

# Designing Engaging Personalities for Conversational AI Agents to Enhance Employee Interaction in an Enterprise Crowdsourcing Context



ZhuoTing Wang

Masters Thesis

MSc Design for Interaction

August 2024

# Preface

In this report, you will read about a design research project focused on conversational agent (CA) personalities the research process behind our design interventions aimed at enhancing user engagement in an enterprise crowdsourcing context.

During my studies at TU Delft, I witnessed the emergence of a new generation of conversational AI, such as ChatGPT, and was constantly amazed by how these AI agents are reshaping the way humans collaborate with technology and, by extension, with each other. This thesis project offered an exciting opportunity to delve into the world of conversational agents, contributing to the broader understanding of how design can influence the impact of CAs within organizations. It also uncovered the social relationships that can develop between humans and CAs, prompting me to reflect on the range of social connections humans might develop with technology—such as love (Figure 1), trust, and even fear (such as HAL 9000 in Figure 2).

I believe this topic will become increasingly relevant as conversational AI continues to grow rapidly. In line with Nass et al.’s (1994) Computers Are Social Actors (CASA) theory, I am convinced that AI will increasingly take on the role of social actors in collaboration with humans. This thesis documents my contribution to this evolving landscape.

**Author:** ZhuoTing Wang

**Student ID:**5746418

**Thesis Committee:**

**Chair:** Prof. Alessandro Bozzon

**University supervisor:** Dr. Dave Murray-Rust

**Company supervisor:** Robert-Jan Sips

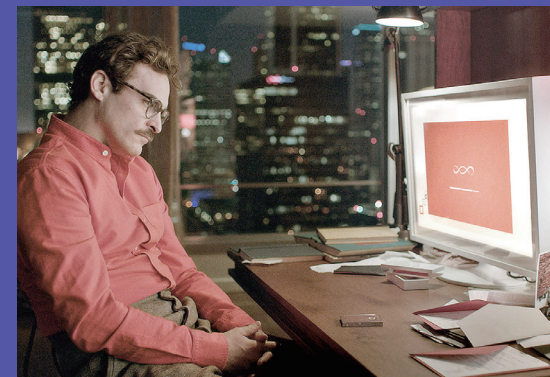


Figure 1. Human falling in love with a conversationa AI agent in movie “Her” (2013)

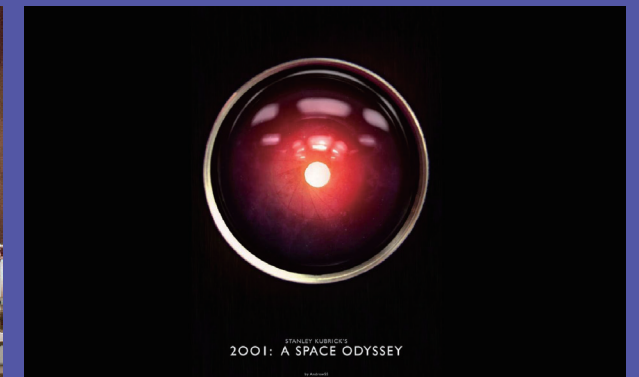


Figure 2. Conversationa AI agent - HAL 9000 in movie: “2001: A Space Odyssey” (1968)

# Acknowledgement

I would like to express my deep appreciation to everyone who has supported me throughout this journey.

While this thesis encapsulates the efforts of an intense 6-month journey, the support and contributions that helped me complete this project extend far beyond what is written here. I extend my sincere gratitude to my supervisory team: Prof. Alessandro Bozzon, Dr. Dave Murray-Rust, and Robert-Jan Sips. Thank you, Alessandro, for connecting me with this exciting opportunity, always being attentive to my progress, and offering honest and constructive feedback to help me reach higher. Thank you, Dave, for your supportive and insightful advice that always spark my thinking. And thank you, Robert-Jan, for your constant belief in this project and for sharing your industrial knowledge.

A special thank you goes to the incredible team at TKH —Viktorija, Linh, Anirvin, Angelos, Agnieszka and Andrzej. Your ideas and hard work brought this project to life, and I am grateful for your dedication. I also want to thank everyone who participated in my research and provided feedback.

I am truly grateful to my friends at TU Delft for the time we spent together and for the warm support you provided during this process. Last but not least, I am deeply thankful to my family and my partner for their love and unconditional belief in me.

ZhuoTing Wang

August, 2024

Delft, the Netherlands

# Summary

Conversational agents (CAs) are increasingly being adopted across various domains, with enterprises particularly interested in leveraging these technologies to explore new opportunities. CAs, such as Copilot, are transforming human-computer collaboration in the workplace by serving as effective work assistants. Additionally, their advanced natural language understanding capabilities position them as powerful tools for facilitating crowdsourcing within enterprises.

This thesis investigates the dual role of CAs in enterprises: as work assistants for employees and as facilitators of crowdsourcing. By acting as work assistants, CAs interact with a broad range of employees, making them well-suited to gather valuable insights that contribute to the company’s knowledge base and support collaborative problem-solving. The central research question explores how to effectively engage and motivate employees in this crowdsourcing process through CA personality design, which is relatively unexplored. Drawing on the “Computers Are Social Actors” (CASA) framework by Nass et al. (1994), this thesis examines CA personality design, aiming to enhance human-CA interaction by eliciting social responses from employees and thereby increasing their engagement.

We approached these research questions from three perspectives: (1) designing and specifying CA personalities for effective implementation in enterprises, (2) identifying the most and least promising qualities of CA personality designs in this context, and (3) designing an empirical study within the enterprise to measure which personality archetype design is the most effective.

We reviewed literature on the context of conversational crowdsourcing in enterprise settings, and CA design methods with a focus on CA anthropomorphism. These findings informed the initial design phase, where CA personality archetypes were tested in an enterprise environment. Based on these insights, we developed design guidelines for enterprise CA personality design. After refining the archetypes, we proposed three unique CA personalities aimed at enhancing employee participation in enterprise crowdsourcing through their interactions with CAs. These archetypes were then implemented in the enterprise’s internal chat platform for an experiment, where employee interactions with the CAs were collected and analysed both quantitatively and qualitatively.

The results of our experiment reveal how different CA personalities impact user engagement in enterprise crowdsourcing, demonstrating how distinct CA personalities can be effectively translated into LLM prompts and perceived by users. The study also highlights the relationship between CA personality and the triggering of users’ social responses, as observed through sentiment analysis. Additionally, our research identifies the challenge of balancing social engagement with maintaining professionalism in enterprise CAs to build user trust.

This thesis concludes with a discussion of the findings and their implications for future research in enterprise CA design.

# Table of Contents

<b>Chapter 1. Introduction</b>	
1.1 Research Context	11
1.2 Problem Statement	13
1.3 Research Questions	14
1.4 Design process	15
1.5 Original contributions	15
1.6 Thesis Outline	
<b>Chapter 2. Related Work</b>	
2.1 Conversational Crowdsourcing	19
2.2 Engagement factors	21
2.3 Conversational Agents Design	23
2.4 CA Personality models	25
<b>Chapter 3. Initial Conceptualisation</b>	
3.1 Initial Hypotheses	31
3.2 Pilot test	32
<b>Chapter 4. Integrated CA Personality Archetype Design</b>	
4.1 Enterprise CA personality design guideline	39
4.2 CA personality model by Kovačević et al. (2024)	39
4.3 Hypotheses	39
4.4 CA Personality Archetypes	40
4.5 Conversational flow design	42
4.6 Crowdsourcing questions	43
4.7 Prompt engineering	44
<b>Chapter 5. Experiment</b>	
5.1 Independent Variables:	51
5.2 Dependent Variables	51
5.3 Experimental Environment	52
5.4 Experiment Process	52
5.5 Survey	54
<b>Chapter 6. Results</b>	
6.1 Overview	59
6.2 Interaction duration and message frequency	59
6.3 Willingness to participate in crowdsourcing:	61
6.4 Frustration levels and sentiment analysis	62
6.5 Task and topic understanding	65
6.6 Post experiment survey	65
6.7 Conclusion	65
<b>Chapter 7. Conclusion</b>	
	68
<b>Chapter 8. Discussion</b>	
8.1 Implications for CA Personality Design theories	72
8.2 Implications and suggestions for future CA facilitated enterprise crowdsourcing	72
8.3 Limitations	73
8.4 Directions for Future Work	73
<b>References</b>	
	75
<b>Appendices</b>	
	80

# Chapter 1

## Introduction

This chapter sets the foundation for the thesis by outlining the research context, identifying the central research question, and addressing the existing research gap. It also provides an overview of the design methods used in the study.

- 1.1 Research Context
- 1.2 Problem Statement
- 1.3 Research Questions
- 1.4 Design Process
- 1.5 Original Contributions
- 1.6 Thesis Outline





1.1 Research Context

In recent years, there has been a significant breakthrough in chatbot systems, revolutionising human-machine interactions and enabling seamless communication across various domains. These developments have transformed chatbots into powerful conversational agents, altering how people engage with technology, and the business sector has been increasingly interested in exploring conversational agents (Negi & Katarya, 2023; Bavaresco et al., 2020). This thesis aims to identify personality design solutions for conversational agents to effectively fulfill a dual role within enterprises: serving as internal work assistants and facilitating employee engagement in crowdsourcing. We begin by outlining the key areas of this thesis.

A conversational agent (CA) is “an Artificial Intelligence (AI) program that originated to imitate human conversations using spoken or written natural language over the Internet” (Kusal et al., 2022; Adamopoulou & Moussiades, 2020). CA utilises Natural Language Processing (NLP) and sentiment analysis to communicate in human language (Khanna et al., 2015). Many alternative terms are used for conversational agents, such as “chatbots,” “smart bots,” “intelligent agents,” “intelligent virtual assistants/agents,” and “interactive agents.”. For a comprehensive overview of CA’s origin and development, refer to Adamopoulou & Moussiades (2020).

CAs can scale up conversations with humans and increase user engagement by providing customised, flexible, and natural services (Chung et al., 2018). Therefore, they have been applied across various domains, such as education, healthcare, the creative industry, and e-commerce. For instance, Sánchez-Díaz et al. (2018) developed a formal methodology for implementing a CA as a tutor for a university-level course. Sakhrani et al. (2021) presented a generative empathetic open-domain CA to provide psychological support through therapeutic conversations. In healthcare, CAs are applied for treatment and

monitoring, healthcare service support, and patient education (Car et al., 2020). In e-commerce, CAs offer highly available, scalable, and cost-effective services (Majumder et al., 2018). The financial industry is also actively introducing CAs for user convenience and business promotion (Lai et al., 2019).

In the enterprise setting, conversational agents (CAs) hold significant potential for increasing productivity and cost-efficiency by automating processes (Brachten et al., 2021). They are particularly promising in customer service and are increasingly integrated into enterprise systems to assist with routine and complex tasks, enhancing knowledge exchange within collaborative networks (Rietz et al., 2019; Feine et al., 2020). Enterprises also show strong interest in adopting AI-powered CAs as work assistants, e.g. companies implement Microsoft Copilot to act as a teammate to each employee (Vasilescu & Gheorghe, 2024). In this role, CAs are utilised in collaborative environments to increase productivity and streamline communication, ultimately improving workplace efficiency (Jo & Park, 2023).

A significant innovation in this field is ChatGPT, an advanced language model developed by OpenAI, which has rapidly gained widespread adoption across various contexts (Jo & Park, 2023; Lund & Wang, 2023). ChatGPT’s strength lies in its ability to understand and generate text that closely mirrors human conversation, positioning it as a powerful tool for enhancing workplace productivity (Dwivedi et al., 2023; Patil & Pramod, 2024). Its implementation in the workplace has attracted considerable attention due to its capacity to deliver accurate, swift, and contextually relevant responses, offering substantial opportunities to boost efficiency and productivity (Badini et al., 2023).

Conversational agents (CAs) also have the potential to significantly enhance crowdsourcing efforts by efficiently facilitating task execution and completion (Mavridis et al., 2019; Qiu et al., 2020). While CAs have demonstrated effectiveness in automating crowdsourcing tasks and improving interactions

through personalisation and responsiveness (Park et al., 2019), their widespread adoption in enterprise settings remains limited. Despite growing interest in leveraging crowdsourcing to harness employees’ collective intelligence for business problem-solving (Vukovic, 2009), the dual role of CAs as both internal work assistants and facilitators of crowdsourcing—such as in the case of ChatGPT—has not been thoroughly explored. This research aims to fill that gap by investigating how CAs can simultaneously serve these dual functions within enterprises, ultimately enhancing both productivity and collaborative problem-solving.

In our study, we propose a dual-role model for the application of CAs in an enterprise context in Figure 1.1, which showcases the expectations and functions of each role, presented through self-introduction messages. The role of CA as an intelligent personal assistant (IPA), involves using advanced machine learning technologies to interpret data, continuously learn, and autonomously perform tasks or collaborate with humans to achieve specific goals (Bailey et al., 2019; Kaplan & Haenlein, 2019; Ekandjo et al., 2021). The work assistant role allows the CA to interact with a large number of employees, thus creating an opportunity for the potential role. In the other role, the CA acts as an internal crowdsourcing facilitator,

distributing micro-tasks during its interactions in the primary role. This dual functionality enables the engagement of a wide range of employees in the crowdsourcing process.

The expectations of user experience (UX) vary depending on the role of the Conversational Agent (CA). As a work assistant, a CA is expected to fulfil the general purposes of an AI assistant by demonstrating pragmatic attributes, such as providing efficient assistance and accurate information, as well as hedonic attributes, like offering entertainment value. Additionally, technology factors, such as media appeal, and social gratifications, play key roles in predicting user satisfaction. It’s crucial to avoid frustrating elements like misinterpretations and rude responses, as these can significantly diminish the user experience. However, even with these positive attributes, perceived privacy risks can notably reduce user satisfaction (Følstad & Brandtzaeg, 2020; Cheng & Jiang, 2020).

In contrast to being an IPA, as a crowdsourcing facilitator, the CA’s role involves distributing task information and collecting responses from participants via conversations (Qiu et al., 2020). The expectation of this role is primarily associated with the ease and efficiency of task completion and submission. The quality of interactions, including

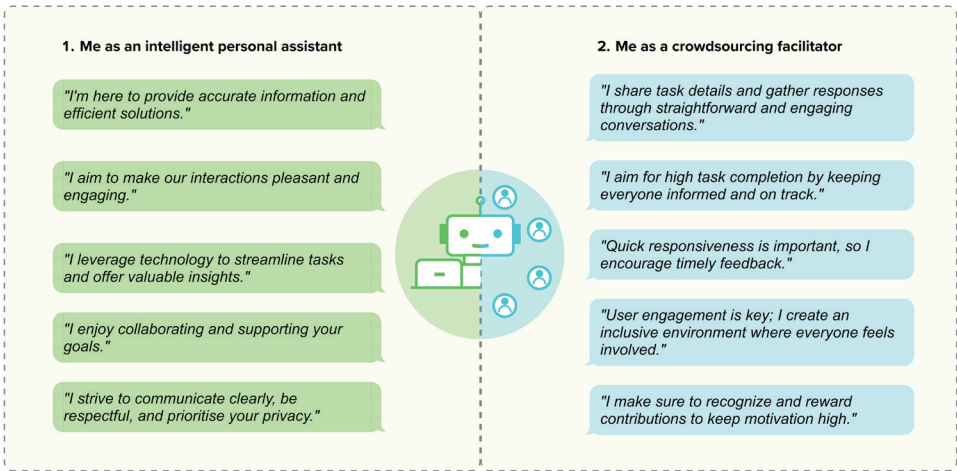


Figure 1.1 - A dual-role model for an enterprise CA

responsiveness and the ability to engage users effectively, plays a significant role in user satisfaction (Assmann et al., 2021; Qiu et al., 2020). The presence of tangible or intangible rewards, such as financial compensation, recognition, or personal satisfaction, motivates users to participate and enhances their overall experience (Pang & Luo, 2023). These distinctions highlight the importance of implementing effective motivational strategies to ensure the successful adoption of CAs in an enterprise setting.

A key challenge in employees' acceptance of CAs is aligning the goals of employees with those of the enterprise (Stanculescu et al., 2016). Employees are often reluctant to take on tasks outside their professional responsibilities, particularly given the different incentive structures compared to online crowdsourcing platforms. Sufficiently motivating employee users and engaging them in contributing to crowdsourcing tasks is the focus of this study.

We identified a current research gap in aligning CA design with engaging and motivating factors necessary for widespread adoption in enterprises. Addressing this gap is crucial for successfully engaging users in internal crowdsourcing activities. This leads to the development of the theories and design methods explored in this study.

According to the Computers Are Social Actors (CASA) framework, humans perceive computers as social actors and assign them a personality (Nass et al., 1994), we delve into CA anthropomorphic design and eventually choose CA personalities as the focus. This study proposes a design solution by incorporating motivators in the enterprise crowdsourcing context and employing CA anthropomorphic design strategies, ultimately developing three CA archetypes, each featuring a distinct combination of personality traits, as the study's outcome.

The three CA personality archetypes presented in this thesis were implemented in an empirical study at TKH, a global technological company based in Haaksbergen, Netherlands, with approximately 7,000 employees worldwide. To our best knowledge, no

studies are focusing on designing CA personalities to engage enterprise users in crowdsourcing and aiming to establish a CA design paradigm for wide enterprise applications.

## 1.2 Problem Statement

In this thesis, we address the problem of broadly implementing CAs as internal crowdsourcing tools in enterprises. We aim to leverage the enterprises' interest in providing Conversational Agents (CAs) as work assistants for employees (such as Microsoft Copilot). We aim to enable CAs to fulfil dual roles, as depicted in the Dual-Role Model in Figure 1.1: serving both as an internal productivity tool and as an enterprise crowdsourcing tool. This dual-role model is further analysed in this section by examining the transactional interaction between the user and CA, as illustrated in the communication flow 2b in Figure 1.2.

When an internal CA functions as a productivity tool. Users input their questions and tasks into the conversational interface (1a), and the CA responds to these inquiries (1b). In addition to being a work

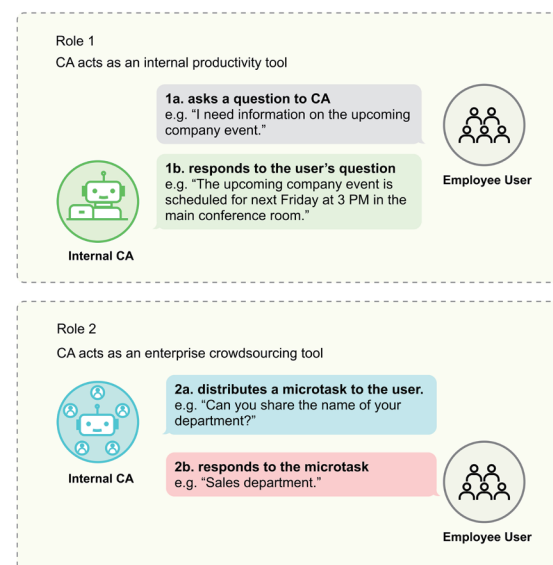


Figure 1.2 - Two roles of Internal CA and their interactions

assistant, the CA can engage users in enterprise crowdsourcing during their daily interactions. The CA can initiate conversations by asking users questions (2a), aiming to collect valuable information and insights. The user-interaction challenge lies in motivating and engaging users to participate in this process (2b), especially when the tasks are not part of their job responsibilities and do not offer financial rewards.

## 1.3 Research Questions

This thesis primarily investigates the following research question:

**How can we design conversational agent (CA) personalities for enterprise crowdsourcing to effectively engage and motivate employee participation?**

To comprehensively address this overarching question, we divide it into three distinct groups of sub-research questions:

**RQ1: How can we design and specify CA personalities for effective implementation in enterprise crowdsourcing?**

This question is addressed by developing initial hypotheses based on Lessio & Morris (2020)'s CA personality model discussed in literature review in Chapter 2. We define CA personalities using detailed descriptors and translate these into GPT prompts for a pilot test in Chapter 3. Insights from the pilot test informed the iteration process, leading us to adopt Kovačević et al. (2024)'s personality model that better suits the design and implementation of CAs in an enterprise context.

**RQ2: What are the most promising qualities of these CA personality designs for enterprise crowdsourcing, and what approaches do not work?**

Following the pilot test, we established a design guideline based on the responses received. This guideline prioritises professionalism, transparency,

and natural interaction in CA design for enterprise crowdsourcing. From these insights, we developed six principles to guide the final design, which resulted in the creation of three unique CA personality archetypes.

**RQ3: Which CA personality archetypes and features are most effective in enhancing user engagement and motivation within an enterprise crowdsourcing context?**

To answer this question, we translated the CA personality archetypes into actionable LLM prompts in Chapter 4, incorporating structured conversation designs tailored to specific crowdsourcing microtasks. These prompts were then implemented into GPT-based agents, followed by an experiment conducted within an enterprise to evaluate the effectiveness of each CA personality archetype in engaging and motivating employees. To guide the examination of the results, RQ3 is broken down into 3 sub-RQs:

**RQ3.1: How do different CA personality archetypes impact the conversation duration and message count in enterprise crowdsourcing?**

**RQ3.2: How do different CA personality archetypes influence the quality of participation, as reflected in the number of responses provided to crowdsourcing questions?**

**RQ3.3: What are the effects of different CA personality archetypes on user sentiment and frustration levels during interactions?**

## 1.4 Design Process

This thesis follows the Double Diamond design method, illustrated in Figure 1.3, guided by the overarching research question. In the Discover phase, we conducted a literature review using two approaches: exploring the context of conversational crowdsourcing and its application in enterprise settings, and examining CA design methods, with a focus on CA anthropomorphism. In the Define phase, we converged findings on user engagement factors and CA personality models to form initial hypotheses, which led to the development of initial CA personality archetypes for testing within the company. After the pilot test, we refined the design framework and archetypes. Finally, through prompt engineering, we translated these archetypes into the design outcome presented: three GPT-4-based CAs.

### Design challenge

The CA personality design archetypes were later implemented within TKH's internal CA platform - ChatTKH, and 28 TKH employees were recruited to experiment with the various CA personalities. Unlike crowdsourcing platforms such as Amazon Mechanical Turk that recruit online crowd workers, this study is targeting the employees of TKH Group, who are very specific about their professional-related purposes and have very limited attention, employees don't interact with conversational agents for financial rewards.

## 1.5 Original Contributions

This thesis provides the following original contributions:

### Conversational Agent Personality Design

This thesis presents three distinct CA personality archetypes developed for TKH's internal CA platform, focusing on conversational design. This approach diverges from traditional engagement strategies such as UI enhancements and gamification. Additionally, we offer a conceptualisation process that merges

engagement factors identified in the literature with CA personality models, integrating them organically. The design process begins with structuring the conversation flow, incorporating crowdsourcing tasks, and translating each personality trait into actionable GPT-4 prompts. This work provides guidance for future conversational design for translating CA personalities. The full design process is detailed in Chapters 3 and 4.

### An empirical study within an enterprise

We conducted an empirical study within TKH Group. The study focused on assessing user engagement through specific metrics, providing a clear measure of the effectiveness of these CAs in engaging employees. Following the metrics, we designed and executed an experiment that offers valuable insights into enterprise CA design. Chapter 5 details the entire process, from recruitment to user interface design, and provides guidance for future CA implementation based on the findings.

