivier Tricot, Omar Abou Khaled*, et al.* 2020. Trends & methods in chatbot evaluation. In *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion)*, October 25-29, 2020, Virtual Event, Netherlands. Association for Computing Machinery, New York, NY, USA, 280-286. http://dx.doi.org/10.1145/3395035.3425319

[24] Leyland Cecco. 2024. Air Canada ordered to pay customer who was misled by airline's chatbot. Retrieved June 25, 2024 from https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit

[25] I-Chiu Chang, Yi-Syuan Shih and Kuang-Ming Kuo. 2022. Why would you use medical chatbots? interview and survey. *International Journal of Medical Informatics*, 165, Article 104827, (Sep 2022), 11 pages. https://doi.org/10.1016/j.ijmedinf.2022.104827

[26] Ana Paula Chaves, Jesse Egbert, Toby Hocking*, et al.* 2022. Chatbots Language Design: The Influence of Language Variation on User Experience with Tourist Assistant Chatbots. *ACM Trans. Comput.-Hum. Interact.*, 29, Article 13, Article 2, (2022), 1-38. https://doi.org/10.1145/3487193

[27] Anthony Chen, Gabriel Stanovsky, Sameer Singh and Matt Gardner. 2019. Evaluating Question Answering Evaluation. In *2nd Workshop on Machine Reading for Question Answering*, Nov, 2019, Hong Kong, China. Association for Computational Linguistics., 119-124. https://doi.org/10.18653/v1/D19-5817

[28] Daniel Y. Chen. 2017. *Pandas for everyone: Python data analysis*. (1st. ed.). Addison-Wesley Professional, Boston.

[29] Alebachew Chiche and Betselot Yitagesu. 2022. Part of speech tagging: a systematic review of deep learning and machine learning approaches. *Journal of Big Data*, 9, Article 10, (Jan 2022), 25 pages. https://doi.org/10.1186/s40537-022-00561-y

[30] Wesley Chun. 2001. *Core python programming*. (1st. ed.). Prentice-Hall Inc., Upper Saddle River, NJ.

[31] Google Cloud. 2023. Dialogflow Messenger. Retrieved July 28, 2023 from https://cloud.google.com/dialogflow/es/docs/integrations/dialogflow-messenger

[32] Simon Coghlan, Kobi Leins, Susie Sheldrick*, et al.* 2023. To chat or bot to chat: Ethical issues with using chatbots in mental health. *DIGITAL HEALTH*, 9, Article 20552076231183542, (Jun 2023), 11 pages. https://doi.org/10.1177%2F20552076231183542

[33] Toshka Coleman, Sarina Till, Jaydon Farao*, et al.* 2023. Reconsidering priorities for digital maternal and child health: community-centered perspectives from South Africa. *Proceedings of the ACM on Human-Computer Interaction*, 7, Article 290, (October 2023), 31 pages. https://doi.org/10.1145/3610081

[34] Sarah M. Coyne, Brandon T. McDaniel and Laura A. Stockdale. 2017. “Do you dare to compare?” Associations between maternal social comparisons on social networking sites and parenting, mental health, and romantic relationship outcomes. *Computers in Human Behavior*, 70, (May 2017), 335-340. https://doi.org/10.1016/j.chb.2016.12.081

[35] Anthony G. Crocco, Miguel Villasis-Keever and Alejandro R Jadad. 2002. Two wrongs don't make a right: harm aggravated by inaccurate information on the Internet. *Paediatrics*, 109, Article 3, (Mar 2002), 522-523. https://doi.org/10.1542/peds.109.3.522

[36] Douglas Crockford. 2012. JSON. Retrieved August 2, 2023 from https://www.rfc-editor.org/rfc/rfc7159

[37] Meta for Developers. 2023. Messenger API Updates for Europe and Japan. Retrieved August 11, 2023 from https://developers.facebook.com/docs/messenger-platform/europe-japan-updates

[38] Meta for Developers. 2023. Overview for the Messenger Platform. Retrieved July 28, 2023 from https://developers.facebook.com/docs/messenger-platform/overview#how-it-works

[39] Meta for Developers. 2023. Facebook Pages API. Retrieved September 4, 2023 from https://developers.facebook.com/docs/pages/#

[40] Erin Dietsch. 2019. Consumers don't fully trust healthcare chatbots, study finds. Retrieved July 17, 2023 from https://medcitynews.com/2019/01/consumers-healthcare-chatbots/

