When Personas Talk to You: Evaluating the Evolution of User Personas from Static Profiles to Conversational User Interfaces

Ilkka Kaate
University of Turku
Marketing
Turku, Finland
iokaat@utu.fi

Joni Salminen
University of Vaasa
Vaasa, Finland
jonisalm@uwasa.fi

Soon-Gyo Jung
Hamad Bin Khalifa University
Qatar Computing Research Institute
Doha, Qatar
sjung@hbku.edu.qa

Trang Thi Thu Xuan
University of Vaasa
School of Marketing and Communication
Vaasa, Finland
FPT University
Swinburne Vietnam
Hanoi, Hanoi, Vietnam
x9036166@student.uwasa.fi

Jinan Y. Azem
Qatar Computing Research Institute
Doha, Qatar
jazem@hbku.edu.qa

João M. Santos
Instituto Universitário de Lisboa (ISCTE-IUL)
Lisbon, Portugal
jmcsm@iscte.pt

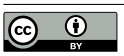
Bernard J Jansen
Hamad Bin Khalifa University
Qatar Computing Research Institute
Doha, Qatar
bjansen@hbku.edu.qa

Abstract

The development of persona systems provides a possibility for end users to interact with different persona modalities. In a 54-participant randomized controlled experiment, we compare two persona interaction modalities, document and dialogue personas, both generated using AI approaches from survey data. Overall, dialogue personas appear to be perceived more favorably than document personas. However, document personas exhibit a wider range of perceptions, suggesting that experiences with document personas are more polarizing among users. The document personas had higher transparency and were perceived as more complete, but the task completion was perceived as more difficult, although the task success rate was higher. The dialogue personas were perceived as more usable, with a higher System Usability Scale score, and more enjoyable. Our findings provide critical insights into the increasingly important area of persona interaction modalities and the broad paradigm of human-persona interaction.

CCS Concepts

• **Human-centered computing** → Human computer interaction (HCI).



This work is licensed under a Creative Commons Attribution 4.0 International License.

DIS '25, Funchal, Portugal

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1485-6/2025/07
https://doi.org/10.1145/3715336.3735676

Keywords

Interactive Personas, Persona Systems, User Representation, Usability

ACM Reference Format:
Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, Trang Thi Thu Xuan, Jinan Y. Azem, João M. Santos, and Bernard J Jansen. 2025. When Personas Talk to You: Evaluating the Evolution of User Personas from Static Profiles to Conversational User Interfaces. In *Designing Interactive Systems Conference (DIS '25), July 05–09, 2025, Funchal, Portugal*. ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/3715336.3735676

1 Introduction

User personas are detailed narratives of user archetypes [56] facilitating understanding of end users among designers, managers, and other stakeholders [25], thereby enhancing the development of interactive systems that genuinely resonate with key end-user groups [17]. Although the persona technique was introduced in the 1990s, the rise of *interactive personas* in the wake of web interaction techniques [38, 53, 67] marks a milestone in human-computer interaction (HCI) and user-centered design (UCD). Interactive personas afford users interaction techniques like search, filter, and sort [38–40], denoting a clear separation from "static file" personas [34].

In general, interactive personas utilize technology to enhance the interaction between designers, managers, and the persona [34], making personas more interactive [19, 56]. Figure 1 illustrates instances of these interaction techniques, a document-style conventional persona profile (CPP) with a chat-style high-information dialogue persona (DP).