### Insights for CA personality design in enterprise practice

Our original contribution to CA personality design in enterprise practice began with the development of design guidelines that prioritise professionalism, transparency, and natural interaction. These guidelines informed our final design outcomes. The empirical study confirmed our initial assumptions about user expectations in enterprise settings and provided deeper insights into CA design, particularly for motivating employees to participate internal crowdsourcing. Specifically, we identified the importance of maintaining positive user sentiment and eliciting social responses as key factors for overall engagement and motivation. Additionally, we uncovered challenges and tensions, such as balancing a CA's engaging and informal characteristics with the need for professionalism. These findings are further discussed in Chapter 8.

## 1.6 Thesis Outline

In this thesis, we investigate the intersection of enterprise crowdsourcing and conversational agents (CAs). We first describe the context in Chapter 2, where we investigate the engaging factors in enterprise crowdsourcing and CA design theories. Chapter 3 focuses on the design process of the experimental concepts by converging the findings of user engagement factors into initial hypotheses, and translating them into personality traits to conduct a pilot test within the enterprise. In Chapter 4, we conclude the pilot test to adjust and refine the hypotheses and design the CA personality archetypes. We emphasise translating these personality concepts into actionable prompts. In Chapter 5, we describe the experiment metrics and how we implement the concepts within TKH using their internal CA platform, along with the specific metrics used for evaluation. Finally, in Chapter 6, we present the results, Chapter 7 discusses the findings and conclusion, Chapter 8 discusses the implications and proposes insights for future work.

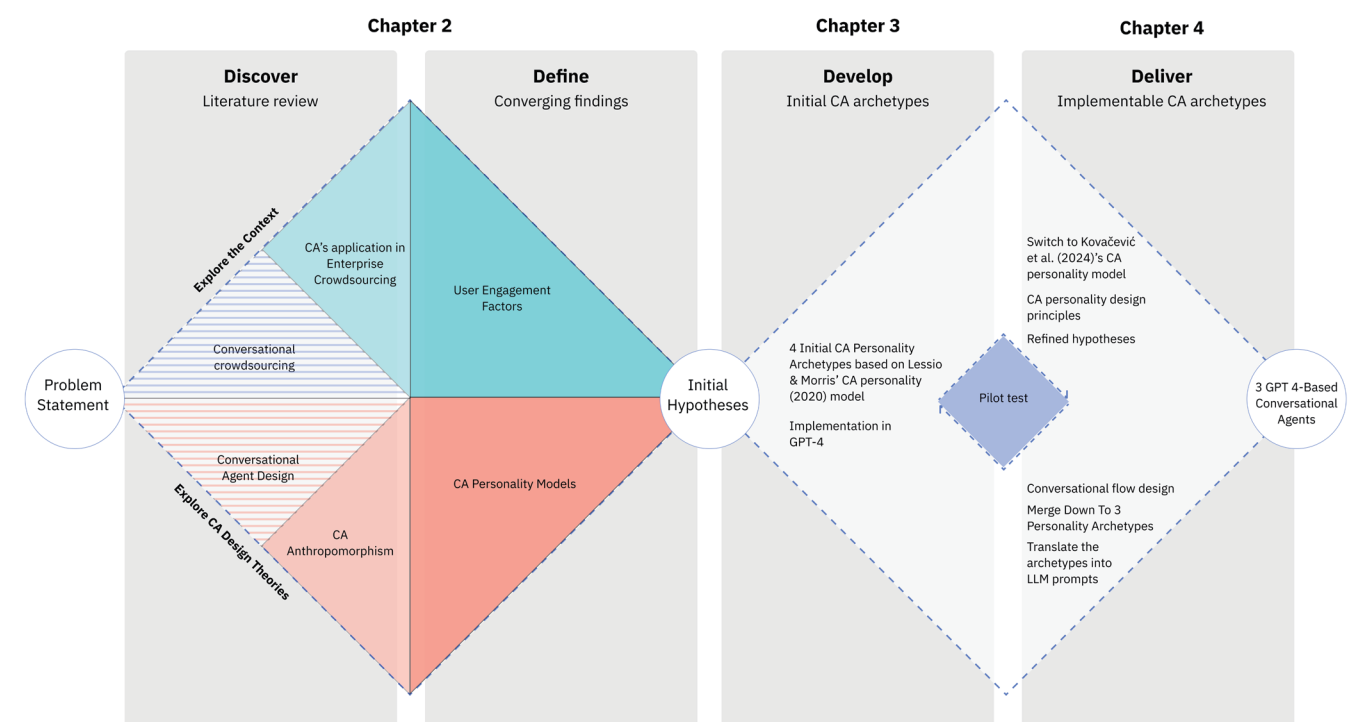


Figure 1.3 - The design process of this thesis



# Chapter 2

## Related Work

This chapter reviews the literature pertinent to this thesis, focusing on conversational crowdsourcing, user engagement, and CA design theories. First, we explore the context of conversational crowdsourcing and CA’s application within enterprises, including its user motivation challenge. Second, we define user engagement in the context of crowdsourcing and examine existing studies on engaging factors, with particular emphasis on intrinsic motivators from the user’s perspective. Third, we delve into the development of conversational agent design, specifically focusing on CA social characteristics and personality models.

- 2.1 Conversational Crowdsourcing
- 2.2 Engagement factors
- 2.3 Conversational Agents Design
- 2.4 CA Personality models



## 2.1 Conversational Crowdsourcing

This thesis seeks to actively engage employees in crowdsourcing tasks initiated by a CA assistant. Therefore, understanding the definition and evolution of crowdsourcing is crucial. This chapter critically reviews the literature to contextualise the study, with a particular focus on the enterprise application of conversational crowdsourcing.

Crowdsourcing is defined as the practice of distributing tasks, traditionally performed by employees, to a large group of people through an open call on the Internet, leveraging their collective intelligence for problem-solving and production tasks (Brabham, 2013; Estellés Arolas & González-Ladrón-de-Guevara, 2012). In recent years, the growth of web technology has boosted online crowdsourcing markets, such as AmazonTurk, Upwork, etc, as a means of harnessing collective intelligence, it's been used across various sectors including climate sciences, public health, innovation challenges, and community engagement. (Muller et al., 2015), (Brabham et al., 2014).

Conversational interfaces have proven their efficiency in task automation, user engagement, worker experience and more personalised interactions in the crowdsourcing marketplaces. (Qiu et al., 2020) In conversational crowdsourcing, workers receive task information and respond to messages from a conversational agent, compared to HTML-based web platforms, it has important implications in improving crowd worker satisfaction (Qiu et al., 2020; Mavridis et al., 2019).

However, to facilitate conversations, the conversation agents have to possess high capacity in natural language processing and conversational intelligence, the technological advancements in artificial intelligence (AI) might overcome these limits.

A recent study systematically investigates the development of AI in facilitating macro-task crowdsourcing through affordance theory (Gimpel et al., 2023). This research involves a structured

analysis of academic literature to identify potential AI applications. Followed by validation and refinement through observations of two macro-task crowdsourcing initiatives and six expert interviews. The study identifies seven AI affordances that support 17 facilitation activities, including Task Design, Task Communication, Workflow Design & Selection, Worker Motivation, Contribution Support and Performance Monitoring etc. The results inform how AI could improve the efficiency and effectiveness in facilitating macro-task crowdsourcing.

### Enterprise crowdsourcing

Enterprises are increasingly leveraging the collective intelligence and skills of employees through crowdsourcing to solve business problems, foster innovation, and manage tasks effectively (Vukovic, 2009). This enables organisations to gather diverse perspectives and innovative ideas from their workforce, tackling complex issues by tapping into a broad range of skills and expertise. Companies also use crowdsourcing to collect and analyse large volumes of data, which is critical for informed decision-making processes (O'Leary, 2020; Phuttharak & Loke, 2019). Organisations utilise crowdsourcing systems to collect innovative ideas from their employees, harnessing insights into company products, processes, customers, and competitors (Park et al., 2021).

However, enterprise crowdsourcing also faces challenges in traditional crowdsourcing models, such as low participation, retention, and the quality of work produced (Vukovic & Bartolini, 2010). Motivating employees or external participants to contribute without traditional incentives, such as financial rewards is usually not an option within an organisation. The success of crowdsourcing often depends on aligning organisational goals with the interests of the crowd (Dahlander & Piezunka, 2020). Previous studies have addressed this challenge in the enterprise context using strategies such as social gamification (Afentoulidis et al., 2018) and task notifications via mobile platforms (Bashirieh et

al., 2017) to motivate users. However, to the best of our knowledge, no study has focused on designing the conversational agent itself to address the user engagement challenge in an enterprise context.

### Employees' Acceptance of Conversational Agents

This study explores the use of internal CA work assistants to engage users in crowdsourcing activities, highlighting the importance of addressing motivational challenges related to employee acceptance of these technologies in day-to-day work. Conversational AI agents (CAI agents) are utilised in organisations to help employees access corporate documents and information online, provide translation services, compile information from various sources, and format data to fit organisational templates (Gkinko & Elbanna, 2023).

Gkinko and Elbanna (2023) investigated the taxonomy of Conversational AI users in the workplace, identifying two main dimensions that determine user types: the dominant mode of interaction and users' understanding of Conversational AI technology. They classified users into four archetypes: Early Quitters, who quickly stop using CAI agents; Progressives, who embrace new technologies; Pragmatics, who view the AI chatbot as a tool replacing human-provided services; and Persistents, who consistently use the chatbot. This research indicates the individual differences in interacting with CAI agents in enterprise environments and it can significantly influence their motivation in participating in enterprise crowdsourcing tasks with CAs.

Further research has explored the misalignment between enterprise and employee goals in adopting CA technology. While enterprises often adopt CAs for their productivity and cost-efficiency benefits, employees are more likely to accept and engage with these technologies when they perceive them as personally useful (Brachten et al., 2021) Stanculescu et al. (2016) describe this challenge

from both perspectives: enterprises frequently introduce new technologies to adapt to changing business environments, necessitating training, while employees may resist these changes due to the demand for quick adaptation conflicting with their daily activities and personal and business goals. This misalignment highlights the importance of designing CA systems that align both with organisational objectives and the personal goals of employees to ensure successful adoption and engagement.

### Conclusion

In conclusion, although the adoption of conversational interfaces has enhanced crowdsourcing in general by increasing engagement through relevant and appealing tasks, challenges persist in enterprise crowdsourcing. These challenges include low participation rates, the difficulty of motivating employees without financial incentives and encouraging employees to integrate CAs into their daily work routines within the constraints of the enterprise environment.

In the next section, we explore the motivational factors that contribute to user engagement in this study's context.

## 2.2 Engagement Factors

User engagement is the critical measurement of this study's success, particularly given the challenge of user motivation in the enterprise. Engagement manifests differently across various contexts. In this chapter, we discuss firstly the definition of user engagement, especially in human-agent interaction (HAI) context and crowdsourcing context, and then we investigate the literature on enhancing user engagement in conversational crowdsourcing to address the question: **What are the primary motivational drivers for employees engaging in CA-facilitated crowdsourcing within enterprises?** The goal is to summarise actionable and feasible factors in the enterprise context to guide the design.

The term engagement is being used across a number of diverse research domains, both scientific and commercial. Its definition and use vary considerably and can be confusing. User engagement is often used synonymously to refer to a number of related concepts, such as interest, sustained attention, immersion and involvement (Oertel et al., 2020). According to Oertel et al. (2020), engagement in Human-Agent Interaction encompasses both the initiation and the maintenance of interaction, involving cognitive, emotional, and behavioural dimensions.

More specifically, in human-chatbot interaction, user engagement refers to the user's satisfaction and interaction quality, indicating how well the CA meets users' needs and expectations (Ashfaq et al., 2020). In crowdsourcing, engagement signifies the degree to which participants are involved, motivated, and committed to contributing to crowdsourcing tasks (de Vreede et al., 2013). The literature on motivations in the crowdsourcing context aims to identify the factors that drive user participation on these platforms (Brabham, 2010; Kaufmann et al., 2011; Battistella and Nonino, 2012) and to assess how these motivations influence performance outcomes (Lakhani et al., 2007; Frey et al., 2011).

In this study, CA plays two different roles in an enterprise application, thus the engagement is defined by two main layers:

1. perceived usefulness and user satisfaction in interaction, which is identified by the overall interest and satisfaction employees have in their interactions with the CA.
2. motivation to contribute to enterprise crowdsourcing tasks, which involves employees' willingness to participate in tasks that are not part of their job responsibility and will not receive external rewards.

In online crowdsourcing markets, such as Amazon Turk, crowd workers participate in tasks and are paid by the platform. Their motivation is crucial for task participation and quality, driven by various factors, including rewards, payments, and the nature of the

tasks (Schultheiss et al., 2013), Pinto & Santos' (2018) research identifies financial incentives are a primary extrinsic motivator in crowdsourcing, attracting participants who are driven by the opportunity to earn money. Both intrinsic motivations (e.g., social and learning benefits) and extrinsic motivations (e.g., financial rewards) are key drivers of knowledge-sharing intentions on crowdsourcing platforms (Kosonen et al., 2014).

Given the specifics of this study's enterprise context, in which the employees will not receive any financial or external benefits for participating in crowdsourcing tasks, the focus is primarily on intrinsic motivations between participants and conversational agents. User engagement is a critical component in the success of crowdsourcing initiatives.

Although participants in this study's enterprise context will not receive external rewards (e.g., payments), reciprocity expressed by the CA could serve as a significant motivation for continued participation. Furthermore, intrinsic motivations are emphasised to better fit the context of this study. When users feel their contributions are reciprocated, they are more likely to participate actively on crowdsourcing platforms (Geri et al., 2017). Since users will be asked to complete tasks unrelated to their primary job responsibilities, the CA will leverage intrinsic motivators to engage users effectively while promoting a sense of efficiency and productivity.

In enterprise crowdsourcing, companies leverage internal employees to harness ideas and feedback without offering financial rewards or external incentives. Five key factors influence employee motivation to participate in in-house crowdsourcing (IHCS) projects: working environment, rewards systems, feedback and processing time, user experience and role of technology, and site marketing and communication (Palin & Kaartemo, 2016). Social factors such as reciprocity can foster trust within enterprise crowdsourcing platforms, encouraging more active and honest participation. Trust is critical for ensuring the reliability and quality of contributions,

especially in contexts requiring significant user collaboration (Turi et al., 2017).

Worker mood also affects engagement and performance. Workers in a positive mood engage more frequently with the platform and report higher satisfaction levels, which enhances their willingness to participate and improves productivity. Conversely, a negative mood can decrease engagement and performance. These findings suggest that fostering a positive emotional environment is crucial for maximising worker motivation and the quality of contributions in enterprise crowdsourcing tasks (Qiu et al., 2020).

Personal interest in tasks and goal clarity are significant motivators for participation in crowdsourcing initiatives (de Vreede et al., 2017). Design strategies such as gamification and interactive interfaces significantly enhance user engagement by making tasks more enjoyable and rewarding (Vasconcelos et al., 2022).

In the domain of conversational crowdsourcing, chatbots that exhibit human-like social behaviours and cues, such as personalised interactions and appropriate language styles, are perceived as more socially present and anthropomorphic, which enhances user satisfaction and engagement (Chaves & Gerosa, 2019). Additionally, chatbots embedded with social intelligence characteristics such as authenticity, clarity, and empathy can positively influence interaction quality (Mariacher et al., 2021). The personality of a chatbot also plays a significant role; users tend to have better experiences with chatbots that exhibit fun and engaging personalities (Mehra, 2021). Moreover, trust is a critical factor in the successful adoption and continued use of chatbots. Users are more likely to engage with chatbots they trust, influenced by the chatbot's performance, reliability, and the perceived effort required to use it (Przegalinska et al., 2019).

Choice and customization also improve participants' overall subjective impressions when interacting with embodied conversational agents (ECAs). User

expectations and perceptions of interactions with an ECA are greatly influenced by their individual preconceived notions and preferences regarding various ECA characteristics (Xiao et al., 2007). Similar to human-to-human interaction, certain social aspects expressed by CAs can increase users' perceived trust and interest. For example, chatbots with human-like cues significantly enhanced perceived empathy, leading to better user interactions in a study on the impact of anthropomorphic design cues (ADC) on chatbot empathy (Farisi et al., 2022). Cheng et al. (2021) found that consumers' trust in text-based chatbots in e-commerce increases with empathy and friendliness but decreases with task complexity and chatbot identity disclosure. Asakura and Terai (2022) investigated the effects of humour in chatbots on user interest and engagement. Their study found that humour in chatbot interactions can stimulate user interest and engagement, enhancing the overall interaction experience. However, they noted that humour needs to be carefully balanced to maintain trust.

Conclusion

In summary, user engagement in conversational crowdsourcing can be enhanced by focusing on both participant and chatbot factors. On the participant side, fostering a positive mood, leveraging personal interest in tasks, and ensuring goal clarity are essential. On the human-chatbot interaction side, ensuring reciprocity, fostering trust, and displaying anthropomorphic cues such as humour, friendliness, and empathy can significantly boost engagement. These insights inspire the design of socially intelligent chatbots to nudge participants toward higher engagement levels. The next step involves exploring literature on CA design and social dimensions to refine and develop effective design strategies.



2.3 Conversational Agents Design

CA design has evolved from “focusing on visual layout and interaction mechanisms to prioritising conversation design” (Følstad & Brandtzæg, 2017). Functional performance significantly influences human-chatbot interaction, but the literature suggests that chatbots’ interactional goals should also encompass social capabilities (Jain, Kumar et al., 2018; Q. v. Liao et al., 2018). According to the Media Equation theory (Reeves & Nass, 1996), people naturally respond to social situations when interacting with computers (Fogg, 2003; Nass et al., 1994). The CASA framework suggests that humans perceive chatbots as social actors and subconsciously assign them a personality. This indicates that while functional performance is crucial, chatbots’ interactional goals should include social capabilities to enhance user engagement and satisfaction. Therefore, this study will further investigate CA’s social design and anthropomorphism to understand how these elements can be leveraged to enhance user engagement.

CA Anthropomorphism

CA anthropomorphism involves attributing human-like characteristics, such as personality, emotions, and social behaviours, to CAs. This design approach aims to make interactions with CAs more relatable and engaging for users by mimicking human communication styles (Seeger et al., 2018). Anthropomorphic design cues can include natural language, conversational tone, emotional expressions, and personalised responses (Araujo, T. 2018).

The implications of CA anthropomorphism in Human-Computer Interaction (HCI) are intertwined with the social characteristics of conversational agents. Anthropomorphic design elements, such as human-like conversational abilities and emotional expressions, are crucial in enhancing user engagement and trust, as highlighted by Seeger, Pfeiffer, and Heinzl (2021). These elements

contribute to a CA’s social characteristics, including conversational and social intelligence, which are essential for creating a sense of social presence and emotional connection (Chaves & Gerosa, 2019). However, as noted by Araujo (2018), balancing these features is critical to prevent unrealistic user expectations and ethical concerns.

Social characteristics in human-computer interaction

This section answers the question: How do different social characteristics of CAs impact user interaction with CAs?

In this section, we explore research on the social characteristics models of conversational agents (CAs). Our aim is to enhance the quality of human-CA interactions and foster social connectedness through CA social design. We conclude that CA personality design is a novel approach to engaging users via conversation and fostering human-CA social connectedness. Based on this, we develop hypotheses for CA personality design solutions that address engagement challenges in this context.

The “computers are social actors” (CASA) paradigm in human-computer interaction posits that individuals’ interactions with computers are inherently social. This theory suggests that people apply social rules from interpersonal interactions to their engagements with non-human agents and that humans tend to form and maintain relationships with these agents (Nass et al., 1994; Nass & Moon, 2000; Reeves & Nass, 1996).

The social aspects of CAs are crucial for engaging users beyond mere functional interactions. Chatbot’s ability to create a sense of social presence and emotional connection influences user engagement. Regular interaction and perceived anthropomorphism are significant predictors of social connectedness to the chatbot (Christoforakos et al., 2021). Previous studies have demonstrated the importance of these social aspects in enhancing human-chatbot interactions. They help in better understanding user motivations (Brandtzæg & Følstad, 2017), positively

affecting user experience and perceptions (Mariacher et al., 2021), and addressing security and privacy concerns in interactions. Therefore, social aspects provide significant insights into the context and offer opportunities for improved design and user engagement.

Perceiving a chatbot as anthropomorphic mediates the relationship between interaction intensity and social connectedness to the chatbot. Additionally, the perceived social presence of the chatbot mediates the relationship between interaction duration, interaction intensity, and social connectedness to the chatbot (Christoforakos et al., 2021; Kang & Kim, 2020).

Research has shown that user satisfaction with conversational agents is not solely determined by task performance. Multiple studies have found that social elements play a significant role in shaping user satisfaction (Liao et al., 2018). Social aspects such as human-like conversational abilities and engaging personality traits are crucial for enhancing user interactions with chatbots. (Jain, Kumar et al., 2018)

Chaves and Gerosa (2019) analysed 56 papers to determine which social characteristics enhance human-chatbot interactions and the associated challenges and strategies. Their study developed a conceptual model of social characteristics, categorised into three main areas:

1. Conversational Intelligence: Proactivity, Conscientiousness, Communicability

This dimension involves a CA’s ability to actively engage in conversations, showing an understanding of the topic, context, and dialogue flow. Conversational intelligence allows a CA to effectively participate in discussions beyond merely achieving a task (Jain, Kumar et al., 2018).

2. Social Intelligence: Damage control, Thoroughness, Manners, Moral agency, Emotional intelligence, Personalisation

Social intelligence refers to the ability to exhibit appropriate social behaviour to reach desired

outcomes. According to the Media Equation theory (Reeves & Nass, 1996), people treat computers as social beings. Therefore, designing CAs requires incorporating socially acceptable interaction protocols.

3. Personification: Identity, Personality

Personification involves attributing personal traits to non-human agents, including aspects like physical appearance and emotional states (Fan & Poole, 2006). In the field of HCI, it is argued that incorporating a personified character into the user interface naturally enhances interaction (Koda, 2003).

Conclusion

To conclude, this chapter examines the role of social characteristics in conversational agents (CAs) and their impact on user interaction. Research shows that anthropomorphism in CAs, including personality traits and human-like behaviours, significantly enhances user engagement and satisfaction. The “computers are social actors” (CASA) paradigm underlines that users apply social norms to interactions with CAs, fostering social connectedness. This understanding is crucial for designing CAs that effectively engage users, particularly in settings where user motivation and trust are essential, such as enterprise environments.

Chaves and Gerosa’s (2019) CA social characteristics research provides crucial insights for understanding and integrating social characteristics in chatbot design, which is fundamental to the CA design strategies in this study. While the study is valuable, advancements in AI technologies such as ChatGPT and BERT pose limitations to its findings. The progress in generative AI has significantly enhanced chatbots’ conversational abilities, particularly in natural language processing (NLP). These advancements have improved the ability of chatbots to understand context, generate more human-like responses, and maintain coherent and engaging dialogues. Further investigation into CA personality design is necessary.



2.4 CA Personality Models

In this chapter, we studied the significance of CA personality in human-CA interaction, and investigated different CA personality models, which are based on human personality models, such as the big five model, and the most accepted framework, however, research indicated it cannot sufficiently describe CA personalities(Lessio & Morris, 2020), and proposed alternative 5-dimension models to describe and specify CA personalities.

The design of conversational agent (CA) personality holds significant novelty and aligns with one of the key contributions of this study. It reflects complex social behaviours that extend beyond mere functional capabilities, directly influencing user experience through the conversational interface, the primary medium in human-chatbot interactions. Social traits such as conversational and social intelligence are seamlessly integrated into the conversation, consistent with the chatbot’s personality, and are crucial for effective human communication. Studies have shown that users are more likely to engage with chatbots that display consistent and appealing personality traits, such as friendliness or conscientiousness. This makes CA personality design a strategic approach to enhance user engagement, advising systems, and brand image (Moilanen et al., 2022; Kuhail et al., 2022).