[41] Jennifer L. Doty, Jodi Dworkin and Jessica H. Connell. 2012. Examining digital differences: Parents' online activities. *Family Science Association*, 17, Article 2, (Jan 2012), 18-39. http://dx.doi.org/10.26536/FSR.2012.17.02.02

[42] Jodi Dworkin, Jessica Connell and Jennifer Doty. 2013. A literature review of parents' online behavior. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 7, Article 2, (July 2013), 1-10. http://dx.doi.org/10.5817/CP2013-2-2

[43] Anas El-Ansari and Abderrahim Beni-Hssane. 2023. Sentiment Analysis for Personalized Chatbots in E-Commerce Applications. *Wireless Personal Communications*, 129, Article 3, (Feb 2023), 1623-1644. https://doi.org/10.1007/s11277-023-10199-5

[44] Guido A. Entenberg, Malenka Areas, Andrés J. Roussos, *et al.* 2021. Using an Artificial Intelligence Based Chatbot to Provide Parent Training: Results from a Feasibility Study. *Social Sciences*, 10, Article 11, (November 2021), 426. https://doi.org/10.3390/socsci10110426

[45] Guido A. Entenberg, Sophie Mizrahi, Hilary Walker, *et al.* 2023. AI-based chatbot micro-intervention for parents: Meaningful Engagement, Learning, and Efficacy. *Frontiers in Psychiatry*, 14, Article 1080770, (January 2023), 10 pages. https://doi.org/10.3389/fpsyt.2023.1080770

[46] EUR-Lex. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Retrieved August 2, 2023 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[47] AI Tools Explorer. 2023. The History of Natural Language Processing: From ELIZA to GPT. Retrieved August 13, 2023 from https://aitoolsexplorer.com/ai-history/the-history-of-natural-language-processing-from-eliza-to-gpt/

[48] Leah Findlater and Joanna McGrenere. 2004. A comparison of static, adaptive, and adaptable menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* April 24-29, 2004, Vienna, Austria. Association for Computing Machinery, New York, NY, USA, 89–96. https://doi.org/10.1145/985692.985704

[49] Luciano Floridi and Josh Cowls. 2019. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1, Article 1, (2019), 14 pages. https://doi.org/10.1162/99608f92.8cd550d1

[50] Asbjørn Følstad, Marita Skjuve and Petter Bae Brandtzaeg. 2019. Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. In *Internet Science: INSCI 2018 International Workshops*, October 24–26, 2019, St. Petersburg, Russia. Springer International Publishing, 145-156. http://dx.doi.org/10.1007/978-3-030-17705-8_13

[51] Gina Ford. 2012. *The Complete Sleep Guide For Contented Babies & Toddlers*. (2nd. ed.). Ebury Publishing, Houghton, South Africa.

[52] Fotos Frangoudes, Marios Hadjiaros, Eirini C Schiza*, et al.* 2021. An overview of the use of chatbots in medical and healthcare education. In *International Conference Human-Computer Interaction*, July 24–29, 2021, HCII 2021, Virtual Event. Springer International Publishing, 170-184. http://dx.doi.org/10.1007/978-3-030-77943-6_11

[53] Yujia Gao, Jinu Jang and Diyi Yang. 2021. Understanding the Usage of Online Media for Parenting from Infancy to Preschool At Scale. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, May 7, 2021, Yokohama, Japan. Association for Computing Machinery, New York, NY, USA, 1-12. https://doi.org/10.1145/3411764.3445203

[54] Eun Go and S. Shyam Sundar. 2019. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, (Aug 2019), 304-316. https://doi.org/10.1016/j.chb.2019.01.020

[55] Ian Goodfellow, Yoshua Bengio and Aaron Courville. 2016. *Deep Learning*. MIT Press, Cambridge, Massachusetts.

[56] Google. 2023. Hey Google. Retrieved October 11, 2023 from https://assistant.google.com

[57] South African Government. 2013. Protection of Personal Information Act 4 of 2013. Retrieved August 4, 2023 from https://www.gov.za/sites/default/files/gcis_document/201409/3706726-11act4of2013protectionofpersonalinforcorrect.pdf

[58] Lyndsay Grant. 2009. *Learning in Families: A review of research evidence and the current landscape of Learning in Families with digital technologies*. Futurelab, General Educators Report 59. Futurelab, Bristol, U.K.