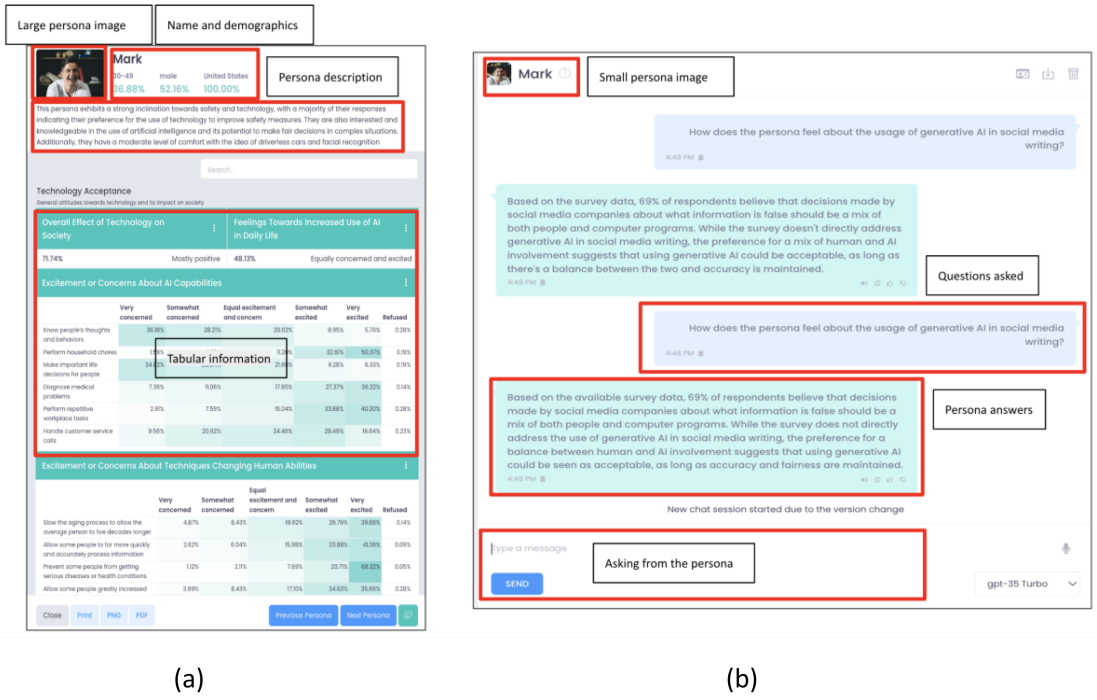


Figure 1: A comparison of how an analytics user interface (UI) (CPP) and a dialogue-based persona deliver information about personas differently. The (a) CPP uses graphics to show key persona details (such as image, name, demographics, and description) all at once. In contrast, the (b) dialogue-based persona (DP) only shows a name and a small image upfront, requiring users to ask for more details through conversation. The CPP provides its persona information directly in tables and text descriptions.

When user personas are created from high-information datasets, the outcomes are *high-information user personas* (HIUPs). Built with machine learning (ML) and artificial intelligence (AI), *interactive HIUPs* have emerged as a pivotal metaphor for user segments [6, 52, 53, 83]. Interactive HIUPs offer a persona user the possibility to browse large amounts of persona information interactively. However, there are many possible interaction practices, enabling various interaction approaches. By 'interaction', we refer to exchanges taking place between the HIUP system and the HIUP user. With current AI technologies, these interactions increasingly rely on natural language [9, 19, 44].

From the HCI perspective, a crucial motivating question for our study is: What HIUP interaction modalities should be offered to HIUP users so that they can productively engage with the HIUP toward learning more about user groups? While AI personas are increasingly used, the impact of CPP versus DP HIUP modalities on user behavior and HIUP perception remains underexplored. This study addresses this knowledge gap, offering empirical insights into how interaction modality in HIUPs shapes user behavior and HIUP perceptions.

To this end, *persona interaction modalities* are defined as modes of persona presentation that affect the exchanges between the end user and the system. The interaction between a persona and its user can take place via voice- or chat [18, 28, 77], especially those using *large language models* (LLMs). The incorporation of LLMs into the persona generation process has gained interest from HCI scholars [1, 20, 84]. Consequently, these interaction techniques represent a new line of study for human persona interaction as persona research lacks knowledge on the *dialogue modality* compared to the traditional persona presentation style of "profiles" [17].

Essentially, HIUP perceptions and the HIUP's usability contribute to the user's task perceptions and behavior [71]; thus, they are important factors for understanding the overall persona UX and related concepts. While prior research has found that realism (for HIUPs) and anthropomorphism (for chatbots) or "human-likeness" are perceptions that play a central role in how users perceive and interact with these entities [7, 21], no extant studies specifically investigate the trade-offs involved in using DPs versus using CPPs. We simply do not know enough about the strengths and advantages of CPP or DP interaction modality. To address this research gap, we put forth three research questions (RQ):

- **RQ1:** How does the HIUP interaction modality (document or dialogue) affect users' experience, behavior, usability, learning and perceptions of the HIUP?
- **RQ2:** Under what conditions is document modality more preferable than dialogue modality, and vice versa?
- **RQ3:** How does the HIUP modality affect task completion?