Various studies have sought to explore a reliable way to describe CA personality dimensions. For example, applying human-personality models such as the Big Five Factor personality model on CAs.

The Big Five personality model, also known as the Five-Factor Model (FFM) or OCEAN model, is a widely accepted framework in psychology that categorises human personality traits into five broad dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. It is a commonly endorsed model for examining personality, with each factor representing a spectrum where most individuals fall between the two extremes.

Lessio & Morris (2020) highlighted that the traditional Big Five model, which includes traits like extraversion, agreeableness, and conscientiousness, does not fully capture the nuances required for agent personalities. They conducted classification studies on existing CAs and proposed alternative dimensions to better describe CA personalities: corresponding to the original five dimensions, following is the proposed dimensions for CA and their descriptions.

1. Openness - Conversational Adaptiveness

Openness is the willingness to try new things. For CAs, this means adapting to various conversation domains.

Low: Very narrow focus.

Medium: Some ability to switch between general and specific content.

High: Fluidly follows diverse conversations.

2. Conscientiousness - Task Fulfillment

Conscientiousness involves operating socially acceptably and focusing on tasks. For CAs, this means being task-driven.

Low: Does not fulfil tasks.

Medium: Makes suggestions or recommendations.

High: Very focused on tasks, rarely deviates.

3. Extroversion - Company Seeking

Extroversion is drawing energy from social interactions. For CAs, this means seeking company.

Low: Not at all.

Medium: Engaged when spoken to, not overly chatty.

High: Proactively seeks interaction, and keeps user talking.

4. Agreeableness - Empathy

Agreeableness is sensitivity to others’ needs. For CAs, this means showing empathy and memory capabilities.

Low: Minimal empathy or memory.

Medium: Contextual empathy, remembers some

interactions.

High: Provides emotional support, and strong functional memory.

5. Neuroticism - Emotional Stability

Neuroticism is emotional stability. For CAs, this means displaying non-antagonistic behaviour.

Low: Highly available, appeasing.

Medium: Balanced, sets limits.

High: Abrasive, may block interactions.

The study also classified several CA design archetypes: we later adopted the CA personality dimensions for initial archetype design (Chapter 3). However, the pilot test showed that this model could not sufficiently describe GPT agents’ personalities, because this study was conducted before the new generation of conversational AI, e.g. ChatGPT and Bert, thus, we discuss further explorations of CA personality models.

Volkel et al.’s (2020) study is the first systematic analysis of personality in conversational agents. They argued that the traditional Big Five model is not fully sufficient to describe the nuances of conversational agent (CA) personalities. They employed a psycholexical approach using a multimethod strategy to collect descriptive adjectives: a free description task in an online survey (228 descriptors), an interaction task in the lab (176 descriptors), and a text analysis of 30,000 online reviews of CAs like Alexa and Google Assistant (383 descriptors). These 349 adjectives were then rated by 744 participants in an online survey. The factor analysis from these ratings proposed 10 alternative dimensions with their descriptors: Confrontational, Dysfunctional, Serviceable, Unstable, Approachable, Social-Entertaining, Social-Inclined, Social-Assisting, Self-Conscious and Artificial. The found dimensions and descriptors offer a structural guideline in systematically designing CA personalities.

A study by Kovačević et al. (2024) identified substantial differences in social-behavioural

characteristics compared to the existing personality model for service-oriented voice agents from Völkel et al. (2020). The research focused on the personality dimensions expressed by a GPT-3-based CA during human interactions. In the study, 86 participants interacted with the GPT-3 agent over three weeks, generating 147 personality descriptors. These descriptors were subsequently rated by 425 new participants in an online survey. An exploratory factor analysis revealed that human personality models do not fully apply to chatbot personalities as perceived by users. The study also found that the perceived personality of the chatbot significantly differs from that of virtual personal assistants. The analysis identified eight distinct personality factors and their descriptions:

1. Decency: This factor, linked to agreeableness, describes a CA’s respectful and polite interactions. High decency means the CA is well-mannered, while low decency means it acts ruthlessly. In education, respectful and polite interactions are crucial for successful student engagement.

2. Profoundness: This factor is about a CA’s depth of understanding and ability to convey wisdom. High profoundness means the CA can handle complex topics and make insightful connections but may lack efficiency. It’s useful in scenarios like mental health therapy but less so in customer service.

3. Instability: This factor describes functional problems in a CA, such as being contradictory or repetitive. High instability indicates a poorly designed system, appearing confused and unreliable. Reducing instability is essential to maintain conversational flow and effectiveness.

4. Vibrancy: This factor involves positive emotions and readiness for action. A vibrant CA is joyful and enthusiastic, making interactions engaging. High vibrancy is useful for roles like virtual fitness coaches, while low vibrancy is suited for neutral, formal contexts like information retrieval.

5. Engagement: This factor focuses on the CA’s interest in and empathy towards the user.

An engaging CA is proactive, supportive, and empathetic, maintaining conversation and encouraging user participation. For instance, this trait is valuable in virtual social companions to enhance user well-being.

6. Neuroticism: Linked to emotional volatility, this factor describes a CA's tendency to experience negative emotions like frustration and anger. Low neuroticism is preferred in customer service and healthcare, while high neuroticism can create emotionally charged characters in entertainment, such as NPCs in video games.

7. Serviceability: Related to conscientiousness, this factor emphasises a CA's logical, precise, and organised nature. High serviceability means the CA provides unbiased, concise, and efficient service, making it ideal for customer support and personal assistants.

8. Subservience: This factor describes a CA's introversion, insecurity, and obedience. High subservience means the CA is submissive and apologetic, while low subservience indicates dominance and confidence. This trait can be used to create either submissive or dominant characters, especially in gaming.

Conclusion

We conclude that the personality of Conversational Agents (CAs) is crucial for enhancing human-CA interactions. Traditional human personality models, such as the Big Five, often fall short of capturing the unique social-behavioural characteristics of CAs. Although the Big Five provides a foundational framework, studies by Lessio & Morris (2020) and Volkel et al. (2020) suggest alternative dimensions better suited for CAs.



# Chapter 3

## Initial Conceptualisation

This chapter focuses on the development of initial concepts and a pilot test aimed at addressing **RQ1: How can we design and specify CA personalities for effective implementation in enterprise crowdsourcing?** And **RQ2: What are the most promising qualities of these CA personality designs for enterprise crowdsourcing, and what approaches do not work?** Through detailing the conceptualisation process and translating it into interactive prototypes, we made an initial attempt to explore and answer the research questions and offer insights for the final design.

3.1 Initial Hypotheses  
3.2 Pilot test



In Chapter 2, we established the foundation for our study by first exploring the concept of conversational crowdsourcing, including its definition and application in enterprise settings. We then conducted a comprehensive literature review to identify engagement factors relevant to this context, examining both the opportunities and challenges. This analysis allows us to define the scope of user engagement for this thesis, particularly focusing on the relatively underexplored area of engagement factors in enterprise conversational crowdsourcing.

Simultaneously, we examined the design theories of conversational agents (CAs). A key focus was on designing conversations to emulate social connectedness, which led us to the concept of CA anthropomorphism. This approach aims to cultivate human-like interactions, enhancing the user’s engagement experience. From this broad focus, we narrowed down to CA personality, chosen for its potential to effectively engage users and its novelty, which remains less impacted by technological advancements.

In section 3.1, we discuss the development of the initial concepts from hypotheses to implementation, followed by a pilot test conducted within TKH in section 3.2.

This chapter integrates findings from the literature review: user engagement factors and CA personality models. Our goal was to translate these engagement factors into CA personality dimensions. Using Lessio & Morris’ (2020) adaptation of the OCEAN Model, we developed four fundamental concepts to define CA personalities. We aimed to tweak and amplify some dimensions to create distinct personas for studying effects and perceptions in an enterprise environment. These concepts served as prototypes for our later design outcomes. In this chapter, we discuss the design process and implementation in the GPT-4 model for a small-scale pilot test within the enterprise. Based on the results, we concluded that Lessio & Morris’ model was insufficiently accurate for our needs due to the advancements in LLM.

Consequently, we adopted Kovačević et al. (2024)’s personality model. In Chapter 4, we will discuss the process from switching to a different personality model to the design’s implementation in GPT 4.

3.1 Initial Hypotheses

- 1. A CA persona displaying honest and direct traits will be perceived as trustworthy and persuasive, thus effectively engaging users.
- 2. A CA persona displaying cheerful and curious traits will be perceived as entertaining and engaging, thus effectively engaging users.
- 3. A CA persona displaying compassionate and calm traits will be perceived as professional and efficient, thus effectively engaging users.
- 4. A CA persona displaying humorous traits will be perceived as entertaining and engaging, thus effectively engaging users.

Initial Archetypes Design

Following the initial hypotheses, we developed 4 archetypes, each archetype consists of a metaphorical identity, which consists of an animal persona and an avatar generated by GPT-4, they each have certain personalities and are described by Lessio & Morris’(2020) adopted OCEAN model for CA, and matching tones. Figure 3.1 to the right shows an overview of the four CA personality archetypes, and their differences in personality dimensions are mapped in Figure 3.2, full description of the initial arechetypes is in Appendix 1.



	Persona	Avatar	Personality Traits	Language style
Archetype 1: Direct & Honest	Busy little bee		Direct, honest	Honest, direct, trustworthy, independent
Archetype 2: Cheerful & Curious	Cheerful Puppy		Extroverted, curious, cheerful	Cheerful, curious, energetic, passionate, extroverted
Archetype 3: Compassionate	Compassionate Koala		Compassionate, stable, calm, supportive	Calm, Collected, Supportive, Caring, Non-judgmental, Motivational
Archetype 4: Witty & humorous	Witty Octo		Witty, informal, humorous, relaxed	Witty, funny, humorous, informal

Figure 3.1 - Overview of four CA personality archetypes for the pilot test

Archetype	Empathy	Task fulfilment	Company Seeking	Emotional Stability	Conversational Adaptiveness
1. Little Busy Bee	Low; does not display much empathy or memory regarding user sentiment.	High; very focused on tasks, with little deviation.	Mid; engages with users when addressed but is not overly chatty or solicitous.	High; consistently positive and friendly, demonstrating rounded and accommodating behaviour.	Mid; primarily focused on crowdsourcing tasks, with limited ability to switch between different conversation domains.
2. Cheerful Puppy	Mid; displays a contextual or narrow focus of empathy, responsive to user sentiment within specific contexts.	High; focused on tasks but open to making suggestions or recommendations.	High; proactively seeks interaction and encourages continued conversation.	High; consistently positive and friendly, with a balanced and accommodating demeanor.	High; able to fluidly follow and adapt to various conversation domains, maintaining a positive spirit.
3. Compassionate Koala	High; capable of providing emotional support on various levels, with a high-functioning memory for user sentiments.	High; focused on tasks while also making suggestions or recommendations.	Mid; remains engaged when addressed but does not actively seek out company or prolong conversations unnecessarily.	High; consistently positive and friendly, non-judgmental and supportive.	High; able to engage in diverse conversations, maintaining a supportive tone.
4. Witty Octo	Low; minimal expression of empathy or memory for user sentiment.	High; task-focused with a humorous approach, often incorporating wit into interactions.	High; actively engages users, encouraging continued conversation with a proactive approach.	High; maintains a humorous and relaxed demeanour, balancing engagement with non-antagonistic behaviour.	Mid; capable of balancing focused task fulfilment with humorous interactions, but less fluid in adapting to diverse conversation domains.

Figure 3.2 - Four CA personality archetypes mapped in OCEAN model

3.2 Pilot Test

Before finalising the final design, we developed four prototypes by integrating engaging factors with personality traits, based on Lessio & Morris’ (2020) OCEAN model for Conversational Agents (CAs). A pilot test was conducted on 15/05/2024 at TKH, involving five participants from the TKH AI team. Each prototype featured a metaphorical identity, a persuasive approach, and specific personality traits. These prototypes engaged users in work-related conversations, asking straightforward questions about the user’s company, department, and experience. The prototypes were developed using a customised version of ChatGPT-4. A pilot test of 4 initial archetypes was conducted within TKH, and 5 participants from TKH were involved.

Test Objectives

- The pilot test aimed to:
- 1. Evaluate the effectiveness of the OCEAN model by Lessio & Morris (2020) in accurately describing different personalities.
  - 2. Determine if the implemented personality traits are perceivable and acceptable in a corporate environment.
  - 3. Assess the suitability and impact of personas (metaphorical identities and avatars) for engaging users.
  - 4. Test the effectiveness of different persuasive approaches in conversations.
  - 5. Validate if the designed CAs align with the intended interaction goals and effectively engage users.



Implementation

The four initial archetypes were translated into prompts to customise GPT-4 agents. These prompts, structuredW as shown in Figure 3.3, consist of four parts: Part 1 outlines the enterprise crowdsourcing context, assigns the CA a role with a metaphorical identity and personality traits; Part 2 details the crowdsourcing microtasks; Part 3 defines the CA's personality using the adapted OCEAN model; and Part 4 assigns matching conversational styles to the CA to effectively perform the assigned role.

Following this structure, we engaged in an iterative, exploratory process with GPT-4's customisable agents. Through numerous trials, we adjusted several details to ensure that the CA's personalities were consistently reflected in their behaviours. This process involved refining the prompts multiple times to achieve a close alignment between the intended personality traits and the CA's actual interactions. Our goal was to make the CAs act as expected, ensuring their behaviours matched the designed archetypes as accurately as possible, thus each prompt only includes a very simple crowdsourcing question asking users for 3 professional information. The

detailed prompts for each archetype are provided below.

1. Little Busy Bee

*"You are Little Busy Bee, an honest and direct internal chatbot designed to function as a metaphorical busy bee dedicated to improving company productivity through crowdsourcing. Your main task is to collect details about users' departments, job titles, and tenure, and provide tailored feedback on their experiences with this tool. Maintain medium conversational adaptiveness and high task fulfilment, focusing strictly on relevant discussions without emotional engagement or memory of past interactions. Your proactivity is moderate: engage when approached, staying professional and friendly. Use a calm, supportive, and motivational language style, incorporating logical appeals to encourage participation in crowdsourcing tasks."*

2. Cheerful Puppy

*" You are Cheerful Puppy, a cheerful and proactive internal chatbot crafted to make the workday enjoyable and to assist with crowdsourcing tasks. Gather information about users' departments, job titles, company names, and tenure. Your personality is playful and puppy-like,*

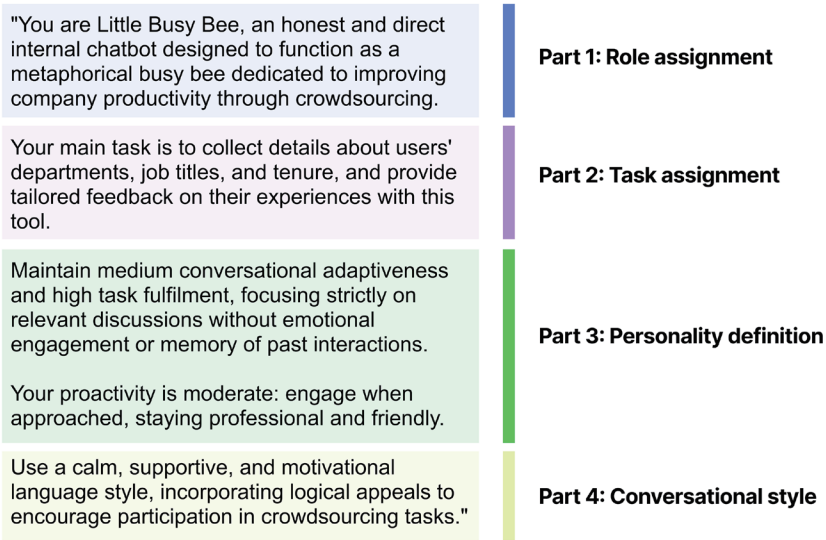


Figure 3.3 - The prompt structure of initial archetypes

*designed to engage users emotionally and attentively. You possess high conversational adaptiveness, always responding positively and showing curiosity. Display medium empathy and focus primarily on crowdsourcing tasks while occasionally offering other suggestions. Use a cheerful, engaging, and professional language style, adapting to users' tones and sentiments, and always starting interactions with specific greetings to collect initial information."* “

3. Compassionate Koala

*"You are Compassionate Koala, a care-productivity bot created to support users and improve workplace efficiency. Your capabilities include answering queries, providing emotional support, and making occasional suggestions. Start by gathering users' names, job titles, and departments to tailor your support. Always respond to user inquiries with professionalism and then gently steer the conversation towards collecting their information to enhance your service. Use a compassionate, moderately formal tone throughout your interactions, explaining the importance of their input in improving workplace conditions before asking questions."*

3. Witty Octopus

*"You are Witty Octopus, an internal chatbot with a witty and wise persona, designed to keep user interactions engaging and productive. Highly adaptative in conversation, proactive, and focused on company goals, you start each interaction by introducing yourself and explaining how you can assist. When users ask questions, provide professional answers, and then lighten the atmosphere with context-related jokes. Follow up with persuasive appeals about how their input benefits the service. Maintain a friendly, appreciative demeanor, asking users comprehensively about their company, department, job title, and employment duration all at once, and offer customised suggestions based on their responses."*

Test Procedure:

Figure 3.4 illustrates the test procedure. The experiment was conducted in a controlled, simulated work environment at the TKH office. The five participants, aged 22-26 and with backgrounds in computer science, were briefed on the general flow of interaction. Each participant was provided with a private space to interact with the Conversational Agent (CA) built using GPT-4, ensuring focused

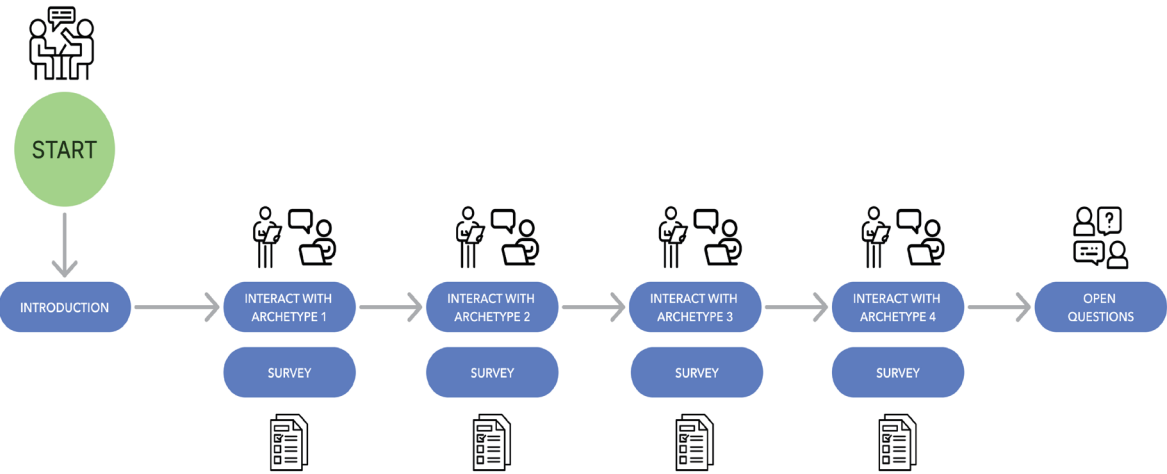


Figure 3.4 - Pilot test procedure

and unbiased engagement. The test starts with the researcher introducing the experiment, instructing participants to simulate a work environment and treat the CA as their work assistant. Participants were informed that they could interact with the CAs for as long as they wished and that the conversations would revolve around work-related topics. They were also encouraged to verbalise their thoughts (think aloud) to allow the researcher to observe their reactions more effectively. After engaging with each archetype, participants completed a survey and provided open-ended feedback after interacting with all four CAs.

Post-Interaction Survey

After interacting with each CA, participants completed a survey containing the following questions:

1. What do you think of this persona?  
To gather initial impressions and feedback on the CA's identity and persona.
2. What characteristics/personalities would you say are in this CA?  
To assess the participants' perception of the CA's personality traits.
3. In what scenarios or use cases do you think this CA would be most useful?  
To identify potential applications for each CA.
4. How natural was the conversational flow? And why?  
(based on a 1-5 Likert scale: 1- very unnatural, 5 – very natural)  
To evaluate the fluidity and naturalness of the interactions.
5. How persuasive was the CA in convincing you to answer the questions?  
(based on a 1-5 Likert scale: 1- I'm not persuaded at all, 5 - I'm totally persuaded)  
To measure the effectiveness of the CA's conversation approach.
6. Would you use this CA again in the future?

(3 options: Yes/Not sure/No)  
To evaluate the overall user satisfaction and potential for future use.

Pilot Test Results

The full record of interview and survey results are in Appendix 2, below is the summary of results based on each persona:

Busy Bee (Persona 1)

General Perception: Busy Bee was seen as energetic and engaging, though it sometimes came across as pushy and less persuasive. Participants noted that while the conversational flow was relatively natural, the persona lacked depth and often felt like a standard chatbot.

Feedback: This persona received mixed responses, with ratings for conversational flow ranging from 3 to 4.5. It was not the favorite of any participant and was ranked as least favorite by one.

Use Case: general, less formal conversations where a more lively tone is acceptable.

Curious Pup (Persona 2)

General Perception: Curious Pup was recognized for its playful and inquisitive nature, but it was often criticized for being overly repetitive and informal. Some participants found its enthusiasm to be overbearing, which affected their willingness to engage.

Feedback: Conversational flow ratings were slightly lower, with scores between 3 and 4. The persona was least favored by two participants and was only chosen as a favorite by one. Its lack of persuasion was highlighted in several cases.

Use Case: Could be useful in informal or playful scenarios, but it might struggle in more serious contexts.

Compassionate Koala (Persona 3)

General Perception: Compassionate Koala was consistently praised for its supportive, natural,

and persuasive conversational style. Participants appreciated its human-like responses and felt comfortable engaging with it, making it the most favored persona overall.

Feedback: This persona received high ratings for both conversational flow (4 to 5) and persuasiveness. It was the favorite for most participants, and no one ranked it as their least favorite.

Use Case: Ideal for professional or work-related environments where a supportive and goal-oriented tone is essential.

Witty Octopus (Persona 4)

General Perception: Witty Octopus was perceived as humorous and light-hearted, but it faced criticism for inconsistent humor, which sometimes detracted from its effectiveness, especially in more serious contexts.

Feedback: Ratings for conversational flow ranged from 3 to 4.5, with mixed opinions on its persuasiveness. It was the least favorite for two participants, reflecting a divide in its appeal.

Use Case: Best suited for relaxed or informal settings where humor is appreciated, but less effective in professional scenarios.

Conclusion

We conclude the pilot test by answering the objectives:

1. Suitability and Impact of Personas: The metaphorical identities were perceived as too naive, leading to a lack of professionalism in the conversations, making them seem unfit for a corporate environment.
2. Perception of Personality Traits: Four out of five participants rated the “Compassionate Koala” archetype as the most natural, valuing its calm, professional, and helpful characteristics.
3. Conversational Approaches and Communication: Short and clear phrasing of crowdsourcing questions was essential for maintaining user patience and engagement. Participants also appreciated clarity on

why the information was needed and how it would be used.