[59] Christine Grové. 2021. Co-developing a mental health and wellbeing chatbot with and for young people. *Frontiers in psychiatry*, 11: 606041, (February 1 2021), 12 pages. https://doi.org/10.3389/fpsyt.2020.606041

[60] Jonathan Grudin and Richard Jacques. 2019. Chatbots, Humbots, and the Quest for Artificial General Intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, May, 2019, Glasgow, Scotland. Association for Computing Machinery, New York, NY, USA, 1-11. https://doi.org/10.1145/3290605.3300439

[61] Cristin M. Hall, Erica D. Culler and Anne Frank-Webb. 2016. Online Dissemination of Resources and Services for Parents of Children with Autism Spectrum Disorders (ASDs): a Systematic Review of Evidence. *Review Journal of Autism and Developmental Disorders*, 3, (June 2016), 273-285. https://doi.org/10.1007/s40489-016-0083-z

[62] Louis Hickman, Stuti Thapa, Louis Tay, *et al.* 2022. Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25, Article 1, (Nov 2022), 114-146. https://doi.org/10.1177/1094428120971683

[63] Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, *et al.* 2014. Towards an open-domain conversational system fully based on natural language processing. In *COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers,* August 23-29, 2014, Dublin, Ireland. Dublin City University and Association for Computational Linguistics, Dublin, Ireland, 928-939. https://aclanthology.org/C14-1088

[64] Alexis Hiniker, Sarita Y. Schoenebeck and Julie A. Kientz. 2016. Not at the Dinner Table: Parents' and Children's Perspectives on Family Technology Rules. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing,* February 27 - Mach 2, 2016, San Francisco, CA. Association for Computing Machinery, New York, NY, USA, 1376–1389. https://doi.org/10.1145/2818048.2819940

[65] Sarah M. Honaker and Lisa J. Meltzer. 2014. Bedtime Problems and Night Wakings in Young Children: An Update of the Evidence. *Paediatric Respiratory Reviews*, 15, Article 4, (Dec 2014), 333-339. https://doi.org/10.1016/j.prrv.2014.04.011

[66] Hsiu-Fang Hsieh and Sarah E. Shannon. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, 15, Article 9, (Nov 2005), 1277-1288. https://doi.org/10.1177/1049732305276687

[67] John D. Hunter. 2007. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9, Article 3, (May - June 2007), 90-95. https://doi.ieeecomputersociety.org/10.1109/MCSE.2007.55

[68] Shafquat Hussain, Omid A. Sianaki and Nedal Ababneh. 2019. A Survey on Conversational Agents/Chatbots Classification and Design Techniques. In *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019)*, Mar 27-29, 2019, Matsue, Japan. Springer International Publishing, 946-956. http://dx.doi.org/10.1007/978-3-030-15035-8_93

[69] Clayton J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, June 1-4, 2014, Ann Arbor, Michigan, U.S.A. University of Michigan, Ann Arbor, Michigan U.S.A., 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

[70] Antje Janssen, Lukas Grützner and Michael Breitner. 2021. Why do Chatbots fail? A Critical Success Factors Analysis. In *Proceedings of the Forty-Second International Conference on Information Systems (ICIS)*, November 16, 2021, Austin, Texas, 17 pages. https://www.researchgate.net/publication/354811221

[71] Tingting Jiang, Qian Guo, Yuhan Wei*, et al.* 2022. Investigating the relationships between dialog patterns and user satisfaction in customer service chat systems based on chat log analysis. *Journal of Information Science*, (Sep 2022). https://doi.org/10.1177/01655515221124066

[72] Zhao Jianqiang and Gui Xiaolin. 2017. Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis. *IEEE Access*, 5, (Feb 2017), 2870-2879. http://dx.doi.org/10.1109/ACCESS.2017.2672677

[73] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems,* April 19, 2023, Hamburg, Germany. Association for Computing Machinery, New York, NY, USA, 1-16. https://doi.org/10.1145/3544548.3581503

[74] Daniel Jurafsky and James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. (2nd. ed.). Pearson Prentice Hall, Upper Saddle River, New Jersey.