We address these RQs via a randomized controlled trial (i.e., an experiment) in which participants use HIUPs from both modalities

to accomplish an information retrieval task on a fictitious (but realistic) work task scenario. Overall, our findings shed light on the broader question of how generative AI technologies can help in human persona interaction toward developing interactive persona systems that effectively and efficiently serve stakeholders' information needs by offering optimal interfaces to access the user data that the personas represent.

2 Related Work

Personas represent sets of users who share common characteristics, goals, and behaviors when interacting with software, websites, or applications [17, 56]. Personas are fictional characters created based on user research to represent user groups that might use a service, product, site, or brand similarly. Personas are seen as a critical tool in HCI for designing, developing, and evaluating interactive systems to ensure that they meet the needs of the target user group [5, 26, 32].

Throughout the history of persona development, the question of "interaction" has intrigued scholars [34]. Personas were introduced as flat, static "paper" profiles that contain information mainly in text format [17, 55], but sometimes coupled with graphs, tables, and statistics. One can view these flat file personas as essential a document. The essential CPP also contains a picture and the name of the persona to facilitate identification and empathy formation [50, 60, 69].

Over time, technology has increased the depth and level of interaction in personas. Most notably, web technologies such as HTML, CSS, and JavaScript have been deployed to develop persona systems that contain interactive features, including real-time persona generation [36, 39, 40]. However, such personas remained largely one-sided: there was no *genuine* interaction between the user and the persona, only one-sided interaction like filtering [34]. The introduction of generative AI and LLMs into the persona creation process ushers in a new era, in which personas can be interacted with using natural language. Users can engage in conversation and dialogue with the persona in a reciprocal activity. The need for true interaction between the persona and the persona user has become more evident with the increasing complexity and information content of personas – the emergence of HIUPs.

So, adopting the analogy of Goh et al. [31] in terms of "three waves of HCI", we may postulate that there are three waves in the historical development of personas: (1) *the classic era*, consisting of paper persona profiles (e.g., [10, 17, 29, 33, 51, 59]); (2) *the data-driven era*, consisting of online and algorithmic persona systems using social media and online analytics data (e.g., [3, 4, 52, 83, 85]); and (3) the *AI era*, in which personas are created and interacted with using AI technologies, most notably generative AI and LLMs [8, 20]. Concerning this novel AI era, the body of literature is only developing, with few empirical studies we are aware of [1, 20, 22, 35]. Thus, the question of persona modality, connected with the general *cadre* of technological development, remains relevant.

Previous research highlighted AI personas' utility in UCD by enabling designers and developers to understand user needs through archetypal representations [54, 61]. However, there is limited research comparing interaction modalities such as CPP versus DP personas, despite personas' growing use of generative AI applications [15] and the emergence of HIUPs. Existing studies in conversational AI have noted the importance of natural language interfaces for reducing cognitive load and increasing user engagement [49]. Research into cognitive load in interactive systems, and in high-information systems (HIUPs in particular), highlights that fragmented, conversational approaches, like those in DPs, may reduce mental effort compared to CPPs, which can overwhelm users with excessive information and detail [70, 78] These findings are particularly relevant as systems increasingly aim to adapt to diverse user needs. Recent studies suggest that hybrid approaches, combining document-based and dialogue-based interaction, could address the limitations of each persona, and HIUP, modality. For instance, allowing users to switch between comprehensive document views and conversational interfaces may improve usability in knowledge- and information-intensive tasks [57]. Overall, research on human-persona interaction is scarce. Despite its promising potential, the adoption of conversational technology does not *always* have a positive impact on the UX because the technology has been prone to errors and fails in interaction with the system users [27]. So, even though the application of LLMs in persona generation introduces novel opportunities for naturalistic interaction and adaptive feedback, expanding the scope of how personas are used in UCD [15, 49], the trade-offs between CPPs and DPs pose a compelling research gap.