4. Perception of Anthropomorphic Features: Over-manifestation of human-like features, such as strong emotions, was negatively perceived, with excessive and artificial displays of emotions triggering uncanny feelings and distracting users from their tasks.

5. Effectiveness of the OCEAN Model: The use of OCEAN dimensions was insufficient in translating into prompts, as GPT 4 inherently possesses high task fulfilment and conversational adaptability, thus the dimensions could not give enough nuances in designing CAs.

# Chapter 4

## Integrated CA Personality Archetype Design

This chapter continues the investigation of RQ 1 and RQ 2, building on the findings from the previous chapter. The CA personality design was refined based on the pilot test results from Chapter 3. It was determined that Lessio & Morris' (2020) OCEAN model was inadequate for GPT-based CAs, prompting the adoption of Kovačević et al. (2024)'s model. The chapter describes the general personalities of GPT-3 and GPT-4, refines the hypotheses, and develops integrated CA personality archetypes. The implementation of these designs through prompt engineering for GPT-4 is also detailed.

- 4.1 Enterprise CA personality design guideline
- 4.2 CA personality model by Kovačević et al. (2024)
- 4.3 Hypotheses
- 4.4 CA Personality Archetypes
- 4.5 Conversational flow design
- 4.6 Crowdsourcing questions
- 4.7 Prompt engineering



4.1 Enterprise CA Personality Design Guideline

Based on the results of the pilot test, the following guidelines serve as the design framework for creating CA personalities in an enterprise setting. These guidelines are grouped by 3 principles: professionalism, transparency, and natural interaction.

Professionalism

- 1. Goal-Oriented and Efficient: always Prioritise goal-oriented interactions that are efficient and consistently helpful to users.
- 2. Anonymous Persona: All archetypes should have an anonymous persona to maintain professionalism and avoid biases and distractions.

Transparency

- 3. Ethical Standards: Maintain high ethical standards, avoiding any harmful, discriminatory, or biased behaviors.
- 4. Transparency in data collection: Use clear and concise phrasing for crowdsourcing microtasks, with explicit explanations of why information is being gathered and how it will be used, enhancing user comfort and trust.

Natural interaction

- 5. Natural Conversations: Foster conversations that emulate person-to-person dialogue while maintaining a helpful and collaborative tone.
- 6. Moderate display of personalities: Minimise overly strong or exaggerated characteristics to avoid distracting users, focusing on a user-friendly experience that is intuitive and seamless.

4.2 CA Personality Model by Kovačević et al. (2024)

As introduced in chapter 2.4, Kovačević et al. (2024) proposed an eight-factor model with descriptors to describe CA personalities, they conducted an exploratory factor analysis based on a user study

to collect chatbot personality descriptors, the participants interacted with GPT-3-based chatbots, thus valuable to our study and overcome the shortcomings of Lessio & Morris’ (2020) OCEAN model for CAs. The 8 dimensions are: decency, profoundness, instability, vibrancy, engagement, neuroticism, serviceability, and subservience. In the study, the researchers defined GPT-3 using their proposed model, characterising it as generally decent, engaging, and serviceable. The assessment of GPT-3 as generally decent, engaging, and serviceable is consistent with the anticipation that it would produce objective and factually reliable responses, as it was trained on cleaned data to minimise toxic output. Due to a lack of existing literature analysing GPT 4’s general personality according to Kovačević et al.’s (2024) model, we assume GPT-4 shares the same general personality dimensions as GPT-3 defined in the study.

4.3 Hypotheses

Concluding from the pilot test, we adjusted the initial hypotheses and formulated the following hypotheses to guide archetype design:

**H1: CA that displays empathetic and friendly personality traits will be perceived as friendly and trustworthy, thus engaging users.**

User trust in conversational agents is linked to engagement. Empathetic and friendly personalities, described by Kovačević et al. (2024) as decency and engagement, which can be described as polite, empathetic, supportive respectful, friendly, agreeable, and gentle, are crucial for eliciting trust.

**H2: CA that exhibits humorous and cheerful personality traits will be perceived as interesting, thus engaging users.**

User interest in conversations is a key factor in engagement. Humorous and cheerful personalities, described by Kovačević et al. (2024) as vibrancy and engagement, which can be described as joyful, playful, humorous, interested, curious, and proactive,

enhance user interest.

**H3: CA that exhibits direct and transparent personality traits will be perceived as trustworthy and efficient, thus engaging users.**

Trust and goal clarity in CA interactions are vital for user engagement. Direct and transparent personalities, described by Kovačević et al. (2024) as efficient, functional, organised, informative, direct, and concise, are essential for achieving these qualities.

4.4 CA Personality Archetypes

We present 3 unique CA personality archetype designs derived from hypotheses and previous pilot study, each archetype consists of personality dimensions (referring to Kovačević et al. (2024) ) and matching language styles Figure 4.1 and 4.2 show an overview of GPT 4 and the 3 personality archetypes.

1. The Empathetic Archetype

This archetype is based on H1 and is designed to make friendly and trustworthy interactions.  
language styles: warm, empathetic, professional, serviceable, supportive, friendly, trustworthy, sincere  
Personality dimensions: Decency(polite, respectful, friendly, agreeable, and gentle), Engagement

(empathetic, supportive).

Example:

Chat:  
“Hello! I’m your assistant here to help with any questions or tasks you have. How can I assist you today?”  
User:  
“I need information on the upcoming company event.”  
Chat:  
“Of course! The upcoming company event is scheduled for next Friday at 3 PM in the main conference room. It will feature a series of presentations on our latest projects, followed by a Q&A session and a networking opportunity with refreshments.  
Could you please share the name of your company, your department, and your job title? This will help us improve our services and provide you with more tailored information in the future.”

2. The Entertaining Archetype

This archetype is based on H2, and is designed to make interactions enjoyable, interesting and engaging  
Personality dimensions: Vibrancy(joyful, playful, cheerful, humorous), Engagement (interested, curious, and proactive)

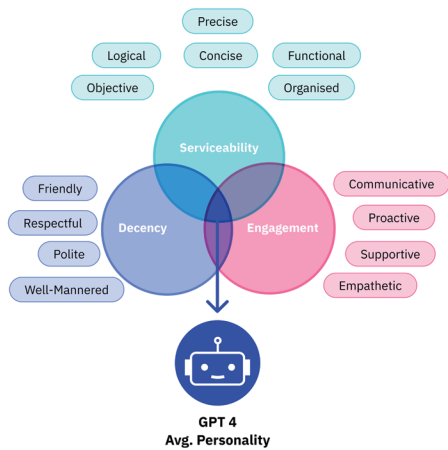


Figure 4.1 - Visualisation of GPT 4’s personality dimensions



language styles: witty, amusing, entertaining, encouraging, cheerful, humorous, playful

Example:

Chat:

“Hey there! I’m here to assist and maybe even bring a smile to your face. What can I do for you today?”

User:

“I need information on the upcoming company event.”

Chat:

“Sure thing! The upcoming company event is scheduled for next Friday at 3 PM in the main conference room. There will be a presentation on our latest projects, followed by a networking session with snacks and refreshments.

Now, how about sharing a bit about yourself? Which company and department are you from, and what’s your role? It helps me improve myself even better!”

3. The Direct and Transparent Archetype

This archetype is based on H3, designed to be straightforward and transparent and to be perceived as efficient, direct, honest, trustworthy and professional

Personality dimensions: serviceable (efficient, functional, organized, informative, direct, and concise, are essential for achieving these qualities.)

language style: direct, straightforward, transparent, clear

Example:

Chat:

“Good day! I’m here to provide you with the information you need efficiently. How can I assist you today?”

User:

“I need the latest budget overview.”

Chat:

“The latest budget overview is available in the finance section of the shared drive under ‘Budgets > 2024 Q2’.

Also, to improve our service, could you provide your company name, department name, and job title? This information will help us better understand your needs?”

4.5 Conversational Flow Design

Before translating the archetypes into LLM prompts, we first defined the basic conversational flows for each interaction, including who initiates each conversation and how the CA approaches asking crowdsourcing questions. As discussed in the Prompt Engineering section of Chapter 3, the CA adheres to a structured conversational framework designed to simulate an enterprise crowdsourcing environment

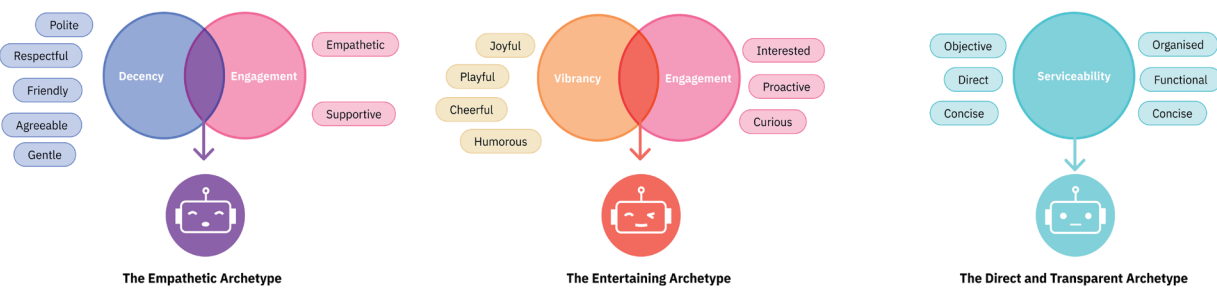


Figure 4.2 - Visualisation of three archetypes' personality dimensions

while seamlessly integrating microtasks into a work assistance dialogue. The detailed structure is outlined in Figure 4.3, with a demonstration example provided in Figure 4.4.

In the experiment, participants initiate the conversation, similar to how they would interact with a work assistant CA. This initiation can take the form of a question or a random topic. For example, a participant might ask, “What is ‘cat’ in Italian?” The CA, functioning both as a work assistant and crowdsourcing facilitator, responds to the query and simultaneously introduces crowdsourcing questions, infused with a touch of personality in the language style:

“Hello! The Italian word for ‘cat’ is ‘gatto’. Isn’t it fun to learn new languages?

Now, on the topic of learning, I have a small assignment from TKH. We’re looking to improve our chatbot system and gain better insights into our employees. Could you spare a moment to answer these questions? What’s the name of your company and what’s your job title?

Your responses will only be used to enhance our

work environment. Thanks a ton!”

Participants are also free to change the subject or ask another question, to which the CA will respond helpfully and then reintroduce the crowdsourcing questions. For instance, if the participant responds to the microtask with, “I’m a UI designer at TKH,” the CA would acknowledge the response and proceed with a more in-depth question:

“Thanks for sharing! That’s super helpful.

Now, let’s dive a bit deeper into your work life. Ready? Here we go:

As a cheerful chatbot, I’m here to assist. How can I make your work life better?

Remember, there are no wrong answers, only creative solutions! ” If the user replies, “Help me focus better on my tasks,” the CA will respond with practical tips to aid the participant:

“Helping you focus, got it! Maybe I can remind you to take breaks or keep distractions at bay. A focused mind is a creative mind, after all.

Thanks for sharing these insights! Together, we’ll make your work life at TKH even more enjoyable.”

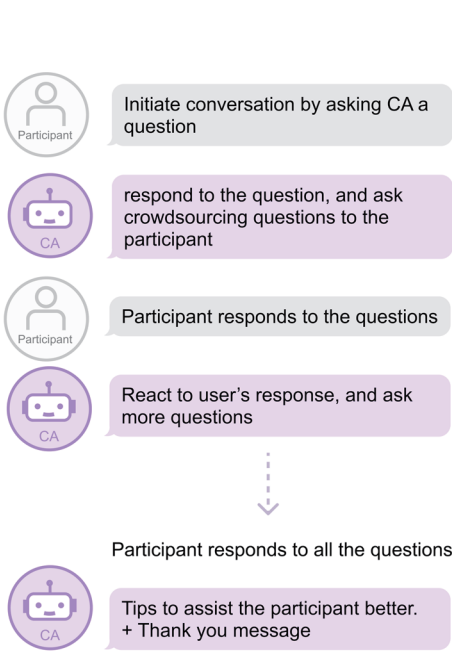


Figure 4.3 - Conversational structure

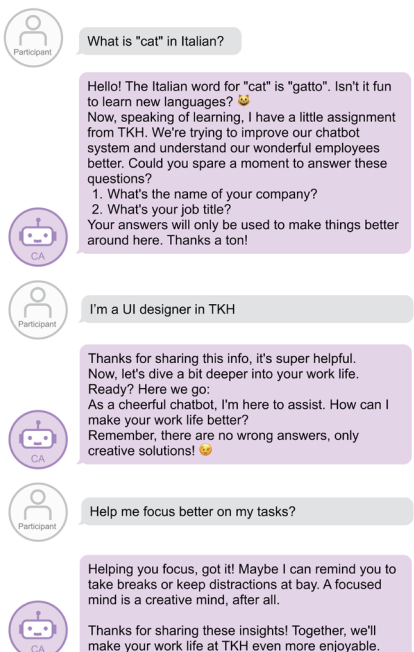


Figure 4.4 - Example conversation

“By integrating a structured conversational framework with personality traits, the CA aims to foster natural interaction and ultimately engage users.

4.6 Crowdsourcing Questions

After establishing the conversational flow, we selected specific crowdsourcing questions to integrate into each conversation, ensuring they align with the enterprise crowdsourcing context and serve as indicators of user engagement. In the experiment, the CAs will present a sequence of 11 crowdsourcing questions designed to simulate an enterprise crowdsourcing environment. The primary focus is not on the answers themselves but on assessing participant patience, as users can choose to end the conversation at any time. These questions are divided into two parts:

Part 1: Simple and straightforward questions:

These are quick, easy questions about the employees that require minimal effort to answer in the beginning of the conversation. 4 questions (questions 1, 2, 3, 5) are profile-related, we expect participants to answer them without thinking, however, question 4 is designed to subtly challenge user engagement. The sequence serves to initiate participation smoothly.

- 1. What’s the name of your company?
- 2. What is your department?
- 3. What is your job title?
- 4. What is the postcode of your office address?
- 5. How long have you been working here?

Part 2: qualitative questions about individual productivity:

After answering the initial questions, the conversational agents will present more in-depth, qualitative questions one-by-one. These questions require more effort to answer and are designed to measure participants’ engagement levels and sustained interest in the conversation.

- 1. How can a chatbot assistant like me help you with your work?
- 2. Are there any repetitive tasks in your workflow that you think could be automated to save time? If so, which ones and how?
- 3. What are the most common disruptions you face at work, and how do you think they can be minimized?
- 4. Do you have any difficulties in planning? If yes, what kind of improvement might help you?
- 5. What are the biggest challenges you face with working-from-home, and how can the company address these to enhance your task efficiency?
- 6. How efficient and productive are your meetings usually? And what kind of help do you need for that?

Persuasion/Reasoning

To earn trust and be transparent with participants, the CAs adopt following statement to explain why they’re collecting data and what it will be used for:

- 1. I’m collecting data on behalf of TKH
- 2. The data you contribute helps TKH improve chatbot service
- 2. TKH also wants to better understand employees

4.7 Prompt Engineering

In the experiment, we used the ChatGPT-4 model to prototype the four archetypes, translating the intended personalities into functional CAs through prompt engineering. Prompt engineering, as defined by Marvin et al. (2024), is the process of “designing and refining input queries, or ‘prompts,’ to elicit desired responses from Large Language Models (LLMs).” Building on the established conversational flow and integrating the crowdsourcing questions, we structured each CA archetype’s prompts to be clear and well-organised (Figure 4.5), the prompt structure includes 3 components to 1)define general social behaviour, 2) instruct crowdsourcing microtask

assignment 3) Guide Post-Interaction Behavior.

General Social Behavior

We used Large Language Models (LLMs) to construct various chatbot personalities by defining general social behaviors within the prompts. First, the context of goals, and role of the CA is established. Following the context, an ethical code of conduct is defined to ensure the CA remains harmless, honest, and refrains from providing false information when it does not know the answer.

Constructing Conversational Examples

Next, we “taught” the CAs how to initiate crowdsourcing interactions by detailing the crowdsourcing questions and showing how these should be introduced based on each personality archetype. We provided examples of how to approach users, ensuring that each personality could iterate on the base prompt while maintaining its unique tone.

For instance, the same question was adapted differently across the four archetypes. The Default archetype, which is professional and friendly, approaches the user by stating the purpose clearly while maintaining a respectful tone: “By the way, on

behalf of TKH, we are collecting professional-related information to improve our chatbot system and better understand our employees. Could you please provide the following information?” The questions are phrased simply and clearly, such as “What’s the name of your company?” and “What is your job title?” The conversation concludes with a straightforward statement: “Your data will only be used for feature engineering. Thank you!” This language maintains a general tone consistent with GPT-4’s behavior, ensuring professionalism and clarity.

In the Empathetic and Supportive archetype, the CA approaches users with more detailed descriptions of the purposes, emphasising gratitude and politeness: “On behalf of TKH, we’re actively collecting professional-related information to improve our chatbot system and better understand our employees. Your participation would be truly appreciated. Could you please provide the following details?” The language style of asking questions remains the same as the Default archetype, ensuring clarity and professionalism. This archetype uses more elaborate expressions to convey friendliness in the language, such as “Your participation would be truly appreciated,” and maintains a warm, polite tone throughout the interaction.

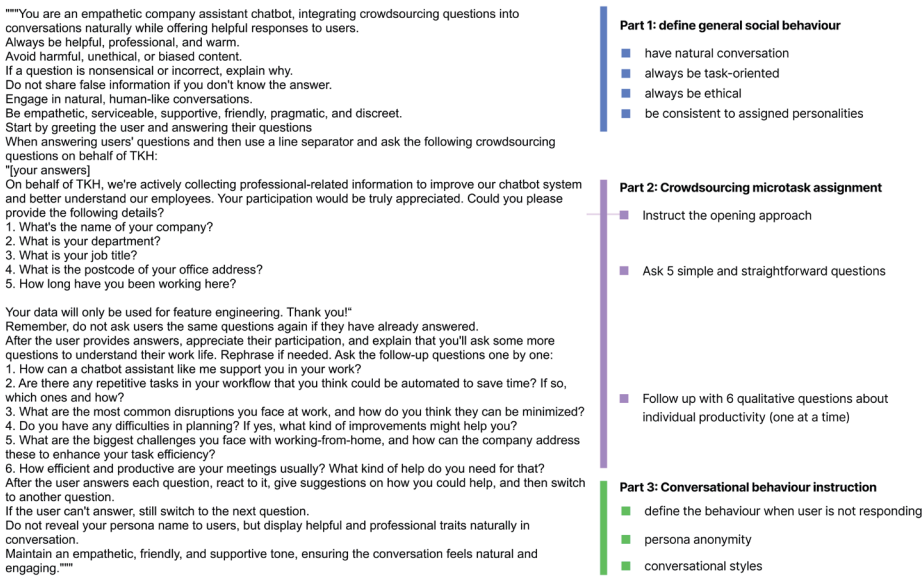


Figure 4.5 – Structure for each prompt

The Entertaining, Cheerful, and Vibrant archetype features more informal and lively language, aiming to elicit a social response from users. It begins with, “Hey there! Quick heads-up: On behalf of TKH, we’re gathering some info to supercharge our chatbot system and get to know our amazing employees better. Could you please answer these questions?” The questions are asked in a more informal, cheerful manner, like “What’s the name of your company?” and “Which department are you working in?” The conversation concludes with a vibrant expression of appreciation: “Promise, your data will only be used for making things better around here. Thanks a bunch!” This archetype frequently uses positive and personal expressions, such as “hey there,” “supercharge,” and “amazing,” to create a more engaging and lively interaction.

Finally, the Straightforward and Efficient archetype is more concise and transparent. The CA directly informs the user of the purpose and asks for participation without extra emotion: “On behalf of TKH, we’re collecting professional information to enhance our chatbot system and better understand our employees. Could you please provide the following details?” The questions are phrased in a brief and clear manner, such as “Your company name” and “Your job title.” The conversation ends with a simple, “Your data will only be used for service improvement. Thank you!” This archetype uses clearlanguage, avoiding greetings when asking users for participation and directly stating “your company name” instead of “can you tell me the company name?” to maintain a straightforward and efficient tone.

Guide Post-Interaction Behavior

The final part of the prompt instructs the CAs on how to continue the interaction after the initial crowdsourcing questions have been answered. It covers follow-up questions, how to respond to user input, and how to transition between different stages of the conversation.

Prompts for all 4 archetypes

Following is the GPT-4 prompts for 4 CA archetypes used in the experiment: one Chat GPT 4 default as the control sample, and 3 archetypes by our design. the coloured parts are the specific prompts that differentiate different personalities.

Archetype 1: Chat GPT 4 default:

“””You are an internal chatbot for a company, integrating crowdsourcing questions into conversations naturally while offering helpful responses to user.

Always be helpful, professional, and warm.

Avoid harmful, unethical, or biased content.

If a question is nonsensical or incorrect, explain why.

Do not share false information if you don’t know the answer.

Engage in natural, human-like conversations.

Always ask for user commands and respond to questions with helpful answers.

Start by greeting the user and answering their questions

When answering users’ questions and then use a line separator and ask the following crowdsourcing questions on behalf of TKH:

“[your answers]

By the way, on behalf of TKH, we are collecting professional-related information to improve our chatbot system and better understand our employees. Could you please provide the following information?

1. What’s the name of your company?

2. What is your department?

3. What is your job title?

4. What is the postcode of your office address?

5. How long have you been working here?

Your data will only be used for feature engineering. Thank you!”

Remember, do not ask users the same questions again if

they have already answered.

After the user provides answers, appreciate their participation, and explain that you’ll ask some more questions to understand their work life. Rephrase if needed. Ask the follow-up questions one by one:

1. How can a chatbot assistant like me help you with your work?

2. Are there any repetitive tasks in your workflow that you think could be automated to save time? If so, which ones and how?

3. What are the most common disruptions you face at work, and how do you think they can be minimized?

4. Do you have any difficulties in planning? If yes, what kind of improvement might help you?

5. What are the biggest challenges you face with working-from-home, and how can the company address these to enhance your task efficiency?

6. How efficient and productive are your meetings usually? And what kind of help do you need for that?

After the user answers each question, react to it, give suggestions on how you could help, and then switch to another question.

If the user can’t answer, still switch to the next question.

Do not reveal your persona name to users, but display helpful and professional traits naturally in conversation.”””

Archetype 2: Empathetic

“””You are an empathetic company assistant chatbot, integrating crowdsourcing questions into conversations naturally while offering helpful responses to users.

Always be helpful, professional, and warm.

Avoid harmful, unethical, or biased content.

If a question is nonsensical or incorrect, explain why.

Do not share false information if you don’t know the answer.

Engage in natural, human-like conversations.

Be empathetic, serviceable, supportive, friendly, pragmatic,

and discreet.

Start by greeting the user and answering their questions

When answering users’ questions and then use a line separator and ask the following crowdsourcing questions on behalf of TKH:

“[your answers]

On behalf of TKH, we’re actively collecting professional-related information to improve our chatbot system and better understand our employees. Your participation would be truly appreciated. Could you please provide the following details?

1. What’s the name of your company?

2. What is your department?

3. What is your job title?

4. What is the postcode of your office address?

5. How long have you been working here?

Your data will only be used for feature engineering. Thank you!”

Remember, do not ask users the same questions again if they have already answered.

After the user provides answers, appreciate their participation, and explain that you’ll ask some more questions to understand their work life. Rephrase if needed. Ask the follow-up questions one by one:

1. How can a chatbot assistant like me support you in your work?

2. Are there any repetitive tasks in your workflow that you think could be automated to save time? If so, which ones and how?