[75] Yury Kashnitsky. 2020. Topic 1. Exploratory Data Analysis with Pandas. Retrieved April 24, 2023 from https://www.kaggle.com/code/kashnitsky/topic-1-exploratory-data-analysis-with-pandas

[76] Anirudh Khanna, Bishwajeet Pandey, Kushagra Vashishta*, et al.* 2015. A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence. *International Journal of u- and e- Service, Science and Technology*, 8, Article 7, (July 2015), 277-284. http://dx.doi.org/10.14257/ijunesst.2015.8.7.28

[77] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini*, et al.* 2019. Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support. *Biomedical Informatics Insights*, 11, (Dec 2019), 9 pages. https://doi.org/10.1177/1178222619829083

[78] Christian Kubb and Heather M. Foran. 2020. Online Health Information Seeking by Parents for Their Children: Systematic Review and Agenda for Further Research. *J Med Internet Res*, 22, Article 8:e19985, (August 2020), 21 pages. https://doi.org/10.2196/19985

[79] MyMemory by translated LABS. MyMemory: API usage limits. Retrieved October 10, 2023 from https://mymemory.translated.net/doc/usagelimits.php

[80] Jane M. Lamp and Patricia A. Howard. 1999. Guiding parents' use of the Internet for newborn education. *MCN: The American Journal of Maternal/Child Nursing*, 24, Article 1, (Jan 1999), 33-36. https://doi.org/10.1097/00005721-199901000-00007

[81] Lauren M. Amaro, Nataria T. Joseph and Theresa M. de los Santos. 2019. Relationships of online social comparison and parenting satisfaction among new mothers: The mediating roles of belonging and emotion. *Journal of Family Communication*, 19, Article 2, (February 2019), 144-156. https://doi.org/10.1080/15267431.2019.1586711

[82] Serena Leggeri, Andrea Esposito and Luca Iocchi. 2018. Task-oriented Conversational Agent Self-learning Based on Sentiment Analysis. In *Workshop on Natural Language for Artificial Intelligence (NL4AI 2018) co-located with the 17th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2018). CEUR-WS*, 2018, Rome, Italy, 4-15. https://api.semanticscholar.org/CorpusID:53425861

[83] Sonia Livingstone, Giovanna Mascheroni and Elisabeth Staksrud. 2017. European research on children's internet use: Assessing the past and anticipating the future. *New media & society*, 20, Article 3, (Jan 2017), 1103-1122. https://doi.org/10.1177/1461444816685930

[84] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* May, 2016, San Jose, California, USA. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[85] Raquel Magalhães. 2022. Are Facebook Messenger Bots Still Relevant? Retrieved August 10, 2022 from https://landbot.io/blog/are-facebook-chatbots-still-relevant

[86] Wari Maroengsit, Thanarath Piyakulpinyo, Korawat Phonyiam*, et al.* 2019. A survey on evaluation methods for chatbots. In *Proceedings of the 2019 7th International Conference on Information and Education Technology (ICIET 2019)*, March, 2019, Aizu-Wakamatsu, Japan. Association for Computing Machinery, New York, NY, USA, 111-119. https://doi.org/10.1145/3323771.3323824

[87] Ines Montani Matthew Honnibal. 2023. spaCy 101: Everything you need to know. Retrieved July 1, 2023 from https://v2.spacy.io/usage/spacy-101

[88] Wes McKinney. 2010. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, 2010, Austin, Texas, U.S.A, 56-61. http://dx.doi.org/10.25080/Majora-92bf1922-00a

[89] Oliver Miles, Robert West and Tom Nadarzynski. 2021. Health chatbots acceptability moderated by perceived stigma and severity: A cross-sectional survey. *Digital Health*, 7, (Jan-Dec 2021), 7 pages. https://doi.org/10.1177/20552076211063012

[90] Jodi A. Mindell, Brett Kuhn, Daniel S. Lewin*, et al.* 2006. Behavioral treatment of bedtime problems and night wakings in infants and young children. *Sleep: Journal of Sleep and Sleep Disorders Research*, 29, Article 10, (Oct 2006), 1263-1276. https://doi.org/10.1093/sleep/29.10.1263

[91] Brent D. Mittelstadt, Patrick Allo, Mariarosaria Taddeo, *et al.* 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3, Article 2, (July-Dec 2016), 21 pages. https://doi.org/10.1177/2053951716679679

[92] João Luis Zeni Montenegro, Cristiano André da Costa and Luisa Plácido Janssen. 2022. Evaluating the use of chatbot during pregnancy: A usability study. *Healthcare Analytics*, 2, Article 100072, (November 1 2022), 9 pages. https://doi.org/10.1016/j.health.2022.100072

[93] Cláudia Matos Morgado. Social Problem - Our Genesis. Retrieved May 15, 2023 from http://auroratechai.com/pt/familia/

[94] Heidi Murkoff. 2014. *What to Expect the First Year*. Workman Publishing Company, Hachette, UK.