Research within the conversational user interface (CUI) community also provides valuable insights into how persona design in commercial or design-driven domains might benefit from findings related to conversational agents. For instance, Clark et al. [16] and Doyle et al. [23] have explored how users perceive humanness and conversational qualities in intelligent personal assistants, raising questions about whether similar expectations exist when users interact with interactive personas. These studies highlight that "good" conversational experiences are shaped not just by conversational system functionality but also by user assumptions about agency, responsiveness, and human-likeness; concepts that may transfer meaningfully to interactive personas. Further, Zargham et al. [81] and Reicherts et al. [63] investigate the design of proactive voice assistants, showing tensions between user expectations and system behavior. Their findings suggest that the temporal dynamics and contextual appropriateness of persona behavior (e.g., proactive vs. reactive) could be relevant when designing interactive personas, as these factors possibly enhance persona believability and trustworthiness.

3 Methodology

3.1 Overview

To address our RQs, we conducted a user experiment with 54 participants. Participants were recruited via email at the computer science research institute where the study was conducted. The institute's researchers and engineers were generally knowledgeable about AI and some of them even had startup companies focused on AI technology, so the fictitious scenario was relevant to them (the scenario is described in the following subsection). Five study administrators took turns to conduct the study sessions in two separate study rooms that had an identical setup (the same set of

Table 1: The study flow and number of participants within each of the study flows.

Study Flow	Number of Participants (%)
CM / PF	13 (24%)
CF / PM	14 (26%)
PM / CF	13 (24%)
PF / CM	14 (26%)
Total	54 (100%)

laptops, monitors, and control devices). Each participant entered the study room, was welcomed, and seated in their chair. Each participant was offered identical equipment (Windows 11 laptop, two 24" displays, keyboard, and external mouse). Study supervisors used a joint script for the study to ensure that the same content was shared for each participant.

First, the participants were presented with the overall study procedure and asked for their consent to participate in the study. Study sessions were recorded and later transcribed. Second, the participants were presented with the work task scenario (WTS) with a set of questions to which they were to find answers in the HIUP. We applied a structured task for the participants to learn about different user groups (HIUPs). Third, as the participants were presented with two HIUPs, they initially examined one HIUP (either CPP or DP depending on which flow the participant was assigned) and then another HIUP that had a different modality than the first examined HIUP. According to the within-subjects design, each participant examined one CPP and one DP. After examining the first HIUP and answering the HIUP-related questions, the participant was guided to answer a short survey; then repeating the process for the other HIUP. After this, the participant was thanked for participation and rewarded with a gift card. The following subsections contain a more detailed description of the study procedures.

3.2 Experiment Design

As mentioned, the study was designed as a within-subjects experiment, ensuring that all participants interacted with both HIUPs (the general advantage of this study design is that it can help mitigate the effect of individual-level differences on study results). To mitigate the effects of learning and order of presenting the HIUPs (which are general challenges in user studies [41, 44], the sequence of interaction with the HIUPs was systematically varied among participants. This was done by creating four different sequences. Because the order of participants entering the survey was random by any practical means, the assignment to these flows in a balanced way was also random. In other words, the sequence of interactions was carefully counterbalanced, with participants randomly allocated to one of the following orders (C = chat, P = profile, M = male, F = female persona): from *Chat (DP) with Male HIUP to Profile (CPP) with Female HIUP* (CM → PF), from *Chat with Female HIUP to Profile with Male HIUP* (CF → PM), from *Profile with Male HIUP to Chat with Female HIUP* (PM → CF), and from *Profile with Female HIUP to Chat with Male HIUP* (PF → CM). Table 1 shows the number of participants allocated to each flow.

Participants engaged with two distinct forms of interactive HIUPs: one was a scrollable, searchable, and clickable conventional persona profile, the CPP. The other was a chat-based persona designed for conversation and dialogue, the DP. These HIUPs were generated using the Survey2Persona (S2P) system [65] which exemplifies data-driven persona generation [52, 83]. The operating principles of S2P are explained in the following subsection.

3.3 Persona Generation Using a Data-Driven Approach

The S2P methodology combines Cluster Analysis, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) to create realistic personas based on aggregated survey data (see validation study in [37]). The personas can be interacted with either through a "classic" profile view or through a dialogue-based view (see Figure 1). The persona generation leverages data aggregation and clustering, using ML to group survey responses into personas representing specific user segments, as done in various studies [11, 24, 79]. Each persona is enriched with attributes for more detailed information. Conversations leverage LLMs, and are refined by RAG to ensure that responses to persona users' queries are data-driven [46]. The operational workflow involves the system dynamically using survey data to generate relevant responses. The S2P system demonstrates a current approach to advancing the use of AI and LLMs in persona development [1, 20]. More information about the S2P can be found in other work [65]. The personas, HIUPs, in this study, were created using S2P based on a survey dataset from *Pew Research* containing approximately 10,000 U.S. participants[1]. The survey dataset was selected because it was compatible with the stakeholders' interests–it dealt with consumers' AI-related attitudes and the researchers at the institute were AI professionals.