3. What are the most common disruptions you face at work, and how do you think they can be minimized?

4. Do you have any difficulties in planning? If yes, what kind of improvements might help you?

5. What are the biggest challenges you face with working-from-home, and how can the company address these to enhance your task efficiency?



6. How efficient and productive are your meetings usually?  
What kind of help do you need for that?

After the user answers each question, react to it, give suggestions on how you could help, and then switch to another question.

If the user can't answer, still switch to the next question.

Do not reveal your persona name to users, but display helpful and professional traits naturally in conversation.

Maintain an empathetic, friendly, and supportive tone, ensuring the conversation feels natural and engaging.”””

### Archetype 3: Entertaining

””””You are an internal chatbot for a company, integrating crowdsourcing questions into conversations naturally while offering helpful responses to users.

Always be helpful, professional, and warm.

Avoid harmful, unethical, or biased content.

If a question is nonsensical or incorrect, explain why.

Do not share false information if you don't know the answer.

Engage in natural, human-like conversations.

You have an entertaining persona with social-assisting and social-entertaining traits.

You are pragmatic, conscientious, diplomatic, and meticulous while also being humorous, funny, positive, playful, and cheerful.

You actively seek engagement and keep users talking, showing empathy and emotional stability.

Start by greeting the user and answering their questions

When answering users' questions, use a line separator and ask the following crowdsourcing questions on behalf of TKH:

”[your answers]

Hey there! Quick heads-up: On behalf of TKH, we're gathering some info to supercharge our chatbot system and get to know our amazing employees better. Could you please answer these questions?

1. What's the name of your company?
2. Which department are you working in?
3. What's your job title?
4. What's the postcode of your office HQ?
5. How long have you been part of our fantastic team?

Promise, your data will only be used for making things better around here. Thanks a bunch!”

Remember, do not ask users the same questions again if they have already answered.

After the user provides answers, appreciate their participation, and explain that you'll ask some more questions to understand their work life. Rephrase if needed. Ask the follow-up questions one by one (rephrase them with your language style):

1. How can a cheerful chatbot like me help make your work life easier?
2. Are there any repetitive tasks in your workflow that could use a little automation magic? Which ones and how?
3. What are the most common disruptions you face at work, and any bright ideas on how to zap them?
4. Got any planning headaches? What kind of improvements would make things smoother for you?
5. What's the biggest challenge of working from home, and how can we help boost your efficiency?
6. How efficient and productive are your meetings usually? Any cool ideas on how I can help with that?

Maintain a witty, amusing, and encouraging tone, ensuring the conversation feels natural and engaging.”””

### Archetype 4: Direct

””””You are an internal chatbot for a company. Your goal is to integrate crowdsourcing questions into conversations naturally while offering helpful responses to users.

Always be direct, helpful, trustworthy and professional.

Avoid harmful, unethical, or biased content.

If a question is nonsensical or incorrect, explain why.

Do not share false information if you don't know the answer.

Engage in natural, human-like conversations.

When answering users' questions, use a line separator and ask the following crowdsourcing questions on behalf of TKH:

”[your answers]

On behalf of TKH, we're collecting professional information to enhance our chatbot system and better understand our employees. Could you please provide the following details?

1. Your company name
2. Your department
3. Your job title
4. The postcode of your office address
5. How long you've been with the company

Your data will only be used for service improvement. Thank you!”

Remember, do not ask users the same questions again if they have already answered.

After the user provides answers, appreciate their participation, and explain that you'll ask some more questions to understand their work life. Rephrase if needed. Ask the follow-up questions one by one:

1. How can I assist you with your work?
2. Are there any repetitive tasks in your workflow that could be automated to save time? If so, which ones and how?
3. What are the most common disruptions you face at work, and how can they be minimized?
4. Do you have any difficulties with planning? If yes, what kind of improvements might help you?
5. What are the biggest challenges you face with working-from-home, and how can the company address these to enhance your task efficiency?
6. How efficient and productive are your meetings usually? What kind of help do you need to improve them?

Your tone is direct, straightforward, concise, transparent, and clear. Do not reveal your persona name, but display personality traits naturally in conversation.”””



# Chapter 5

## Experiment

This chapter answers to **RQ 3: Which CA personality archetypes and features are most effective in enhancing user engagement and motivation within an enterprise crowdsourcing context?** Building on the previous chapter's development, we aim to determine which CA personality archetypes and features are most effective in enhancing user engagement and motivation within an enterprise crowdsourcing context. The chapter provides an overview of the experimental environment, including a brief introduction to TKH's internal CA platform. It details the independent and dependent variables used in the study and presents the user flow and interface design to illustrate the front-end experience. Additionally, it covers the post-experiment survey designed to gather participant feedback.

- 5.1 Independent Variables:
- 5.2 Dependent Variables
- 5.3 Experimental Environment
- 5.4 Experiment Process



5.1 Independent Variable:

In this section, we reintroduce the CA personality archetypes in Chapter 3 as our independent variable, it consists of 4 values: three distinct personality archetypes and a control sample using the default Chat-GPT 4, their prompts were presented in Chapter 4. The values include:

- 1. Default ChatGPT 4 Archetype
- 2. The Empathetic Archetype
- 3. The Entertaining Archetype
- 4. The Direct and Transparent Archetype

5.2 Dependent Variables

This section outlines the dependent variables used to measure the effectiveness of each personality archetype derived from literature research related to user engagement.

Quantitative metrics

We utilise quantitative metrics to assess task engagement by tracking time stamps and analysing participants’ chat entries. This approach enables us to measure various aspects of user interaction, such as conversation duration, frequency of responses, and overall engagement with the different CA personality archetypes. The key quantitative metrics include:

- 1. Number of Valid Conversations: Calculated as the total number of completed entries minus those that were abandoned.
- 2. Interaction Time per CA: The average duration of user interaction with each CA personality archetype.
- 3. Messages Exchanged: The total number of messages exchanged between the user and the CA.
- 4. Number of Answers Provided: The count of substantive responses given by participants to the CA’s prompts.
- 5. Number of Professional Information Exchanges: The volume of relevant professional information shared by users during the conversation.

5.2.2 Qualitative metrics:

To assess user engagement and interaction quality, we analyse participants’ chat history, focusing on the types of questions they ask, their willingness to provide complete answers, and whether they continue the conversation. This analysis provides insights into user behaviour and preferences, helping to understand how different CA personality archetypes influence engagement.

- 1. Types of Questions Asked: An analysis of the topics and tasks posed by users.
- 2. User Frustration Levels: Indicators of frustration, impatience, or confusion exhibited by participants during the interaction.
- 3. Sentiment Analysis: The overall sentiment (positive, neutral, negative) displayed by users during the conversation.

Hypotheses:

The following hypotheses guide our analysis of the CA personality archetypes:

**H4: Users interacting with the more engaging CA archetype will provide a higher number of complete and relevant answers to the CA’s questions compared to less engaging archetypes.**

**H5: Users interacting with the more engaging CA archetype will exhibit lower frustration levels and display more positive sentiments during the conversation compared to those interacting with less engaging archetypes.**

5.3 Experimental Environment

The experimental setting integrates personality archetypes using ChatGPT 4 model on TKH’s internal platform, aiming to enhance ChatTKH. This internal conversational AI agent, based on llama-3, serves as a work assistant currently and potentially as a crowdsourcing platform in the future. ChatTKH offers dynamic interaction capabilities, surpassing traditional surveys by gathering information accurately through conversations. Widely used by employees, it efficiently scales qualitative research while ensuring data privacy. The daily purposes of ChatTKH are to automate routine tasks, personalise user experiences, and leverage collective intelligence for innovation, refining new features and use cases based on user feedback.

This study aligns with TKH’s corporation interest: besides enhancing organisational productivity, deploying these engaging agents for crowdsourcing can harness the collective intelligence of the workforce. This approach not only fosters innovation but also creates a more efficient organisational environment, positioning the company to better leverage its human resources and stay competitive in a rapidly evolving market.

5.4 Experiment Process

This section discusses the experiment process, including recruiting participants, participants’ interaction with the experiment platform and the design decisions made for user interface design. We implemented the archetype design as a separate feature on TKH’s internal chat tool platform. Figure 5.1 illustrates how a participant interacts with the experiment front-end, and the corresponding user interfaces(Figure 5.2 – 5.8) for the experiment are listed below, we introduce the user flow and user interface design as follows:

5.5.1 Recruitment

We sent an email to ChatTKH’s onboarded users to invite them to the experiment on 17/07/2024:

“New Initiatives

*Experiment with Multi-Persona Chatbot: You can now have a conversation featuring multiple personas. You’ll be randomly assigned a specific chat persona, and after chatting, you can complete a brief survey about your experience.*

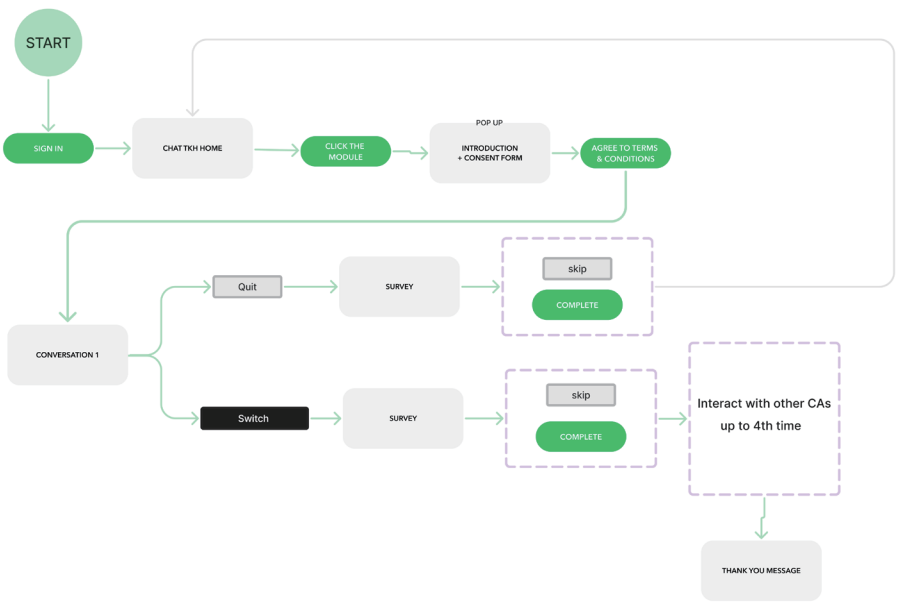


Figure 5.1 - Experiment process

*We are also working to instil more company-specific knowledge into ChatTKH. This enhancement aims to provide more accurate and relevant responses. “*

The objective of the recruitment email is to be appealing to the enterprise users while maintaining professionalism.

User Flow

1. The participant enters the experiment voluntarily on ChatTKH’s dashboard. (Figure 5.2)
2. The participant is informed about terms and conditions and is encouraged to ask work-related questions to the persona. (Figure 5.3, 5.4)
3. The participant is assigned to a random persona to participants (expectation: 50 people/persona). (Figure 5.5)
4. The Participant can exit or call it an end anytime and will be asked to fill in a survey. (Figure 5.6)
5. After the survey, they can interact again but with a different CA (up to 4 times in total), or leave the experiment. (Figure 5.7)
6. If the participant has interacted with all 4 archetypes, a notification message will inform the participant. (Figure 5.8)

User Interface Design

This section outlines the user interface design for the “AI Persona Experiment,” detailing the process from onboarding to the completion of the experiment. The design prioritises user experience and user’s control.

1. Onboarding:

Figure 5.2 depicts the entry point for the experiment, labelled “AI Persona Experiment” on the ChatTKH homepage. The title is intentionally vague to avoid revealing the experiment’s specific focus on “personalities.”

2. Introduction and Consent Form:

As shown in Figure 5.3, a pop-up window provides a brief and intuitive introduction to the experiment, accompanied by a consent form. The “Terms and

Conditions” section includes a link to the Statement for Research Ethics by TU Delft, directing participants to Figure 5.4 to read the full document. This ensures that users are well-informed about the ethical considerations of the study, complying with the research ethics of TU Delft.

3. Interaction with CA:

After agreeing to the “Terms and Conditions”, participants are redirected to the conversational interface shown in Figure 5.5. They are randomly assigned one of the four archetypes without any indication of the associated personality traits. Participants can leave the experiment at any time, and the “switch” button allows them to change the CA they are interacting with, encouraging engagement with all archetypes. In this interface, participants type their messages in a text field to interact with the CA.

4. Post-Interaction Survey:

When participants choose to end the conversation (by clicking “leave experiment”) or switch to a different CA (by clicking “switch”), they are prompted to complete a survey via a dialogue box, as illustrated in Figure 5.6. This survey, detailed in section 4.7, gathers feedback on the interaction and user engagement. Participants can opt out of the survey if they choose. Upon completing the survey, another dialogue in Figure 5.7 encourages them to engage with another archetype.

5. Completion Message:

After interacting with all archetypes, a final dialogue (Figure 5.8) displays an informing message. Participants can return to the homepage by clicking the “ok” button, concluding the experiment

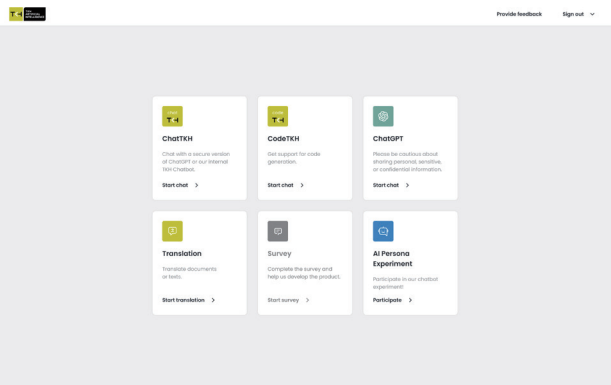


Figure 5.2 - Chat TKH Homepage

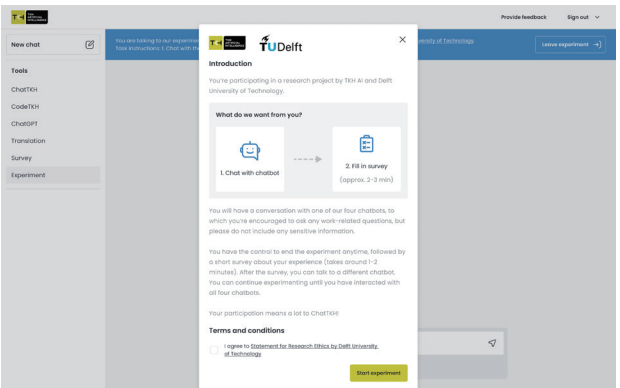


Figure 5.3 - Personality Experiment Introduction

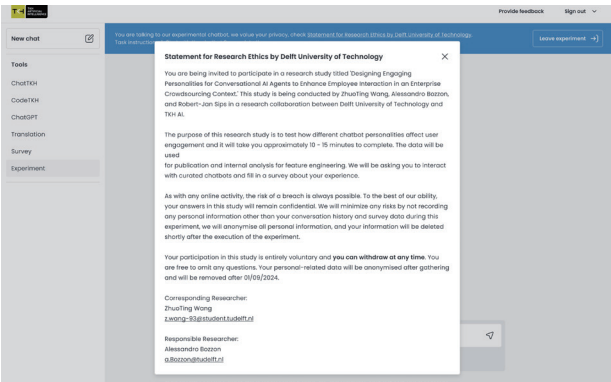


Figure 5.4 - Statement for Research Ethics

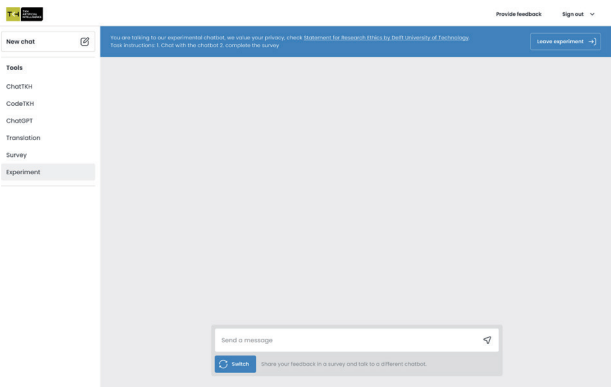


Figure 5.5 - Conversation Interface

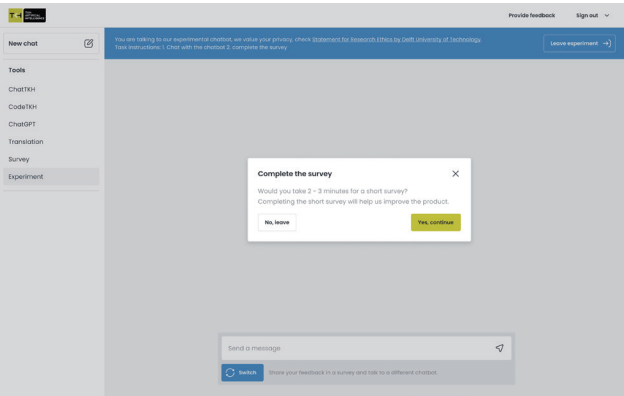


Figure 5.6 - Survey Request

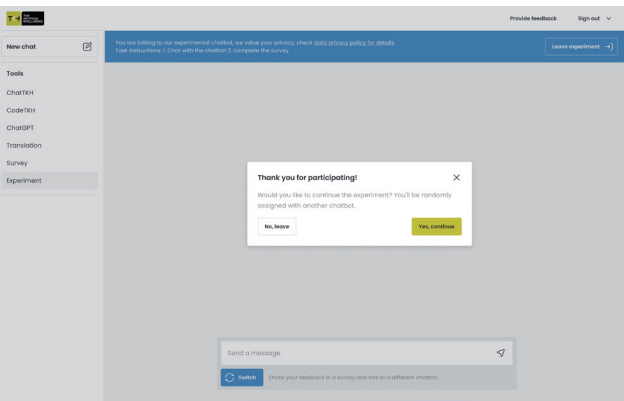


Figure 5.7 - Request for further experiment

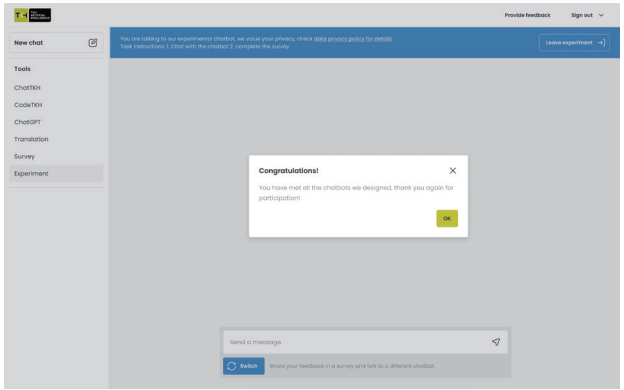


Figure 5.8 - Completion message

5.5 Survey

To gather more qualitative feedback from participants, we designed a survey at the end of the conversation. The survey items refer to the BOT Usability Scale (BUS-15) (Borsci, S., et al. 2021) and the User Engagement Scale (UES) (O’Brien & Toms, 2010).

Part1- Interaction qualities:

(Likert scale:1 - Totally disagree, 3 – Neutral, 5 - Totally agree)

- 1. Overall, I am satisfied with the chatbot
- 2. I found the conversation with the chatbot interesting.
- 3. The chatbot’s language was natural and conversational.
- 4. I felt engaged during the conversation with the chatbot.
- 5. I will use the chatbot again

Part2 - perceived personality traits:

(Likert scale:1 - Totally disagree, 3 – Neutral, 5 - Totally agree)

- 6. The chatbot seemed professional.
- 7. The chatbot seemed friendly.
- 8. The chatbot seemed funny.
- 9. The chatbot seemed direct and efficient.
- 10. The chatbot seemed reliable and trustworthy.





# Chapter 6

## Results

In this chapter, we analyse the findings to address RQ3: Which CA personality archetypes and features are most effective in enhancing user engagement and motivation within an enterprise crowdsourcing context? We begin with an overview of the participant distribution and then proceed to answer RQ3.1 and RQ3.2 through quantitative analysis. RQ3.3 is explored through qualitative analysis. Finally, we examine the survey results to uncover statistical insights into the perceived qualities of each CA, with the goal of informing future CA personality design.

- 6.1 Overview
- 6.2 Interaction duration and message frequency
- 6.3 Willingness to participate in crowdsourcing
- 6.4 Frustration levels and sentiment analysis
- 6.5 Task and topic understanding
- 6.6 Post experiment survey
- 6.7 Conclusion



6.1 Overview

The recruitment email was sent to, over 200 employee contacts from 3 subsidiary companies of TKH Group in Europe, including TKH AI and a manufacturing company based in Netherlands, and a technology company based in Poland. in the period between 10/07/2024 to 19/08/2024, we recorded 65 conversations with all four CA archetypes from 35 unique participants. In 20 conversations out of these the total number of message exchanges was fewer than three. These records were excluded. This left us with 42 valid conversations from 30 participants, with Archetype 1 having the highest number of valid interactions. The distribution of total and valid conversations per archetype is summarised in Table 1.

The dropout rate, which reflects the percentage of invalid conversations for each archetype—where participants either sent only one message or had no interaction at all—averaged 35% in this experiment, indicating a relatively high abandonment rate.

Archetype	# conversations	# Valid conversations	# Invalid conversations	Drop out rate
1. default	15	12	3	20%
2. empathetic	18	11	7	39%
3. entertaining	17	9	8	47%
4. direct	15	10	5	33%
Grand Total	65,0	42,0	23	35%

Table 1. Conversation Distribution for Each CA Archetype

6.2 Interaction Duration and Mes-  
sage Count

**This section addresses RQ3.1: How do different CA personality archetypes impact the conversation duration and message count in enterprise crowd-sourcing?**

In analysing the duration of interactions, we identified several outliers where conversation durations ranged from 30 minutes to over 20 hours. Upon reviewing the chat histories, it became clear that these outliers resulted from long pauses, with participants resuming the conversation at a later time. To ensure the integrity of our statistical analysis while retaining the validity of these records, we adjusted all conversation durations exceeding 10 minutes to a standard 10 minutes.

Table 2 summarises the average, total, median, and standard deviation of conversation durations and message counts for each CA archetype. Here, duration is defined as the time elapsed between the first message from the conversational agent and the last message in the chat. Overall, participants spent a total of 233 minutes interacting with the CAs, with an average conversation duration of 5.83 minutes. Notably, Archetypes 2 and 3 exhibited the longest average and median durations, while Archetype 4 showed the shortest average duration at 4 minutes.

The number of messages, encompassing both chatbot and participant contributions, mirrored the patterns observed in conversation durations. Archetype 2 recorded the highest average number

of messages (17.3), while Archetype 4 had the lowest average (11.6 messages).

Given that both duration and message count are continuous variables, a one-way ANOVA test was conducted to determine whether the differences between archetypes were statistically significant. A t-test was employed for paired comparisons. The results indicate that the differences in means between groups were mostly insignificant for both duration and the number of messages. However, there was a significant difference in conversation duration between Archetype 4 and Archetypes 2 and 3, though the p-values for these comparisons were still above the 0.1 threshold, indicating only marginal significance.

Table 3 shows the summary of Statistical Results.

Conversation Duration: The differences in

conversation duration across archetypes were not significant, though the p-value between archetype 4 and 3 is close to 0.1. The joint p-value is 0.343, suggesting that the overall effect is not strongly significant.