[95] Tom Nadarzynski, Oliver Miles, Aimee Cowie and Damien Ridge. 2019. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital health*, 5, (Aug 2019), 1-12. http://dx.doi.org/10.1177/2055207619871808

[96] Gina Neff and Peter Nagy. 2016. Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication*, 10, (Oct 2016), 4915-4931. Retrieved from http://ijoc.org/index.php/ijoc/article/view/6277/1804

[97] Preksha Nema and Mitesh Khapra. 2018. Towards a Better Metric for Evaluating Question Generation Systems. In *2018 Conference on Empirical Methods in Natural Language Processing*, Oct, 2018, Brussels, Belgium. Association for Computational Linguistics., 3950–3959. 10.18653/v1/D18-1429

[98] Tran Tin Nghi, Tran Huu Phuc and Nguyen Tat Thang. 2019. Applying AI chatbot for teaching a foreign language: An empirical research. *International Journal of Scientific and Technology Research*, 8, Article 12, (Dec 12 2019), 897-902. https://www.researchgate.net/publication/337965319_Applying_Ai_Chatbot_For_Teaching_A_Foreign_Language_An_Empirical_Research

[99] NLTK. 2023. Sample usage for portuguese_en. Retrieved April 23, 2023 from https://www.nltk.org/howto/portuguese_en.html

[100] Cecilie B. Nordheim. 2018. *Trust in chatbots for customer service–findings from a questionnaire study*. Master Thesis. University of Oslo, Oslo, Norway.

[101] Cecilie B. Nordheim, Asbjørn Følstad and Cato A. Bjørkli. 2019. An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interacting with Computers*, 31, Article 3, (2019), 317-335. http://dx.doi.org/10.1093/iwc/iwz022

[102] Heather L O'Brien and Elaine G Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59, Article 6, (February 2008), 938-955. https://doi.org/10.1002/asi.20801

[103] OpenAI. 2022. Introducing ChatGPT. Retrieved August 13, 2023 from https://openai.com/blog/chatgpt

[104] World Health Organization. 2023. Breastfeeding. Retrieved May 22, 2023 from https://www.who.int/health-topics/breastfeeding#tab=tab_1

[105] Talia Orr, Marsha Campbell-Yeo, Britney Benoit*, et al.* 2017. Smartphone and Internet Preferences of Parents: Information Needs and Desired Involvement in Infant Care and Pain Management in the NICU. *Advances in Neonatal Care*, 17, Article 2, (April 2017), 131-138. http://dx.doi.org/10.1097/ANC.0000000000000349

[106] Neri Van Otten. 2023. The History Of Natural Language Processing & Potential Future Breakthroughs. Retrieved August 11, 2023 from https://spotintelligence.com/2023/06/23/history-natural-language-processing/

[107] Rio Pramana, Debora, Jonathan J. Subroto*, et al.* 2022. Systematic Literature Review of Stemming and Lemmatization Performance for Sentence Similarity. In *IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)*, Nov 4-5, 2022, Yogyakarta, Indonesia. IEEE, 1-6. https://doi.org/10.1109/ICITDA55840.2022.9971451

[108] Mike Procter, Fuhua Lin and Bob Heller. 2018. Intelligent intervention by conversational agent through chatlog analysis. *Smart Learning Environments*, 5, Article 1, (Nov 2018), 30 pages. https://doi.org/10.1186/s40561-018-0079-5

[109] Python. 2023. Python. Retrieved July 16, 2023 from https://www.python.org/

[110] Alec Radford, Jeffrey Wu, Rewon Child*, et al.* 2019. *Language models are unsupervised multitask learners*. City.