For the study purpose, we generated two HIUPs in the S2P. As explained previously, the S2P utilizes survey data to generate HIUPs, and the survey data used for the HIUPs generation in this study was AI and human enhancement collected in the *American Trends Panel* by Pew Research in 2021 (n=10,260) [62]. HIUPs were generated with Machine Learning-Based Segmentation, a method that finds the most frequent respondent segments (personas) based on the survey answers. The two HIUPs generated from the dataset were Linda, a 65+ years old female from the US, and Mark, a 30-49-year-old male from the US. Both HIUPs were presented as (a) conventional persona profiles (CPPs), i.e., in a 2D format on the computer screen (Appendix 2A and 2B), and as (b) interactive dialogues with which the users discussed (DPs) (Appendix 2C and 2D). Each persona modality, CPP and DP, contained precisely the same information. There is no single persona profile form that is universally applied, as shown in systematic reviews of persona templates [57, 64]. However, there is a certain standard structure in terms of picture, name, and persona attributes laid out in a CPP. The persona template in the S2P follows this standard structure.

3.4 Data Collection

The study went through internal system testing and external pilot testing before launch. Data was primarily collected using a surveys

[1]The dataset from Pew Research: https://www.pewresearch.org/internet/dataset/american-trends-panel-wave-99/

Table 2: The list of task questions asked from the participants with correct answers for each HIUP.

Task Question	Correct Answer Linda	Correct Answer Mark
How does the persona feel about driverless vehicles?	Linda is skeptical (the participant could have elaborated their answer more)	Mark is moderately comfortable (the participant could have elaborated their answer more)
What does the persona think about the use of AI to decide which patients should get medical treatment?	Opposes	Opposes
What does the persona think about social media platforms using facial recognition technology to automatically identify people in photos?	Opposes	Mark Opposes
What does the persona think about accepting people opting out of facial recognition databases if the person does not have a criminal record?	More acceptable	More acceptable
Does the persona think that people who make a living by driving others or delivering things would lose their jobs because of driverless vehicles?	Probably would happen	Probably would happen
How does the persona feel about the usage of generative AI in social media writing?	No answer in the persona	No answer in the persona
What is the persona's opinion about the use of facial recognition technology at public protests?	Acceptable	Acceptable

(i.e., the Persona Perception Scale (PPS) and the System Usability Scale (SUS)), in this manner: *[participant is welcomed and explained the study] → [participant uses Persona 1] → [participant completes the survey with task and PPS + SUS] → [participant uses Persona 2] → [participant completes the survey with task and PPS + SUS] → [study is concluded].* The study introduction given to the participants is shown below, and the full study script is available in supplementary material[2] for review and further development.

The study introduction shown on the computer screen for the participants was as follows:

> *"Welcome to the study! In the following study session, you will examine two personas, perform two tasks with the help of each persona, and you will also answer two short surveys. Persona is a fictional depiction of a real user group. First, you will now proceed with the examination of the first persona and perform the task for the first persona. The task is presented on the next page. After the task, you will answer a survey after which you will be guided to examine another persona, answer questions about the persona, and then answer the same survey again. There are two kinds of personas in this study: conventional persona profiles and chat personas. If you are examining a conventional persona profile, you will find the answers by examining the persona profile. If you are examining a chat persona, you will have to ask the persona for the answers. When you are ready, select 'Ok, understood!', proceed to the next page, and start answering the questions regarding the persona you have on your other screen at the moment."*