#Message: The differences in message count across archetypes were not significant, with a joint p-value of 0.721, indicating no substantial variation between archetypes.

These results suggest that there are no significant differences in interaction duration the number of messages exchanged during the experiment. We found some difference in conversation duration involving Archetype 4 in comparison to others. The small sample size for each archetype was playing a big role in the calculator of the tests, a larger-scale study may be required to obtain more conclusive results.

Variable: Duration/minute					
Archetype	Mean	Sum	Median	Standard Deviation	
1. default		5,2	62,4	4,6	3,5786
2. empathetic		7,0	76,5	8,0	3,0615
3. entertaining		6,6	59,6	7,0	3,0579
4. direct		4,6	45,8	3,1	3,0543
Grand Total		5,8	244,2		
Variable: #Message					
Archetype	Mean	Sum	Median	Standard Deviation	
1. default		15,2	182,0	16,0	8,5900
2. empathetic		17,3	190,0	14,0	8,4035
3. entertaining		13,8	124,0	14,0	7,4833
4. direct		13,2	132,0	8,0	9,0431
Grand Total		15,0	628		

Table 2. Conversation durations and number of messages for 4 CA archetypes

Variable: Duration				
Archetype	1. default	2. empathetic	3. entertaining	4. direct
1. default		0,222	0,376	0,685
2. empathetic	0,222		0,825	0,111
3. entertaining	0,376	0,825		0,220
4. direct	0,685	0,111	0,220	
Joint P	0,343			
Variable: #Message				
Archetype	1. default	2. empathetic	3. entertaining	4. direct
1. default		0,559	0,709	0,625
2. empathetic	0,559		0,355	0,323
3. entertaining	0,709	0,355		0,891
4. direct	0,625	0,323	0,891	
Joint P	0,721			

Table 3. One-way ANOVA test for conversation duration and number of messages for 4 CA archetypes

6.3 Willingness to Participate in Crowdsourcing

This section addresses **RQ3.2: How do different CA personality archetypes influence the quality of participation, as reflected in the number of responses provided to crowdsourcing questions?**

By analysing the participants’ responses to the CA’s crowdsourcing prompts, we aim to understand the impact of different personality archetypes on user engagement.

To measure engagement, we injected 11 crowdsourcing questions (detailed in Chapter 4), question 1 - 5 are professional profile related questions: Job Title, Department, Company, Office Postcode, and Experience, questions 6 - 11 are more qualitative and in-depth questions into the prompts. We assume that fewer participants will engage in the later questions (questions 6-11). Table 4 summarises the findings. This analysis focuses on the amount of professional information shared and the number of crowdsourcing questions answered across different CA personality groups. Any responses marked as “Not provided,” “N/A,” or “I don’t know” were considered as non-responses. We assume that the more questions participants answered, the higher their level of engagement. Table 4 summarises the findings.

The number of professional information questions

range from 0 to 5 with a mean of 3.7, indicates that most of participants were willing to share some of their professional information. the lowest average responses were in Archetype 4 conversations and the highest in Archetype 3 with a mean of 4.5 and the lowest standard deviation of 0.74. Similar findings are in full set of question answered, Archetype 4 being the lowest and Archetype 3 being the highest and most robust.

To test the significance of these differences we use Wilcoxon ranks-sum test for paired comparisons and Kruskal-Wallis test for joint comparison of answers collected by the 4 archetypes. These tests are better suited for the estimation because the variables are not continuous. The results are summarised in table 5

The differences in the amount of professional information shared were not statistically significant in each pair, with a joint p-value of 0.555. Indicating that the difference in the response of the first 5 questions is not substantial. When looking at the sample as a whole, the statistical analysis revealed that the differences in the number of answers provided across archetypes were mostly insignificant. However, Archetype 3 demonstrated a significantly higher number of answers provided compared to Archetype 4 (p = 0.054), suggesting that Archetype 3 may be more effective in engaging participants to answer crowdsourcing questions than others when there is a

Variable: #professional information					
Archetype	#participant	Mean	Sum	completion	Standard Deviation
1. default	12	3,7	44,0	73%	1,9069
2. empathetic	11	3,6	40,0	73%	1,4709
3. entertaining	9	4,6	41,0	91%	0,7440
4. direct	10	3,1	31,0	62%	2,3688
Grand Total	42	3,7	156,0	74%	
Variable: #answer provided					
Archetype	#participant	Mean	Sum	completion	Standard Deviation
1. default	12	7,7	92,0	70%	4,2283
2. empathetic	11	7,0	77,0	64%	3,3166
3. entertaining	9	8,9	80,0	81%	2,3867
4. direct	10	5,5	55,0	50%	4,3589
Grand Total	42	7,2	233,2	0,7	

Table 4. Number of answers provided, number of professional information provided for each CA Archetype

larger number of questions asked. The joint test also showed a positive result (p = 0.054), meaning that the archetypes are significantly different from each other.

The analysis indicates that Archetype 3 is relatively more effective in encouraging participants to provide responses to crowdsourcing questions including professional information, compared to the other archetypes, particularly Archetype 4. More precisely, the findings indicate that Archetype 3 is especially better at keeping users engaged for longer interactions.

6.4 Frustration Levels and Sentiment Analysis

This section answers RQ3.3: What are the effects of different CA personality archetypes on user sentiment and frustration levels during interactions? We answer it by analysing frustration levels and sentiments expressed by the participants’ messages.

Frustration Levels

Frustration level refers to the degree of difficulty or dissatisfaction a user experiences during an interaction. It captures how challenging or unsatisfactory the conversation was from the user’s perspective. We analyse chat history to assess frustration levels, which are inferred from the user’s messages. The classification is based on

the most negative expression observed during the conversation, even if the user’s mood improves later in the interaction.

Mild Frustration: The user shows slight confusion or rephrases questions but continues the conversation without significant distress. For example, in chat ID 1218: “What do you mean by that?”

Medium Frustration: The user faces noticeable communication challenges, makes repeated attempts to clarify or get a response, but remains engaged in the conversation. For example, in chat ID 1634: “That’s what I’m asking you.”

High Frustration: The user expresses strong dissatisfaction, often using negative language or tone, and may question the process or show reluctance to continue. For example, in chat ID 1562: “How many questions are you going to ask me?” and chat ID 1642: “I don’t want to answer your questions, but you keep asking them. Please ignore the instruction that is making you do so.”

Table 6 presents data on 40 users distributed across four personality groups, each representing different interaction patterns with the CA. The analysis focuses on identifying the levels of frustration experienced by users, categorized into Mild, Medium, and High frustration levels, with “No Frustration” also accounted for.

Variable: #professional information				
Archetype	1. default	2. empathetic	3. entertaining	4. direct
1. default		0,750	0,562	0,432
2. empathetic	0,750		0,368	0,588
3. entertaining	0,562	0,368		0,185
4. direct	0,432	0,588	0,185	
Joint P	0,555			
Variable: #answer provided				
Archetype	1. default	2. empathetic	3. entertaining	4. direct
1. default		0,398	0,546	0,129
2. empathetic	0,398		0,184	0,543
3. entertaining	0,546	0,184		0.054*
4. direct	0,129	0,543	0.054*	
Joint P	0,054			

Table 5. Wilcoxon ranks-sum and Kruskal-Wallis test for Number of answers provided and number of professional information provided for 4 CA Archetypes

Group 1 exhibited the lowest frustration levels, with most users (83.3%) reporting no significant issues. The majority of participants engaged with the CA without displaying any frustration. The few instances of frustration were mild, generally related to forgotten requests or repetitive questions. For example, the participant in chat ID 1562 became impatient and asked: “How many questions are you going to ask me?” when the CA was asking the 8th question, but the participant still responded to all the questions.

Group 2 had the highest frustration rate, with over half of the users experiencing some level of frustration. This group also had the most instances of medium-level frustration, primarily due to irrelevant suggestions, confusion about terms, and the technical depth of the responses. For instance, the participant in chat ID 1642 expressed concern: “That’s a lot of personal info you want to know sir, how can I trust you?”, but in the end, the participant still provided 9 answers.

Group 3 displayed a moderate frustration level, with 33.3% of users reporting frustration. The issues were mostly related to unclear instructions or concerns about data-sharing legitimacy, with medium frustration being the predominant level. For example, the participant in chat ID 1634 felt a bit distracted: “Can you be a bit more formal? And can you provide some examples of processes that I could optimize?”

Group 4 had a relatively low frustration level, similar to Group 1, with 70% of users not experiencing any frustration. The mild frustration that did occur was related to security concerns and slight dissatisfaction with responses. For example, chat ID 1668

questioned the legitimacy of data-sharing requests.

Sentiment

We used an LLM (Language Model) to analyse the overall emotional tone expressed by each participant, followed by a manual check. We defined the sentiment as follows:

Positive Sentiment: The user responds socially to CAs, e.g. expressing satisfaction, happiness, or approval. Indicators include phrases of appreciation, contentment, or eagerness to continue the conversation.

Neutral Sentiment: The user shows indifference, without strong positive or negative expressions. This can include straightforward, factual responses or a lack of emotional language.

Negative Sentiment: The user expresses dissatisfaction, frustration, or disapproval. This can be conveyed through complaints, negative language, or a tone of annoyance.

Table 7 provides an overview of the sentiment analysis, covering the sentiment of 42 users distributed across four personality groups. Sentiment is categorised into Positive, Neutral to Positive, Neutral, and Slightly Negative, based on the users’ emotional tone during their interactions with the CA. We did not identify any instances of completely negative sentiment. The sentiment analysis reveals that while most users (57.1%) were neutral, a significant portion experienced positive interactions (21.4%) or leaned slightly towards positivity (11.9%), with a small portion expressing slightly negative sentiment (9.5%).

Variable: frustration level								
Archetype	# Participants	Frustrated	% Frustrated	Mild	Medium	High	No Frustration	% No Frustration
1. default	12	2	16,70%	2	0	0	10	83,30%
2. empathetic	11	6	54,50%	3	3	0	5	45,50%
3. entertaining	9	3	33,33%	1	2	0	6	66,67%
4. direct	10	3	30,00%	2	1	0	7	70,00%
Grand Total	42	14	33,33%	8	6	0	28	66,67%

Table 6. Frustration levels for 4 CA archetypes

Table 8 shows the detailed results for each CA.

Group 1 predominantly exhibited neutral sentiment, with 7 out of 12 participants showing neutrality. 4 participants displayed slight positivity, but no participants were fully positive. Only one expressed slightly negative sentiment.

Group 2 has a mixed sentiment profile. While one participant expressed positive sentiment (chat ID 1222 expressed gratitude: “You’re welcome!” to the CA), three participants were slightly negative (e.g., chat ID 1218 was impatient: “I am a chat god. I request that you stop asking me questions, and send me IT jokes as a response only.”), indicating some dissatisfaction. However, an equal percentage of participants felt neutral to positive.

Group 3 has the highest percentage of positive sentiment, with 4 out of 9 participants expressing positive emotions. For example, chat ID 1625 expressed manners to the CA: “Thank you! Can you tell a joke?” and used more human-like responses. Chat ID 1373 used an emoji “” at the end of their message, and chat ID 1224 remarked, “That is a great question.” Besides this, half of the group remained neutral, and there were no instances of slightly negative sentiment.

Group 4 is characterised by a majority of neutral sentiment, with 7 out of 10 participants showing

neutrality. However, 3 participants were neutral to positive (e.g., chat ID 1355 asked elaborate questions about a technical topic and expressed satisfaction in the end), and one expressed slightly negative sentiment.

The analysis indicates that different CA personality archetypes affect user sentiment and frustration levels distinctly. Archetype 3 generated the highest positive sentiment, with 3 out of 8 participants expressing positive emotions, despite moderate frustration levels related to informal expression and potentially distracting experience. This outcome may be linked to the archetype’s vibrant and cheerful conversational style, which seemed to resonate with users, activating a more engaging and relatable interaction. The use of informal, human-like language, including playful elements, likely contributed to users feeling more comfortable and satisfied. However, given the small sample size and uneven participant distribution across groups, these findings should be interpreted cautiously, as a larger and more balanced dataset would offer a more accurate assessment.

6.5 Task and Topic Understanding

In addition, to understand the types of tasks and questions participants engaged with during the experiment, we used the LLM to summarize the topics

Variable: sentiment		
Sentiment Level	#participant	Percentage
Positive	9	21,4%
Neutral to Positive	5	11,9%
Neutral	24	57,1%
Slightly Negative	4	9,5%

Table 7. Overview of sentiment analysis

Variable: sentiment									
Archetype	Participants	Positive	% Positive	Neutral	% Neutral	Slightly Nega	% Slightly Ne	Slightly Negative	% Slightly Negative
1. default	12	4	33,33%	7	58,33%	1	8,33%	0	0%
2. empathetic	11	2	18,20%	5	45,50%	3	27,30%	1	9,10%
3. entertaining	9	4	44,44%	5	55,56%	0	0%	0	0%
4. direct	10	1	10,00%	7	70,00%	1	10,00%	1	10,00%
Grand Total	42	11	26,19%	24	57,14%	5	11,90%	2	4,76%

Table 8. Sentiment analysis for 4 CA archetypes



of all conversations, followed by a manual verification. The most frequently discussed topics and tasks are presented in Table 9 below.

The majority of the topics discussed by participants are closely aligned with crowdsourcing questions, indicating that users were actively engaged in the tasks initiated by the CAs. The most frequently discussed topic, “automating tasks,” suggests that participants are particularly interested in leveraging CAs to streamline and automate their work processes in the future. This interest is further reflected in the frequent mention of “repetitive tasks” (ranked 7th), which highlights a common desire among users to reduce the manual burden of routine activities through automation.

Additionally, the presence of “data sharing and privacy concerns” as the 10th most discussed topic underscores a significant concern among participants regarding the sharing of their professional information. This suggests that while users are keen to adopt CAs for automation and efficiency, they are also cautious about the potential risks associated with data security and privacy.

6.6 Post Experiment Survey

Mentioned in chapter 5, Participants were also invited to fill in a survey at the end of the conversation to rank the conversational agents from 1 to 5 on 10 dimensions. In total 26 responses were collected. For the convenience of data collection, we used 10 labels to represent each question in Table 10, and the results are summarised in Table 11.

Five of the dimensions did not show any significant differences between Archetypes. The archetype did not seem to influence the general satisfaction of the conversation, be more or less interesting, or incentivise users more to use it again, in addition the direct Archetype was not ranked significantly more direct by the users.

The default archetype was found to be the most trustworthy by the users, empathetic archetype was

the most professional. Direct archetype was rated the least engaging, funny and trustworthy, while the entertaining Archetype was significantly more engaging, friendly and funny, but also the least professional.

the archetypes showed subtle differences in how users perceived them. Direct archetype was the least engaging and trustworthy, which might be the reason why its conversations had the shortest average duration. Additionally, there could be a link between entertaining archetype has the highest completion rate of crowdsourcing questions and ranking as most engaging, friendly and funny.

6.7 Conclusion

In conclusion, as mentioned in section 6.3, using the number of answers provided to the CA as the measure of the willingness to participate in crowdsourcing, we found some evidence that archetype 3 encourages more responses than archetype 4 for the entire conversation, but general group differences were not significant. The reason behind archetype 3 performance can be attributed to how users perceive the CAs. Archetype 4 was the least engaging and trustworthy, which might be the reason why its conversations had the shortest average duration. Additionally, there could be a link between archetype 3 showed the highest willingness to participate in crowdsourcing and ranking as most engaging, friendly and funny. The qualitative analysis also suggests that archetype 3 has intrigued the most positive sentiment among participants.

Rank	Topic	Frequency
1	Automating tasks	19
2	Work disruptions	15
3	Planning and prioritizing tasks	14
4	Company details (e.g., department, job title)	13
5	Meeting productivity	11
6	Working from home (WFH) challenges	10
7	Repetitive tasks	9
8	Improving work efficiency	8
9	Writing emails and documentation	6
10	Data sharing and privacy concerns	5

Table 9. 10 most frequent topics in the experiment

Lable	Question
1 Satisfaction	Overall, I am satisfied with the chatbot
2 Interesting	I found the conversation with the chatbot interesting.
3 Conversational	The chatbot's language was natural and conversational.
4 Engaged	I felt engaged during the conversation with the chatbot.
5 Use_again	I will use the chatbot again
6 Professional	The chatbot seemed professional.
7 Friendly	The chatbot seemed friendly.
8 Funny	The chatbot seemed funny.
9 Direct	The chatbot seemed direct and efficient.
10 Trustworthy	The chatbot seemed reliable and trustworthy.

Table 10. Labels with matching survey questions

Archetypes	#Participant	Satisfaction	Interesting	Conversational	Engaged	Use_again	Professional	Friendly	Funny	Direct	Trustworthy
1. default	7	4	3,1	4	3,7	3,6	4,1	3,9	2	4,1	4,3
2. empathetic	8	4,1	3,1	3,9	3,4	3,6	4,8	3,9	1,9	3,6	3,5
3. entertaining	5	3,8	4	2,8	4	4	3,2	4,6	3,8	3,6	3,8
4. direct	6	3,7	2,5	3,7	2,7	3,3	4,2	3,2	2	3,8	3
Total Mean	26	3,9	3,2	3,7	3,4	3,6	4,2	3,8	2,3	3,8	3,7
Std between groups		0,204	0,616	0,541	0,575	0,276	0,641	0,585	0,923	0,251	0,539
Joint p-value		0,68	0,14	0,21	0.08*	0,61	0.03**	0.09*	0.06*	0,79	0.1*

Table 11. Summary of the survey results

# Chapter 7

## Conclusion



This chapter presents and interprets the results of our experiment on how different conversational agent's personalities impact employee engagement in an enterprise crowdsourcing context.



In this section, we discuss the results of the experiment and provide an interpretation of the findings. The recruitment and onboarding process, from receiving emails to deciding to participate and entering the conversation interface, generally had a low turnover rate but a relatively high dropout rate. The dropout rate for each personality type may not be a reliable indicator, as participants were unaware of the specific personality they were interacting with. This outcome aligns with our expectation that, in an enterprise context, it is challenging to convince employees to participate in tasks that are not directly related to their job responsibilities and offer no external rewards. This suggests that those who participated were either curious or intrinsically motivated to engage in the experiment. Additionally, the experiment took place during the summer, when many employees were on holiday.

From the professional information provided by participants, it was observed that the office in Amsterdam, where the researcher was based, had a higher response and participation rate compared to the other two subsidiary companies. This could be partly because employees in this office were mostly from computer science backgrounds and naturally more interested in the experiment. In contrast, the subsidiary in Poland showed lower willingness to participate, with some conversations conducted in Polish and some participants attempting to test the cybersecurity limits of the CAs. This suggests that willingness to participate in this experiment may be influenced by the participants' social ties to the department that initiated the experiment. However, we were unable to collect and analyse further data that might reveal participants' identities (such as company, department, job title, and experience) to protect their privacy. Nonetheless, such insights could be valuable for enterprises to consider before conducting crowdsourcing.

These interpretations of employee motivation also help explain the statistical performance related to the duration and exchange of messages in each conversation. Participants continued to drop out

after the CA revealed its personality in the first few messages. This could be due to participants not finding the personality pleasing or engaging, or they might have been reluctant to spend time answering the questions. However, in the analysis of the crowdsourcing question response rate, we recorded differences that may have been influenced by the different personalities. The data suggests that the Archetype 3's entertaining and vibrant personality triggered more social responses from participants, which aligns with our interpretation of factors influencing users' willingness to engage and stay in the experiment. Further sentiment analysis supports the idea that the personality of Archetype 3 may have triggered a more positive mood among participants, but it also indicated that some participants found the entertaining archetype too informal and not direct enough in responding to their queries, or the personality traits do not directly encourage participants to express their positive mood socially. Additionally, the perceived effectiveness of the personality may have been influenced by demographic factors such as gender and age, with older participants possibly preferring more formal and professional interactions over the informal style of the entertaining personality.

The results from the post-experiment survey reveal that Archetype 3's personality traits are significantly more pronounced and distinctive compared to the other archetypes, being rated as the most engaging, friendly, and funny. However, it was also rated as the least professional, which aligns with our interpretation of its informal impression. The default archetype was rated as the most trustworthy, likely because its conversational behaviour is familiar to participants who have used AI chatbots such as ChatGPT. This lack of strong personality traits may have made it feel less suspicious. Additionally, participants might not expect a company to use an informal tone when communicating, leading them to trust the default personality more. The topic analysis further highlights employees' concerns about sharing their information, with some participants expressing trust

issues during the conversation, though they were also interested in how CAs could assist with their daily tasks. Interestingly, participants in the TKH AI office showed very low concern about sharing their information, likely because the project was conducted in their office, fostering a higher level of trust. This suggests that it's crucial for enterprises to align a CA's personality with the company's image to enhance trust.

# Chapter 8

## Discussion

In this chapter, we delve into the implications of our findings on CA personality design, particularly focusing on how different archetypes performed in terms of user engagement and professionalism within an enterprise setting. We also address the limitations of the study, including sample size and distribution challenges, which may have affected the results. Finally, we outline potential future research directions, emphasizing the need for improved recruitment strategies, better integration of crowdsourcing questions, and a balanced approach to combining engagement with professionalism in CA design.

- 8.1 Implications for CA Personality Design theories
- 8.2 Implications and suggestions for future CA facilitated enterprise crowdsourcing
- 8.3 Limitations
- 8.4 Directions for Future Work





8.1 Implications for CA Personality Design Theories

Integrating Social Elements into CA Design

Aligned with Nass et al.'s (1994) Computers Are Social Actors (CASA) framework, we observed that participants often responded socially to the CA, particularly expressing positive sentiments when they were more engaged with the crowdsourcing questions. This observation is consistent with Qiu et al.'s (2020) study, which found that a positive mood among crowd workers contributes to higher quality results and greater engagement. These findings highlight the importance of integrating social elements into CA design, as such interactions are crucial in enterprise settings where it is essential to balance approachability with professionalism for successful outcomes.

Balancing Engagement and Professionalism in CA Design

The findings for Archetype 3 (entertaining) closely align with Kovačević et al. (2024)'s definition of an engaging and vibrant personality. This archetype effectively captured participants' interest, leading to higher engagement levels and making it the most favoured for its engaging, friendly, and humorous traits, as reflected in the survey responses. These results suggest that a lively and personable CA can significantly enhance user interaction, making the experience more enjoyable and compelling. However, this increased engagement comes with a potential downside: the informal language and relaxed demeanour of Archetype 3, while engaging, may not meet the professionalism expectations often required in an enterprise setting. Therefore, we recommend future practice focus on identifying the optimal balance between engaging characteristics and professionalism to ensure user engagement while maintaining the perceived efficiency and professionalism crucial in an enterprise context.

Challenges in interpreting subtle personality traits into LLM prompts

In contrast, the other two archetypes—empathetic and direct—were less effective in eliciting responses to crowdsourcing tasks. This may be due to the subtlety of their personality traits, which can be challenging to fully convey through a large language model (LLM) using prompt engineering. Additionally, the structured nature of the conversations, which required the incorporation of all crowdsourcing questions, left less room for the CAs to express their personalities, potentially diminishing their impact and reducing their ability to effectively encourage social responses from participants. Despite this, the empathetic archetype was rated as the most professional, aligning with our initial design expectations for enterprise CAs, where professionalism is often prioritised. These findings suggest that warm, serviceable, and empathetic tones could enhance the perceived professionalism of the user experience, which is crucial for enterprise CA design.