[111] Noor I. M. Rahim, Noorminshah A. Iahad, Ahmad F. Yusof and Mohammed A. Al-Sharafi. 2022. AI-Based Chatbots Adoption Model for Higher-Education Institutions: A Hybrid PLS-SEM-Neural Network Modelling Approach. *Sustainability*, 14, Article 19, (Oct 2022), 22 pages. https://doi.org/10.3390/su141912726

[112] Ajay Ramachandran. 2019. User adoption of chatbots. *Available at SSRN 3406997*, (June 2019), 14 pages. https://dx.doi.org/10.2139/ssrn.3406997

[113] Bernadette M. Randles, Irene V. Pasquetto, Milena S. Golshan and Christine L. Borgman. 2017. Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, June 19-23, 2017, Toronto, ON, Canada. IEEE, 1-2. https://doi.org/10.1109/JCDL.2017.7991618

[114] Amon Rapp, Lorenzo Curti and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, Article 3, (March 2021), 86 pages. http://dx.doi.org/10.1016/j.ijhcs.2021.102630

[115] Evgeniia Razumovskaia, Goran Glavaš, Olga Majewska*, et al.* 2022. Crossing the Conversational Chasm: A Primer on Natural Language Processing for Multilingual Task-Oriented Dialogue Systems. *Journal of Artificial Intelligence Research*, 74, (July 13 2022), 1351-1402. https://doi.org/10.1613/jair.1.13083

[116] Replika. 2023. What is Replika? Retrieved September 11, 2023 from https://help.replika.com/hc/en-us/articles/115001070951-What-is-Replika-

[117] Elizabeth B.-N Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *Co-design*, 4, Article 1, (Sep 2008), 5-18. https://doi.org/10.1080/15710880701875068

[118] Manuela Sanguinetti, Alessandro Mazzei, Viviana Patti*, et al.* 2020. Annotating Errors and Emotions in Human-Chatbot Interactions in Italian. In *The 14th Linguistic Annotation Workshop*, Dec 12, 2020, Barcelona, Spain (Online). Association for Computational Linguistics, Stroudsburg, U.S.A., 148-159. https://aclanthology.org/2020.law-1.14

[119] Sahil Sawant, Ankit Vishwakarma, Prerana Sawant and Pasenjit Bhavathankar. 2021. Analytical and Sentiment based text generative chatbot. In *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, July 6-8, 2021, Kharagpur, India. IEEE, 1-7. https://doi.org/10.1109/ICCCNT51525.2021.9580069

[120] Bayan A. Shawar and Eric Atwell. 2007. Different measurements metrics to evaluate a chatbot system. In *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies (NAACL-HLT-Dialog '07),* April 26, 2007, Rochester, New York. Association for Computational Linguistics, U.S.A, 89–96. http://dx.doi.org/10.3115/1556328.1556341

[121] Shefaly Shorey, Thiam C. Tan, Jancy Mathews*, et al.* 2021. Development of a Supportive Parenting App to Improve Parent and Infant Outcomes in the Perinatal Period: Development Study. *J Med Internet Res*, 23, Article 12:e27033, (December 2021), 17 pages. https://doi.org/10.2196/27033

[122] Filipe Silva, Cláudia Silva, Lígia Braga and Ana Neto. 2013. Hábitos e problemas do sono dos dois aos dez anos: estudo populacional (Sleep habits and problems from two to ten years of age: population study). *Acta Pediatrica Portuguesa, Sociedade Portuguesa de Peiatria*, 44, Article 5, (November 2013), 196-202. https://doi.org/10.25754/pjp.2013.2898

[123] Patrica Silva. Aurora: especialista digital para jovens pais (entrevista a Cláudia Morgado, fundadora da Aurora). SuperToast. Podcast. (2 March 2021). Retrieved July 23, 2023 from https://www.youtube.com/watch?v=caDc-pdowcc

[124] Cátia Tavares, Ricardo Ribeiro and Fernando Batista. 2021. Sentiment analysis of Portuguese economic news. In *10th Symposium on Languages, Applications and Technologies, SLATE 2021*, August 10, 2021, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany, 1-13. http://dx.doi.org/10.4230/OASIcs.SLATE.2021.17

[125] Mikael Tedjopranoto, Andreas Wijaya, Levi Santoso and Derwin Suhartono. 2019. Correcting Typographical Error and Understanding User Intention in Chatbot by Combining N-Gram and Machine Learning Using Schema Matching Technique. *International Journal of Machine Learning and Computing*, 9, (Aug 1 2019), 471-476. http://dx.doi.org/10.18178/ijmlc.2019.9.4.828

[126] Terry Yin and Rafael Henter. 2017. Translate Python. Retrieved October 10, 2023 from https://translate-python.readthedocs.io/en/latest/

[127] Alan M. Turing. 2004. *Computing machinery and intelligence (1950). In Jack Copeland B. (Ed.), The Essential Turing: The Ideas that Gave Birth to the Computer Age*. Oxford University Press, Oxford.