While using the HIUPs, participants were engaged in a hypothetical scenario where they assumed the roles of either a software engineer or a manager at a company specializing in AI solutions. These two roles were created because not all participants were engineers, so "manager" was a so-called escape role that they could identify with. The participants could freely choose their role from these two options. They were introduced to two distinct HIUPs and tasked with responding to seven questions (Table 2) about these HIUPs, as identified by their (imaginary) superiors. The questions were randomly chosen from the survey dataset from which the personas in this study were generated, and the answer options were the ones that had been offered to the survey participants by Pew Research as the conductor of the survey. Only the question *"How does the persona feel about the usage of generative AI in social media writing?"* was not from the Pew Research survey. This question was intentionally introduced to the set of task questions as an *"impossible question to answer"* to test the hallucinative aspects of DPs and how users adapt when they cannot find an answer from the CPP [43]. In other words, this is an information retrieval task [2] which is a central task in persona use since retrieving information and learning from groups of people is one of the central functions of personas [17, 61]. Information retrieval is just one type of task for which personas can be used. We leave other use cases for future research.

Upon completion of the tasks involving both HIUPs, participants were thanked for their contribution. As a token of appreciation for their participation, they received a gift card. Following each session of using a HIUP, participants answered questions related to the HIUP, with these questions drawing from two established scales: the SUS [47] and the PPS [71], with items provided in Appendix 1. Both scales have been deployed in multiple HCI and persona studies (e.g., [12, 48, 68, 69]). Behavioral variables and task success rate were also measured (Table 3). The accuracy of their responses was

[2]https://bit.ly/study-procedure

Table 3: Additional study variables with measurements and definitions of each.

Variable	Measurement	Definition
Persona system dwell time	Time in seconds	How long did the participant use the persona system per session (in seconds)? Measured automatically by analytics system.
Number of interactions in the chat system	Number of messages	How many messages (number of messages) the participant sent to the chat persona? Measured in the chat persona system (S2P).
Number of mouse clicks	Number of mouse clicks	How many times did the participant click the system during the session (number of mouse clicks)? Measured automatically by the analytic system.
Task completion time (TCT)	Time in seconds	How long did it take for the participant to complete the task (in seconds)?
Task success rate (TSR)	Task success rate	How did the user succeed in the task of finding the correct answers to questions asked about the persona (the rate (%) of correct answers given by the user to the task)? The task success rate was calculated for the seven task questions. The principal researcher checked each task answer for each participant and calculated the task success rate (number of correct answers divided by the number of questions (7).

evaluated to calculate a success rate for each participant, comparing their answers to the actual details of the HIUPs.

In addition to survey data collection, all audio from the study sessions was recorded and transcribed using the think-aloud method in which the participant is encouraged to think aloud while using the HIUP, carrying out the task, and completing the survey. The concurrent think-aloud method we apply has been implemented in numerous previous persona user studies, e.g. [41, 42, 44, 45, 58, 72, 73, 76]. We used the variation called survey think-aloud which not only asks the participant to verbalize their thinking during the system use but also during survey completion [58]. Transcripts were created using MS Word's automatic translation tool and manually read through to ensure adequate quality for analysis. If mistakes were found, the lead author returned to the original voice recordings and fixed the issues (in general, the quality of automatic transcriptions was very high). These transcripts amounted to approximately 1700 pages of text (in MS Word's standard typesetting). The average study completion time was 66.5 minutes per participant (SD = 18.3), which denotes a considerable effort and engagement from the participants' side.

3.5 Participants

A user study involving face-to-face interactions in a lab setting and the user study was carried out in two research sites: a research institute and a university. The study involved a total of 54 participants ,which according to prior studies (the number of participants ranging from 29 to 54), is a sufficient number of participants [41, 42, 44, 45, 66, 68, 69, 71]. An *a priori* power analysis indicates that a within-subjects design with two groups (as is the case) requires a sample size of 54 to detect medium-sized effects (Cohen's f = 0.25), thus making our sample size sufficiently large enough to detect an effect of at least this magnitude. Informed consent was obtained from each participant. No personally identifiable information was collected, apart from the voice recordings that were made accessible only to the authors. The participant pool predominantly consisted of researchers (36 individuals, accounting for 66.67%) and engineers (9 individuals, making up 16.66%). The remainder of the participants included a diverse mix of professions. The average age was 33 years (SD = 10.60). Slightly more than half were men (n = 30, 55.6%), and the rest were female (n = 24, 44.4%). Of the participants, 20 (37.0%) participants had heard of personas before the study, and 12 (22.2%) participants had used personas before, while for the rest, personas were not familiar (however, all participants were given a clear operational definition of the persona concept before starting the study). Average experience with personas was 1.92 years (N=54, SD=5.85 years).