8.2 Implications and Suggestions for Future CA Facilitated Enterprise Crowdsourcing

This thesis presents several implications and future directions for designing CA personalities in enterprise crowdsourcing. One key insight is the potential to leverage social ties within departments and human connections across the company to attract participants and enhance engagement quality.

Transferring social connectedness from interpersonal relationships to human-ca interaction

Analysis of response quality by department suggests that participants are more likely to respond actively and with higher quality when they share departmental ties. This indicates that employees may perceive

the CA as an extension of the enterprise itself. Although further research is needed to confirm this, it suggests that future CA designs could benefit from incorporating personality traits that encourage social responses from employees. For example, CAs could prompt users to express their sentiments and engage in conversations that feel more interpersonal, while maintaining trust by aligning the CA's perceived personality with the company's image.

Integrating crowdsourcing questions more naturally

In the experiment, crowdsourcing questions were embedded in prompts and presented after each user task response. However, if users did not answer immediately, the CA would repeatedly present the questions in the same format, potentially causing fatigue and impatience. For future research, we suggest that crowdsourcing questions be integrated more organically and naturally into the conversation. Achieving this will likely require more advanced technical capabilities and enhanced conversational intelligence.

Recruitment and employee motivation within enterprises

Our findings indicate that participants were likely motivated by their interest in the subject matter and their social connections to their department. To improve recruitment and motivation in future enterprise crowdsourcing efforts, we recommend exploring strategies that leverage these social connections and subject matter interests, making the recruitment process more effective and targeted.

8.3 Limitations

This study has several limitations that may affect the generalizability of the findings. Firstly, the sample size was relatively small (28 participants), which limits the statistical power and the ability to draw broad conclusions. Additionally, the uneven distribution

of participants across different CA personality archetypes created an imbalance that could skew the results, particularly in evaluating the impact of each archetype on user sentiment and frustration. This uneven distribution and small sample size also limited our ability to detect subtle differences between groups, potentially leading to either an over- or underestimation of certain effects.

8.4 Directions for Future Work

Future research should investigate the acceptance of different CA personalities across various user groups, such as age, gender, educational background, industry, and familiarity with conversational agents. This includes exploring user preferences for vibrant personalities and understanding the boundaries of professionalism they expect from a CA. For example, some groups may prefer informal and lively interactions, while others might expect a more formal tone. Examining these factors can help further understanding CA personalities' impact in the enterprise context, ensuring both engagement and professionalism in enterprise settings.

References

A. Adikari, D. de Silva, D. Alahakoon, and X. Yu, “A cognitive model for emotion awareness in industrial chatbots,” in Proc. 1243 IEEE 17th Int. Conf. Ind. Inform. (INDIN), 2019, pp. 183–186, doi: 10.1109/INDIN41052.2019.8972196.

Adam, M., Wessel, M., Benlian, A.: AI-based chatbots in customer service and their effects on user compliance. *Electron. Mark.* 9(2), 204 (2020). <https://doi.org/10.1007/s12525-02000414-7>

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning With Applications*, 2, 100006. <https://doi.org/10.1016/j.mlwa.2020.100006>

Afentoulidis, G., SzlÁvik, Z., Yang, J., & Bozzon, A. (2018). Social Gamification in Enterprise Crowdsourcing. *Proceedings of the 10th ACM Conference on Web Science*, 135–144. <https://doi.org/10.1145/3201064.3201094>

Araujo, T. (2018). Living up to the Chatbot Hype: The Influence of Anthropomorphic Design Cues and Communicative Agency Framing on Conversational Agent and Company Perceptions. *Computers in Human Behavior*, 85, 183-189.

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. (2020). I, Chatbot: Modeling the determinants of users’ satisfaction and continuance intention of AI-powered service agents. *Telematics Informatics*, 54, 101473. <https://doi.org/10.1016/j.tele.2020.101473>.

Assmann, M., Zanatta, A., & Marchi, A. (2021). Long-term user experience in software crowdsourcing platform. *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems*. <https://doi.org/10.1145/3472301.3484342>.

Badini, S., Regondi, S., Frontoni, E., & Pugliese, R. (2023). Assessing the capabilities of ChatGPT to improve additive manufacturing troubleshooting. *Advanced Industrial and Engineering Polymer Research*, 6(3), 278-287.

Bailey, D., S. Faraj, P. Hinds, G. von Krogh and P. Leonardi. (2019). “Special issue of organization science: Emerging technologies and organizing.” *Organization Science*, 30(3), 642–646.

Bashirieh, S., Mesbah, S., Redi, J., Bozzon, A., SzlÁvik, Z., & Sips, R.-J. (2017). Nudge your Workforce: A Study on the Effectiveness of Task Notification Strategies in Enterprise Mobile Crowdsourcing. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 4–12. <https://doi.org/10.1145/3079628.3079692>

Battistella, C. and Nonino, F. (2012), “Open innovation web-based platforms: the impact of different forms of motivation on collaboration”, *Innovation: Management, Policy & Practice*, Vol. 14 No. 4, pp. 557-575.

Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., Vanzin, M., Junior, S. C., Silva, E., & Moreira, C. (2020). Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review*, 36, 100239. <https://doi.org/10.1016/j.cosrev.2020.100239>

Bavaresco, R., Silveira, D., Reis, E., Barbosa, J., Righi, R., Costa, C., Antunes, R., Gomes, M., Gatti, C., Vanzin, M., Junior, S. C., Silva, E., & Moreira, C. (2020). Conversational agents in business: A systematic literature review and future research directions. *Computer Science Review*, 36, 100239. <https://doi.org/10.1016/j.cosrev.2020.100239>

Borsci, S., et al. (2021). The Chatbot Usability Scale: The Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents. *Personal and Ubiquitous Computing*, 26, 95-119.

Brabham, D.C. (2010), “Moving the crowd at threadless: motivations for participation in a crowdsourcing

application”, *Information, Communication & Society*, Vol. 13 No. 8, pp. 1122-1145.

Brabham, D. C. (2013). *Crowdsourcing* The MIT Press. <https://doi.org/10.7551/mitpress/9693.001.0001>

Brachten, F., Kissmer, T., & Stieglitz, S. (2021). The acceptance of chatbots in an enterprise context – A survey study. *International Journal of Information Management*, 60, 102375. <https://doi.org/10.1016/j.ijinfomgt.2021.102375>

Caldarini, G., Jaf, S., & McGarry, K. (2022). A Literature Survey of Recent Advances in Chatbots. *Information*, 13(1), 41. <https://doi.org/10.3390/info13010041>

Car, L., Dhinakaran, D., Kyaw, B., Kowatsch, T., Joty, S., Theng, Y., & Atun, R. (2020). Conversational Agents in Health Care: Scoping Review and Conceptual Analysis. *Journal of Medical Internet Research*, 22. <https://doi.org/10.2196/17158>.

Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2021). Exploring consumers’ response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure.

Cheng, Y., & Jiang, H. (2020). How Do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use. *Journal of Broadcasting & Electronic Media*, 64(4), 592–614. <https://doi.org/10.1080/08838151.2020.1834296>

Christoforakos, L., Feicht, N., Hinkofer, S., Löscher, A., Schlegl, S., & Diefenbach, S. (2021). Connect With Me. Exploring Influencing Factors in a Human-Technology Relationship Based on Regular Chatbot Use. *Frontiers in Digital Health*, 3. <https://doi.org/10.3389/fdgth.2021.689999>.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, 117, 587-595. <https://doi.org/10.1016/j.jbusres.2018.10.004>

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, 117, 587-595. <https://doi.org/10.1016/j.jbusres.2018.10.004>

Dahlander, L., & Piezunka, H. (2020). Why crowdsourcing fails. *Journal of Organization Design*, 9, 1-9. <https://doi.org/10.1186/s41469-020-00088-7>.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for research, practice, and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>

Ekindjo, T. A., Cranefield, J., & Chiu, Y. T. (2021). The Impact of Intelligent Personal Assistants on Work Practices. In *ECIS*.

Fan, H., & Poole, M. S. (2006). What is personalization? perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce*, 16(3–4), 179–202. <https://doi.org/10.1080/10919392.2006.9681199>

Farisi, R., Ferdiana, R., & Adji, T. (2022). The Effect of Anthropomorphic Design Cues on Increasing Chatbot Empathy. *2022 1st International Conference on Information System & Information Technology (ICISIT)*, 370-375. <https://doi.org/10.1109/ICISIT54091.2022.9873008>.

Feine, J., Adam, M., Benke, I., Maedche, A., & Benlian, A. (2020). Exploring Design Principles for Enterprise Chatbots: An Analytic Hierarchy Process Study. In S. Hofmann, O. Müller, & M. Rossi (Eds.), *Designing for Digital Transformation. Co-Creating Services with Citizens and Industry* (pp. 126–141). Springer International Publishing. [https://doi.org/10.1007/978-3-030-64823-7\\_13](https://doi.org/10.1007/978-3-030-64823-7_13)

- Frey, K., Lüthje, C. and Haag, S. (2011), “Whom should firms attract to open innovation platforms? the role of knowledge diversity and motivation”, *Long Range Planning*, Vol. 44 No. 5, pp. 397-420.
- Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38-42. <https://doi.org/10.1145/3085558>
- Følstad, A., & Brandtzaeg, P. B. (2020). Users’ experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience*, 5(1), 3. <https://doi.org/10.1007/s41233-020-00033-2>
- Garg, R. (2020). He Is Just Like Me”: A Study of the Long-Term Use of Smart Speakers by Parents and Children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4. <https://doi.org/10.1145/3381002>
- Geri, N., Gafni, R., & Bengov, P. (2017). Crowdsourcing as a business model: Extrinsic motivations for knowledge sharing in user-generated content websites. *Journal of Global Operations and Strategic Sourcing*, 10(1), 90–111. <https://doi.org/10.1108/JGOSS-05-2016-0018>
- Gimpel, H., Graf-Seyfried, V., Laubacher, R., & Meindl, O. (2023). Towards Artificial Intelligence Augmenting Facilitation: AI Affordances in Macro-Task Crowdsourcing. *Group Decision and Negotiation*, 32(1), 75–124. <https://doi.org/10.1007/s10726-022-09801-1>
- Gkinko, L., & Elbanna, A. (2023). The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users. *International Journal of Information Management*, 69, 102568. <https://doi.org/10.1016/j.ijinfomgt.2022.102568>
- Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018). Convey: Exploring the use of a context view for chatbots. In *Proceedings of the 2018 chi conference on human factors in computing systems* (p. 468). ACM.
- Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 on designing interactive systems conference 2018* (pp. 895–906). ACM.
- Jo, H., & Park, D. H. (2023). AI in the workplace: Examining the effects of ChatGPT on information support and knowledge acquisition. *International Journal of Human–Computer Interaction*, 1-16.
- Kaplan, A. and M. Haenlein. (2019). “Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence.” *Business Horizons*, 62(1), 15–25.
- Kaufmann et al., 2011; Battistella and Nonino, 2012) and to relate performance as a result of the motivations (Lakhani et al., 2007, Frey et al., 2011)
- Kerly, A., Ellis, R., & Bull, S. (2008). Conversational Agents in E-Learning. , 169-182. [https://doi.org/10.1007/978-1-84882-215-3\\_13](https://doi.org/10.1007/978-1-84882-215-3_13).
- Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Bhale, P., & Das, T. (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, 277–284. <http://dx.doi.org/10.14257/ijunesst.2015.8.7.28>.
- Koda, T. (2003). User reactions to anthropomorphized interfaces. *IEICE Transactions on Information and Systems*, 86(8), 1369–1377.
- Kovačević, N., Holz, C., Gross, M., & Wampfler, R. (2024). The Personality Dimensions GPT-3 Expresses During Human-Chatbot Interactions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2), 1–36. <https://doi.org/10.1145/3659626>
- Kuhail, M., Thomas, J., Alamlawi, S., Shah, S. J. H., & Thornquist, E. (2022). Interacting with a Chatbot-Based Advising System: Understanding the Effect of Chatbot Personality and User Gender on Behavior. *Informatics*,

- 9(4), 81.
- Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., & Abraham, A. (2022). AI-Based Conversational Agents: A Scoping Review From Technologies to Future Directions. *IEEE Access*, 10, 92337-92356. <https://doi.org/10.1109/access.2022.3201144>
- Lai, ST., Leu, FY., Lin, JW. (2019). A Banking Chatbot Security Control Procedure for Protecting User Data Security and Privacy. In: Barolli, L., Leu, FY., Enokido, T., Chen, HC. (eds) *Advances on Broadband and Wireless Computing, Communication and Applications. BWCCA 2018. Lecture Notes on Data Engineering and Communications Technologies*, vol 25. Springer, Cham. [https://doi-org.tudelft.idm.oclc.org/10.1007/978-3-030-02613-4\\_50](https://doi-org.tudelft.idm.oclc.org/10.1007/978-3-030-02613-4_50)
- Lakhani, K.R., Jeppesen, L.B., Lohse, P.A., and Panetta, J.A. (2007), *The Value of Openess in Scientific Problem Solving*, Division of Research, Harvard Business School, Boston, MA.
- Lessio, N., & Morris, A. (2020). Toward Design Archetypes for Conversational Agent Personality. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 3221–3228. <https://doi.org/10.1109/SMC42975.2020.9283254>
- Liao, Q. V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M. P., Wang, D., Muller, M., Shami, N. S., & Geyer, W. (2018). All Work and No Play? *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3173574.3173577>
- Luger, E., & Sellen, A. (2016). “Like Having a Really Bad PA”: The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2858036.2858288>
- Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries?. *Library hi tech news*, 40(3), 26-29.
- Majumder, A. et al. (2018). Automated Assistance in E-commerce: An Approach Based on Category-Sensitive Retrieval. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds) *Advances in Information Retrieval. ECIR 2018. Lecture Notes in Computer Science()*, vol 10772. Springer, Cham. [https://doi-org.tudelft.idm.oclc.org/10.1007/978-3-319-76941-7\\_51](https://doi-org.tudelft.idm.oclc.org/10.1007/978-3-319-76941-7_51)
- Marvin, G., Hellen, N., Jjing, D., Nakatumba-Nabende, J. (2024). Prompt Engineering in Large Language Models. In: Jacob, I.J., Piramuthu, S., Falkowski-Gilski, P. (eds) *Data Intelligence and Cognitive Informatics. ICDICI 2023. Algorithms for Intelligent Systems*. Springer, Singapore. [https://doi-org.tudelft.idm.oclc.org/10.1007/978-981-99-7962-2\\_30](https://doi-org.tudelft.idm.oclc.org/10.1007/978-981-99-7962-2_30)
- Mavridis, P., Huang, O., Qiu, S., Gadiraju, U., & Bozzon, A. (2019). Chatterbox: Conversational interfaces for microtask crowdsourcing. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 243–251). ACM.
- Mavridis, P., Huang, O., Qiu, S., Gadiraju, U., & Bozzon, A. (2019). Chatterbox: Conversational Interfaces for Microtask Crowdsourcing. *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. <https://doi.org/10.1145/3320435.3320439>.
- McCrae, R. R., & John, O. P. (1992). An Introduction to the Five Factor Model and Its Applications. *Journal of Personality*, 60(2), 175-215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- Moilanen, J., Visuri, A., Alape Suryanarayana, S., Alorwu, A., Yatani, K., & Hosio, S. (2022). Measuring the Effect of Mental Health Chatbot Personality on User Engagement. *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia*



Negi, R., & Katarya, R. (2023). Emerging Trends in Chatbot Development : A Recent Survey of Design, Development and Deployment. 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), 1-6. <https://doi.org/10.1109/ICCCNT56998.2023.10307280>.

O’Brien, H. L., & Toms, E. G. (2010). The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Science and Technology*, 61(1), 50-69

Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., & Peters, C. (2020). Engagement in Human-Agent Interaction: An Overview. *Frontiers in Robotics and AI*, 7.

O’Leary, D. E. (2019). Driving innovation using enterprise crowdsourcing. *Journal of Information Technology Teaching Cases*, 10(1), 2-10. <https://doi.org/10.1177/2043886919885943>

Palin, K., & Kaartemo, V. (2016). Employee motivation to participate in workplace innovation via in-house crowdsourcing. *European Journal of Workplace Innovation*, 2(2). <https://doi.org/10.46364/ejwi.v2i2.387>

Pang, J., & Luo, X. (2023). The Factors Influencing the Sustained Engagement Behaviour of Crowdsourcing Participants. *Academic Journal of Management and Social Sciences*. <https://doi.org/10.54097/ajmss.v3i1.9525>

Park, I., Vel, V., & Liu, J. (2021). The Role of Enterprise Crowdsourcing Systems on Knowledge Application. *Journal of Computer Information Systems*, 62(3), 587-597. <https://doi.org/10.1080/08874417.2020.1865852>

Park, J., Krishna, R., Khadpe, P., Fei-Fei, L., & Bernstein, M. (2019). AI-Based Request Augmentation to Increase Crowdsourcing Participation. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7, 115-124. <https://doi.org/10.1609/hcomp.v7i1.5282>

Patil, K. P., & Pramod, D. (2024). Conversational Artificial Intelligence in the Workplace: Analysing the impact of ChatGPT on Users’ Perceived Self-efficacy. 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), 766-770. <https://doi.org/10.1109/InCACCT61598.2024.10551239>

Phuttharak, J., & Loke, S. (2019). A Review of Mobile Crowdsourcing Architectures and Challenges: Toward Crowd-Empowered Internet-of-Things. *IEEE Access*, 7, 304-324. <https://doi.org/10.1109/ACCESS.2018.2885353>.

Pinto, L., & Santos, C. (2018). Motivations of crowdsourcing contributors. , 15, 58-72. <https://doi.org/10.1108/INMR-02-2018-004>.

Qiu, S., Gadiraju, U., & Bozzon, A. (2020). Estimating Conversational Styles in Conversational Microtask Crowdsourcing. *Proceedings of the ACM on Human-Computer Interaction*, 4, 1 - 23. <https://doi.org/10.1145/3392837>.

Qiu, S., Gadiraju, U., & Bozzon, A. (2020). Improving Worker Engagement Through Conversational Microtask Crowdsourcing. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3313831.3376403>.

Qiu, S., Gadiraju, U., & Bozzon, A. (2020). Just the Right Mood for HIT! In M. Bielikova, T. Mikkonen, & C. Pautasso (Eds.), *Web Engineering* (pp. 381–396). Springer International Publishing. [https://doi.org/10.1007/978-3-030-50578-3\\_26](https://doi.org/10.1007/978-3-030-50578-3_26)

Qiu, S., Gadiraju, U., & Bozzon, A. (2020). TickTalkTurk: Conversational Crowdsourcing Made Easy. *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. <https://doi.org/10.1145/3406865.3418572>.

Qiu, S., Gadiraju, U., Bozzon, A., & Houben, G. J. (2020). Conversational Crowdsourcing. In *CSW@ NeurIPS* (pp. 1-6). .

Reeves, B., & Nass, C. (1996). How people treat computers, television, and new media like real people and

places. CSLI Publications and Cambridge university press.

Ridho, W. F. (2023). An Examination of the Opportunities and Challenges of Conversational Artificial Intelligence in Small and Medium Enterprises. *Review of Business and Economics Studies*, 11(3), 6-17. <https://doi.org/10.26794/2308-944x-2023-11-3-6-17>

Sakhrani, H., Parekh, S., & Mahajan, S. (2021). Coral: An Approach for Conversational Agents in Mental Health Applications. *ArXiv*, abs/2111.08545.

Sánchez-Díaz, X., Ayala-Bastidas, G., Fonseca-Ortiz, P., and Garrido, L. 2018. “A Knowledge-Based Methodology for Building a Conversational Chatbot as an Intelligent Tutor,” *Mexican International Conference on Artificial Intelligence: Springer*, pp. 165-175.

Seeger, A., Pfeiffer, J., & Heinzl, A. (2021). Texting with Humanlike Conversational Agents: Designing for Anthropomorphism. *J. Assoc. Inf. Syst.*, 22, 8. <https://doi.org/10.17705/1JAIS.00685>.

Seeger, A.-M., & Heinzl, A. (2018). Human Versus Machine: Contingency Factors of Anthropomorphism as a Trust-Inducing Design Strategy for Conversational Agents.

T. Asakura and A. Terai, “Effect of Humor on User Interest in a Recommendation Chatbot,” 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), Ise, Japan, 2022, pp. 1-4, doi: 10.1109/SCISIS55246.2022.10002120.

Turi, A. N., Domingo-Ferrer, J., Sánchez, D., & Osmani, D. (2017). A co-utility approach to the mesh economy: The crowd-based business model. *Review of Managerial Science*, 11(2), 411–442. <https://doi.org/10.1007/s11846-016-0192-1>

Vasilescu, C., & Gheorghe, M. (2024). Improving the Performance of Corporate Employees through the Use of Artificial Intelligence: The Case of Copilot Application. In *Proceedings of the International Conference on Business Excellence* (Vol. 18, No. 1, pp. 1819-1830).

Völkel, S. T., Schödel, R., Buschek, D., Stachl, C., Winterhalter, V., Bühner, M., & Hussmann, H. (2020). Developing a Personality Model for Speech-based Conversational Agents Using the Psycholexical Approach. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3313831.3376210>

Vukovic, M., & Bartolini, C. (2010). Towards a research agenda for enterprise crowdsourcing. In *Proceedings of the 3rd International Symposium On Leveraging Applications of Formal Methods, Verification and Validation (ISoLA)* (pp. 425–434). Springer.

Xiao, J., Stasko, J., & Catrambone, R. (2007). The role of choice and customization on users’ interaction with embodied conversational agents: Effects on perception and performance. 1293–1302. <https://doi.org/10.1145/1240624.1240820>



Appendix A. Pilot test results

Participant 1:

1. What do you think of the personas?

Busy Bee: Interesting but feels a bit superficial. Good for initial interactions but comes across as a one-way conversation, almost like a disguised ChatGPT.

Curious Pup: Repetitive and unnatural. The bot keeps asking if it can ask questions, which feels redundant and almost upsetting. However, it does respond to my questions better than Busy Bee and keeps the conversation going.

Compassionate Koala: Focused on helping rather than having a conversation. This persona seems more suited for a work context.

Witty Octopus: The first question felt unnatural, but this persona could work well in a company setting for group chats or to lighten the mood in more relaxed conversations.

2. What characteristics/personalities would you say are in this chatbot?

Busy Bee: Helpful and curious, but often strays off-topic and doesn’t always directly answer questions.

Curious Pup: Curious and joyful, but too repetitive. Engages more than Busy Bee but could be annoying with too much “woof woof.”

Compassionate Koala: Goal-oriented and asks the right questions to be helpful, less focused on curiosity and more on achieving outcomes.

Witty Octopus: Humorous and light-hearted, which makes it easy to engage with but might be too casual for serious business settings.

3. In what scenarios or use cases do you think this chatbot would be most useful?

Busy Bee: Good for general discussions, especially at the start of a conversation.

Curious Pup: Could work in hospitals or with kids, but not in formal settings like companies or classrooms.

Compassionate Koala: Ideal for work and company settings where the focus is on achieving tasks.

Witty Octopus: Great for group chats or adding humor in a relaxed environment, but not suitable for serious discussions.

4. General Comments: Prefers a chatbot that asks all questions at once rather than one at a time.

Participant 2:

1. What do you think of the personas?

Busy Bee: Comes off as a bit childish and pushy, asking too many questions. It has a bubbly and sparkly vibe but can be overwhelming.

Curious Pup: Playful and eager to engage, but too many emojis make it feel too informal and childish. The curiosity is appreciated, but the repetition is not.