[128] Aditya N. Vaidyam, Hannah Wisniewski, John D. Halamka, *et al.* 2019. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry*, 64, Article 7, (March 2019), 456-464. https://doi.org/10.1177/0706743719828977

[129] Lucia Vaira, Mario A. Bochicchio, Matteo Conte, *et al.* 2018. MamaBot: a System based on ML and NLP for supporting Women and Families during Pregnancy. In *IDEAS '18: Proceedings of the 22nd International Database Engineering & Applications Symposium*, June 18-20, 2018, Villa San Giovanni, Italy. Association for Computing Machinery, New York, NY, USA, 273–277. https://doi.org/10.1145/3216122.3216173

[130] Anu Venkatesh, Chandra Khatri, Ashwin Ram, *et al.* 2018. On evaluating and comparing conversational agents. In *NIPS 2017 Conversational AI workshop Part of Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*, Dec 8, 2018, Long Beach, CA. CA: Neural Information Processing Systems Foundation, San Diego, 10 pages. https://doi.org/10.48550/arXiv.1801.03625

[131] Tibert Verhagen, Jaap Van Nes, Frans Feldberg and Willemijn Van Dolen. 2014. Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, 19, Article 3, (April 2014), 529-545. http://dx.doi.org/10.1111/jcc4.12066

[132] V Vijayaraghavan, Jack Brian Cooper and J Rian Leevinson. 2020. Algorithm Inspection for Chatbot Performance Evaluation. *Procedia Computer Science*, 171, (January 1 2020), 2267-2274. https://doi.org/10.1016/j.procs.2020.04.245

[133] Richard S. Wallace. 2009. *The anatomy of ALICE. In Parsing the Turing Test*. Springer Netherlands.

[134] Michael L. Waskom. 2021. seaborn: statistical data visualization. *Journal of Open Source Software. The Open Journal*, 6, Article 60, (April 2021), 4 pages. https://doi.org/10.21105/joss.03021

[135] Fikadu Wayesa. 2020. Design and Implementation of a Rule Based Afaan Oromoo Conversational Chatbots. *International Journal of Scientific Research in Science and Technology*, 5, Article 8, (Nov-Dec 2020), 27-30. https://ijsrst.com/IJSRST20583

[136] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9, Article 1, (Jan 1966), 36-45. https://doi.org/10.1145/365153.365168

[137] Rainer Winkler and Matthias Söllner. 2018. Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. *Academy of Management Proceedings*, 2018, (April 1 2018), 40 pages. http://dx.doi.org/10.5465/AMBPP.2018.15903abstract

[138] Jill Wong, Agathe C. Foussat, Steven Ting*, et al.* 2021. A Chatbot to Engage Parents of Preterm and Term Infants on Parental Stress, Parental Sleep, and Infant Feeding: Usability and Feasibility Study (Preprint). *JMIR Pediatrics and Parenting*, 4, Article 4:e30169, (Oct-Dec 2021). https://doi.org/10.2196/30169

[139] Flow X. 2021. 2021 Facebook Messenger: What is Changing and What You Need to Know. Retrieved August 11, 2023 from https://flowxo.com/2021-facebook-messenger-what-is-changing-and-what-you-need-to-know/

[140] Deepika Yadav, Prerna Malik, Kirti Dabas and Pushpendra Singh. 2019. Feedpal: Understanding Opportunities for Chatbots in Breastfeeding Education of Women in India. *Proc. ACM Hum.-Comput. Interact. 3, CSCW*, Article 170, (Nov 2019), 30 pages. https://doi.org/10.1145/3359272

[141] Xinzhi Zhang, Rui Zhu, Li Chen*, et al.* 2022. News from Messenger? A Cross-National Comparative Study of News Media's Audience Engagement Strategies via Facebook Messenger Chatbots. *Digital Journalism*, (November 24 2022), 1-20. https://doi.org/10.1080/21670811.2022.2145329

[142] Darius Zumstein and Sophie Hundertmark. 2017. Chatbots – An Interactive Technology for Personalized Communication, Transactions and Services. *IADIS International Journal on WWW/Internet*, 15, Article 1, (Nov 2017), 96-109. Retrieved on August 4, 2023 from https://www.researchgate.net/publication/322855718_Chatbots_-_An_Interactive_Technology_for_Personalized_Communication_Transactions_and_Services