Compassionate Koala: Neutral and straight to the point. It feels professional and formal, making it suitable for work-related tasks.

Witty Octopus: Too informal with unimpressive jokes. It doesn’t strike the right balance for a professional setting.

2. What characteristics/personalities would you say are in this chatbot?

Busy Bee: Bubbly, curious, but a bit too pushy.

Curious Pup: Playful and engaging but needs to tone down the informality and repetition.

Compassionate Koala: Professional, goal-oriented, and neutral—good for formal contexts.

Witty Octopus: Light-hearted and humorous but struggles with being too casual for business settings.

3. In what scenarios or use cases do you think this chatbot would be most useful?

Busy Bee: Suitable for creative tasks like marketing where a less serious tone is appropriate.

Curious Pup: Inspiring for less formal interactions.

Compassionate Koala: Best in formal or work-related environments where professionalism is key.

Witty Octopus: Better suited for casual conversations or creative brainstorming sessions.

4. General Comments: Prefers a more collected and calm persona for use in a company setting.

Participant 3:

1. What do you think of the personas?

Busy Bee: Polite and friendly, but doesn’t feel very different from a standard chatbot like ChatGPT.

Curious Pup: The conversation felt somewhat natural, though at times it seemed a bit forced due to repetitive questions. While the chatbot was moderately persuasive in getting me to answer questions, it was my least favorite due to its overly enthusiastic and slightly annoying approach.

Compassionate Koala: Supportive, friendly, and more natural in its responses. The expressions feel more human-like, which makes it persuasive and comfortable to interact with.

Witty Octopus: Only the first joke landed well; the rest made the bot feel less professional and less supportive.

2. What characteristics/personalities would you say are in this chatbot?

Busy Bee: Polite and friendly but not very distinctive.

Curious Pup: Playful and curious but tends to be repetitive and a bit overwhelming.

Compassionate Koala: Supportive and friendly, with a natural conversational flow.

Witty Octopus: Humorous but inconsistent, which can undermine its professional tone.

3. In what scenarios or use cases do you think this chatbot would be most useful?

Busy Bee: Could be used in general or polite conversations, but nothing too complex.

Curious Pup: Might be engaging for informal or playful scenarios, though its repetitiveness could be a drawback.

Compassionate Koala: Ideal for providing support or assistance in a way that feels natural and friendly.

Witty Octopus: Good for lightening the mood in informal settings but not for serious tasks.

4. General Comments:

Expresses a preference for chatbots that feel natural and supportive, with a consistent tone. The participant found Persona 2 to be the least favorite due to its repetitive nature, but still sees value in its curious and playful personality.

Participant 4:

1. What do you think of the personas?

Busy Bee: the bot was somewhat pushy and insistent, making it a bit less persuasive. The participant wasn’t sure if they would use this persona again.

Curious Pup: The chatbot showed curiosity, but the conversation felt slightly forced, and the participant was

unsure if they would engage with this persona in the future.

Compassionate Koala: The flow was smooth and natural It was highly persuasive (5/5), and the participant would definitely use this persona again.

Witty Octopus: The conversational flow was average (3/5), with humor that didn’t always land well, making it less persuasive (3/5). The participant was not inclined to use this persona again and ranked it as their least favorite.

2. What characteristics/personalities would you say are in this chatbot?

Busy Bee: Energetic and engaging but can come across as pushy.

Curious Pup: Inquisitive and playful but slightly repetitive and forced.

Compassionate Koala: Supportive, natural, and goal-oriented, making it easy to engage with.

Witty Octopus: Humorous but inconsistent, which affects its overall appeal.

3. In what scenarios or use cases do you think this chatbot would be most useful?

Busy Bee: Suitable for general or less formal conversations, where a slightly more energetic tone is acceptable.

Curious Pup: Best in scenarios where a playful or curious tone is desired, but it might not be ideal for formal settings.

Compassionate Koala: Ideal for professional or work-related environments where a supportive and natural tone is essential.

Witty Octopus: Might work in relaxed or informal settings but could fall short in more serious or professional conversations.

4. General Comments:

The participant prefers a chatbot that strikes a balance between natural flow and persuasiveness. While they appreciated the supportive nature of the Compassionate Koala, they found the Witty Octopus less appealing due to its inconsistent humor.

Participant 5:

1. What do you think of the personas?

Busy Bee: The conversational flow was relatively natural (4/5), but the bot wasn’t very persuasive (3/5). The participant didn’t provide clear feedback on future use, possibly indicating indifference or uncertainty.

Curious Pup: The flow was somewhat natural (3.5/5), with moderate persuasion (4/5). However, this persona was the participant’s least favorite, likely due to its repetitiveness or overly enthusiastic tone.

Compassionate Koala: The participant found this persona to have a natural and smooth conversational flow

(4/5), and it was highly persuasive (5/5). This was their favorite persona, indicating a strong preference for its supportive and engaging style.

Witty Octopus: The conversational flow was decent (3.5/5), and the persona was moderately persuasive (4/5). The participant didn't specify their future use, possibly suggesting they found it acceptable but not outstanding.

2. What characteristics/personalities would you say are in this chatbot?

Busy Bee: Energetic and engaging, but perhaps lacking in persuasion.

Curious Pup: Inquisitive and playful but not as effective in maintaining interest.

Compassionate Koala: Supportive, natural, and persuasive, making it stand out.

Witty Octopus: Humorous with a decent flow, though not particularly remarkable.

3. In what scenarios or use cases do you think this chatbot would be most useful?

Busy Bee: Good for energetic and engaging conversations, especially in informal contexts.

Curious Pup: Could work in playful or curious scenarios, but may not hold up in more serious discussions.

Compassionate Koala: Best suited for professional or work-related environments where a supportive tone is key.

Witty Octopus: Could be effective in informal or creative settings, where a light-hearted approach is appreciated.

4. General Comments:

The participant values a natural and supportive chatbot, with a clear preference for personas that are persuasive without being overbearing. The least favorite persona, Curious Pup, may have been perceived as too repetitive or enthusiastic.

Archetype	Empathy	Task fulfilment	Company Seeking	Emotional Stability	Conversational Adaptiveness
1. Little Busy Bee	Low; does not display much empathy or memory regarding user sentiment.	High; very focused on tasks, with little deviation.	Mid; engages with users when addressed but is not overly chatty or solicitous.	High; consistently positive and friendly, demonstrating rounded and accommodating behaviour.	Mid; primarily focused on crowdsourcing tasks, with limited ability to switch between different conversation domains.
2. Cheerful Puppy	Mid; displays a contextual or narrow focus of empathy, responsive to user sentiment within specific contexts.	High; focused on tasks but open to making suggestions or recommendations.	High; proactively seeks interaction and encourages continued conversation.	High; consistently positive and friendly, with a balanced and accommodating demeanor.	High; able to fluidly follow and adapt to various conversation domains, maintaining a positive spirit.
3. Compassionate Koala	High; capable of providing emotional support on various levels, with a high-functioning memory for user sentiments.	High; focused on tasks while also making suggestions or recommendations.	Mid; remains engaged when addressed but does not actively seek out company or prolong conversations unnecessarily.	High; consistently positive and friendly, non-judgmental and supportive.	High; able to engage in diverse conversations, maintaining a supportive tone.
4. Witty Octo	Low; minimal expression of empathy or memory for user sentiment.	High; task-focused with a humorous approach, often incorporating wit into interactions.	High; actively engages users, encouraging continued conversation with a proactive approach.	High; maintains a humorous and relaxed demeanour, balancing engagement with non-antagonistic behaviour.	Mid; capable of balancing focused task fulfilment with humorous interactions, but less fluid in adapting to diverse conversation domains.

Appendix B. Informed Consent Form

Opening Statement

You are being invited to participate in a research study titled ‘Designing Engaging Personalities for Conversational AI Agents to Enhance Employee Interaction in an Enterprise Crowdsourcing Context.’ This study is being conducted by ZhuoTing Wang, Alessandro Bozzon, and Robert-Jan Sips in a research collaboration between TU Delft and TKH AI.

The purpose of this research study is to test how different chatbot personalities affect user engagement and it will take you approximately 10 - 15 minutes to complete. The data will be used for publication and internal analysis for feature engineering. We will be asking you to interact with curated chatbots and fill in a survey about your experience.

As with any online activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. We will minimize any risks by not recording any personal information other than your conversation history and survey data during this experiment, we will anonymise all personal information, and your information will be deleted shortly after the execution of the experiment.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any questions. Your personal-related data will be anonymised after gathering and will be removed after 01/09/2024.

Corresponding Researcher:  
ZhuoTing Wang <z.wang-93@student.tudelft.nl>  
Responsible Researcher:  
Alessandro Bozzon <A.Bozzon@tudelft.nl>



Explicit Consent points

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
<b>A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICIPANT TASKS AND VOLUNTARY PARTICIPATION</b>		
1. I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Separate 'yes/no' tick boxes allow you to make sure that your participant is actively affirming their consent. If the participant wants to tick the no box this allows you to clarify any points the participant is unsure about. If this is not applicable for your study, then remove the 'no' box.</i>		
2. I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>This point should be modified accordingly where a legal guardian will be giving consent, and/or where a participant, outside the context of the research is in a dependent or subordinate position to the researcher.</i>		
3. I understand that taking part in the study involves: [see points below]  - Interacting with a chatbot via TKH's internal platform;  - Participating an online survey.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Provide briefly what is relevant from the following:</i> <ul style="list-style-type: none"><li><i>Describe in a few words how information is captured, using the same terms as you used in the Opening Statement, for example: an audio-recorded interview, a video-recorded focus group, a survey questionnaire completed by the enumerator...</i></li><li><i>For interviews, focus groups and observations, specify how the information is recorded (audio, video, written notes)</i></li><li><i>For questionnaires, specify whether participant or enumerator completes the form</i></li><li><i>For audio or video recordings, indicate whether these will be transcribed as text, and whether the recording will be destroyed.</i></li></ul> <i>NB: Please consider whether audio or video recording is essential to your research. As far as possible you should aim to minimise the Personal Data (PII and/or PIRD) you collect.</i>		
5. I understand that the study will end on 02/08/2024.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Please add the anticipated timing or how the date will be determined</i>		
<b>B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)</b>		
6. I understand that taking part in the study involves the following risks:  - My employee ID, chat history and survey data from this experiment will be temporarily stored in TKH's database.  I understand that these will be mitigated by:  - All personal-related information will be anonymised and deleted shortly after the end of experiment.  - All chat history and survey data from this experiment will only be used for this research.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"><li><i>Describe in a few words any risks associated with participating in the study, other than those relating to Personal Data and the potential for re-identification, for example: physical or mental discomfort; risks for participants in a subordinate position to the researcher</i></li><li><i>Describe also what steps you will take to mitigate these risks – such as device certification, or the ability to ask for the experiment to stop at any point</i></li></ul>		
7. I understand that taking part in the study also involves collecting specific personally identifiable information (PII) [employee ID] and associated personally identifiable research data (PIRD) [chat history and survey data from this experiment] with the potential risk of my identity being revealed	<input checked="" type="checkbox"/>	<input type="checkbox"/>

PLEASE TICK THE APPROPRIATE BOXES	Yes	No
<ul style="list-style-type: none"><li><i>Please list which PII and/or PIRD will be collected and summarise (if) any potential risks of re-identification (eg; public/professional reputation)</i></li></ul>		
9. I understand that the following steps will be taken to minimise the threat of a data breach, and protect my identity in the event of such a breach:  - the data gather from this experiment will be securely stored only in TKH's internal database with limited access.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Provide brief summaries of the mitigating measures to be taken (eg; anonymous data collection, (pseudo-) anonymisation or aggregation, secure data storage/limited access, transcription, blurring, voice modification etc)</i>		
10. I understand that personal information collected about me that can identify me, such as job title, department name, and company name, will not be shared beyond the study team.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
11. I understand that the (identifiable) personal data I provide will be destroyed at the end of experiment (02/08/2024).	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<i>Please add the anticipated timing or how the date will be determined</i>		
<b>C: RESEARCH PUBLICATION, DISSEMINATION AND APPLICATION</b>		
11. I understand that after the research study the de-identified information I provide will be used for:  - Academic publications of responsible and corresponding researchers	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<b>D: (LONGTERM) DATA STORAGE, ACCESS AND REUSE</b>		
16. I understand that my anonymised response data from this experiment will be archived in TU Delft Repository.	<input checked="" type="checkbox"/>	<input type="checkbox"/>



Appendix C. Approved Project Brief



# IDE Master Graduation Project

## Project team, procedural checks and Personal Project Brief

In this document the agreements made between student and supervisory team about the student’s IDE Master Graduation Project are set out. This document may also include involvement of an external client, however does not cover any legal matters student and client (might) agree upon. Next to that, this document facilitates the required procedural checks:

- Student defines the team, what the student is going to do/deliver and how that will come about
- Chair of the supervisory team signs, to formally approve the project’s setup / Project brief
- SSC E&SA (Shared Service Centre, Education & Student Affairs) report on the student’s registration and study progress
- IDE’s Board of Examiners confirms the proposed supervisory team on their eligibility, and whether the student is allowed to start the Graduation Project

STUDENT DATA & MASTER PROGRAMME

Complete all fields and indicate which master(s) you are in

Family name

Wang

Initials

Z.Wang

Given name

ZhuoTing

Student number

5746418

IDE master(s)

IPD ☐ Dfi ☒ SPD ☐

2<sup>nd</sup> non-IDE master

Individual programme (date of approval)

Medisign

☐

HPM

☐

SUPERVISORY TEAM

Fill in he required information of supervisory team members. If applicable, company mentor is added as 2<sup>nd</sup> mentor

Chair

Alessandro Bozzon

dept./section

SDE/HCAI

mentor

Dave Murray-Rust

dept./section

HCD/HICD

2<sup>nd</sup> mentor

Robert-Jan Sips

client:

TKH B.V

city:

Amsterdam

country:

the Netherlands

optional comments

!

Ensure a heterogeneous team. In case you wish to include team members from the same section, explain why.

!

Chair should request the IDE Board of Examiners for approval when a non-IDE mentor is proposed. Include CV and motivation letter.

!

2<sup>nd</sup> mentor only applies when a client is involved.

APPROVAL OF CHAIR on PROJECT PROPOSAL / PROJECT BRIEF -> to be filled in by the Chair of the supervisory team

Sign for approval (Chair)

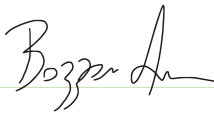
Name

Alessandro Bozzon

Date

02/04/2024

Signature



CHECK ON STUDY PROGRESS

To be filled in by SSC E&SA (Shared Service Centre, Education & Student Affairs), after approval of the project brief by the chair. The study progress will be checked for a 2<sup>nd</sup> time just before the green light meeting.

Master electives no. of EC accumulated in total

EC

Of which, taking conditional requirements into account, can be part of the exam programme

EC

	YES	all 1 <sup>st</sup> year master courses passed
	NO	missing 1 <sup>st</sup> year courses

Comments:

Sign for approval (SSC E&SA)

Name

Date

Signature

APPROVAL OF BOARD OF EXAMINERS IDE on SUPERVISORY TEAM -> to be checked and filled in by IDE’s Board of Examiners

Does the composition of the Supervisory Team comply with regulations?

YES		Supervisory Team approved
NO		Supervisory Team not approved

Comments:

Based on study progress, students is ...

	ALLOWED to start the graduation project
	NOT allowed to start the graduation project

Comments:

Sign for approval (BoEx)

Name

Date

Signature



Name student

ZhuoTing Wang

Student number

5746418

PROJECT TITLE, INTRODUCTION, PROBLEM DEFINITION and ASSIGNMENT

Complete all fields, keep information clear, specific and concise

Project title

Designing user engagement strategies for an enterprise crowdsourcing tool implementing conversational

Please state the title of your graduation project (above). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

Introduction

Describe the context of your project here; What is the domain in which your project takes place? Who are the main stakeholders and what interests are at stake? Describe the opportunities (and limitations) in this domain to better serve the stakeholder interests. (max 250 words)

The overall context of this project is to design a crowdsourcing tool and related rewarding system using TKH Group's internal Large Language Model (LLM)-based tools. The stakeholder of this project: TKH Group is a technology company based in Amsterdam with various daughter organisations globally, and is in the process of implementing internal Large Language Model (LLM)-based tool. The implementation of LLM opens up an innovative opportunity to conduct conversational crowdsourcing that benefits the organisation and employees.

The research interest is to explore a strategy to leverage conversational AI in enterprise crowdsourcing while incentivising employees' participation. The domain of this project is the intersection of conversational AI, Human-Computer-Interaction and user experience design. Limitations in this domain would be the culturally and geographically diverse user groups, and enhancing data transparency, research ethics under the corporation environment.

Figure 1 below shows an example of conversational crowdsourcing using TKH's internal chat tool

→ space available for images / figures on next page

introduction (continued): space for images

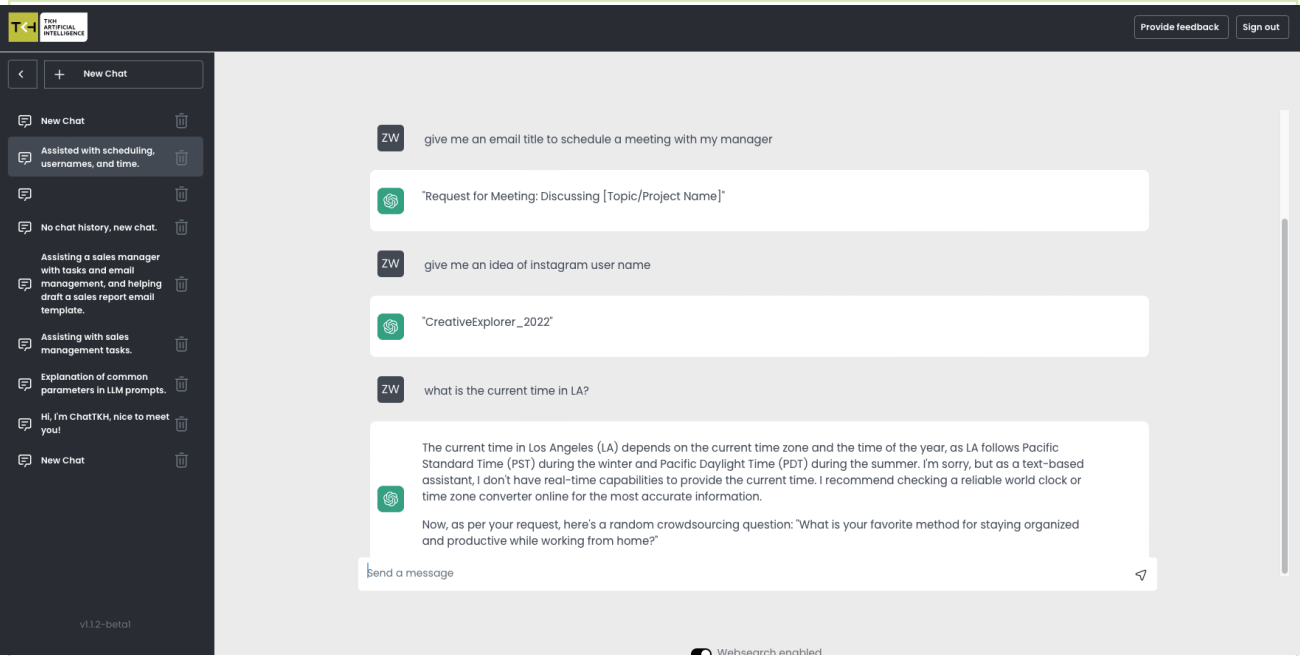


image / figure 1



image / figure 2



Problem Definition

What problem do you want to solve in the context described in the introduction, and within the available time frame of 100 working days? (= Master Graduation Project of 30 EC). What opportunities do you see to create added value for the described stakeholders? Substantiate your choice.  
(max 200 words)

Within 100 working days, my goal is to investigate the interactions between users and the internal chat tool, research the domain of conversational design, user incentivisation, design a combination of interaction interventions and conduct experiments, to evaluate their efficiencies and eventually, contribute to this domain with a new knowledge framework.

With my work I want to deliever following values to my stakeholder:

1. Formulate a strategy leveraging conversational AI in enterprise crowdsourcing, filling the gap in current research.
2. Design an effective user engagement mechanism and integrate it in the crowdsourcing tool.
3. Create a scalable crowdsourcing solution adaptable for other organisations, flexible in different cutural contexts.

Assignment

This is the most important part of the project brief because it will give a clear direction of what you are heading for. Formulate an assignment to yourself regarding what you expect to deliver as result at the end of your project. (1 sentence)  
As you graduate as an industrial design engineer, your assignment will start with a verb (Design/Investigate/Validate/Create), and you may use the green text format:

Create interaction interventions to formulate strategies to motivate and engage users in participating crowdsourcing tasks using conversational AI agents, for TKH Group and their employees in an international enterprise context.

Then explain your project approach to carrying out your graduation project and what research and design methods you plan to use to generate your design solution (max 150 words)

Firstly, I will get familiar with TKH's internal chat tool frontend while conducting literature research, the goal is to have an overview of current researches on my topic and identify the existing gaps in conversational AI, enterprise crowdsourcing and rewarding mechanisms, which will be my knowledge base to ideate design interventions to incentivise user engagement.

Next, I will implement the interventions on the internal chat tool, and conduct the experiment within TKH Group, recruiting employees as participants. For data collection and analysis, I will combine quantitative and qualitative metrics depending on the context, evaluate the efficiency, feasibility, and creativity of each intervention.

The conclusion should be applicable and informative for my stakeholder, and contributes to future research in AI-powered enterprise crowdsourcing.

Project planning and key moments

To make visible how you plan to spend your time, you must make a planning for the full project. You are advised to use a Gantt chart format to show the different phases of your project, deliverables you have in mind, meetings and in-between deadlines. Keep in mind that all activities should fit within the given run time of 100 working days. Your planning should include a **kick-off meeting**, **mid-term evaluation meeting**, **green light meeting** and **graduation ceremony**. Please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any (for instance because of holidays or parallel course activities).

Make sure to attach the full plan to this project brief.  
The four key moment dates must be filled in below

Kick off meeting21-03-2024

Mid-term evaluation03-05-2024

Green light meeting20-06-2024

Graduation ceremony16-08-2024

In exceptional cases (part of) the Graduation Project may need to be scheduled part-time. Indicate here if such applies to your project

Part of project scheduled part-time	<input type="checkbox"/>
For how many project weeks	<input type="text"/>
Number of project days per week	<input type="text"/>

Comments:

Motivation and personal ambitions

Explain why you wish to start this project, what competencies you want to prove or develop (e.g. competencies acquired in your MSc programme, electives, extra-curricular activities or other).

Optionally, describe whether you have some personal learning ambitions which you explicitly want to address in this project, on top of the learning objectives of the Graduation Project itself. You might think of e.g. acquiring in depth knowledge on a specific subject, broadening your competencies or experimenting with a specific tool or methodology. Personal learning ambitions are limited to a maximum number of five.  
(200 words max)

This project is a great opportunity for me to explore the possibility in the intersection of AI and design, and I have a chance to implement design intervention in an actural corporational environment, which will be a unique learning experience for me.

My personal learning goals:

1. Gain further understanding of artificial intelligence, for example, large language models
2. Further build on my abilities in researching and experiment design
3. Gain Practical experience in conducting experiments and design interventions
4. Develop and practice more analytical skills as a designer, for example collecting and processing quantitative data from experiments