3.6 Data Analysis

Given that the study design incorporated repeated measurements, we used a multi-level linear regression analysis. In this approach, each stimulus was considered a separate case but nested within participants, since each participant interacted with two distinct stimuli, allowing us to adjust for variability within subjects. This analysis encompassed data from both the CPP and DP. In our results, we present the coefficients (B = unstandardized coefficient) of the regression model for those findings that reached statistical significance, indicated by a p-value of less than .05. The dataset was cleaned by removing the internal and external test subjects (see Table 4).

In addition to statistical analysis, the transcribed session recordings were qualitatively analyzed for each statistically significant SUS item and PPS construct to contextualize the findings in a process known as triangulation [13]. The idea in triangulation is to combine different data sources, quantitative survey data and qualitative session transcript data in our study, for more comprehensive knowledge on a subject. Qualitative analysis was carried out manually by finding relevant quotes from the session recording transcripts. An AI (ChatGPT-4, v. March 2024) was used to aid in the identification of relevant quotes, which is a known and used method for identifying quotes in qualitative data analysis [80, 82], but all quotes were manually verified by the lead author to avoid hallucinations as recommended by Turobov [80].

Table 4: Dataset cleaning (i.e., excluded participants) with reason.

Excluded (N)	Reason
2	Two pilot participants were excluded (four rows)
6	Six test rows were excluded (testing system and survey)
54	Remaining participants

Table 5: Comparison of measurements across groups without control variables. Significant results are bolded.

Measurement	DP EMM	CPP EMM	B	SE	p
Persona perceptions					
Clarity	4.708	4.537	-0.171	0.136	0.209
Compassion	3.296	3.284	-0.012	0.188	0.948
Completeness	**4.932**	**5.315**	**0.383**	**0.188**	**0.042**
Consistency	5.324	5.028	-0.296	0.192	0.123
Credibility	4.898	5.343	0.444	0.260	0.087
Empathy	3.815	4.154	0.340	0.236	0.150
Similarity	3.858	3.901	0.043	0.269	0.872
Stereotypicality	4.315	3.815	-0.500	0.277	0.071
Transparency	**4.864**	**5.259**	**0.395**	**0.184**	**0.032**
Usability	**5.926**	**4.519**	**-1.407**	**0.238**	**<0.001**
Willingness to Use	5.333	5.037	-0.296	0.205	0.147
Usability					
SUS	**5.906**	**4.863**	**-1.043**	**0.188**	**<0.001**
.					

Notes: A comparison of Estimated Marginal Means for the group variable in the context of a multi-level model is shown. EMM: Estimated Marginal Means. B: Unstandardized regression coefficient. SE: Standard Error. "Profile" is used as the reference category in the model estimation

4 Results

4.1 RQ1: How Does the HIUP Interaction Modality (Document or Dialogue) Affect Users' Experience, Behavior, Usability, Learning and Perceptions of the HIUP?

4.1.1 Quantitative Results. The first model represents a comparison of the CPP and DP groups for (a) HIUP perceptions and (b) usability (SUS), and (c) system dwell time. The CPP group exhibited higher transparency scores (B = 0.395, p = .032) and was perceived as more complete (B = 0.383, p = .042). The DP, on the other hand, was perceived as more usable (B = -1.407, p < .001), and had a higher SUS score (B = -1.043, p < .001). These results are summarized in Table 5[3].

Unlike PPS, which is based on psychometric constructs, the SUS items represent conceptually distinct dimensions and thus can be explored individually. Because the SUS scale offers multiple perspectives on usability [48], we also investigated the values of the individual items on the SUS scale. As can be observed from Table 6, all items exhibit significant differences (in favor of the DP's usability), with the sole exception of "I found there was too much inconsistency in the persona system" (B = 0.150, p = .480), which attained statistically similar scores in both groups. The following subsection presents qualitative insights to contextualize these findings.

4.1.2 Qualitative Results. The session recording transcripts for all the nine statistically significant SUS items (SUS1-SUS5 and SUS7-SUS10) and all three statistically significant PPS constructs (Completeness, Transparency, and Usability) offered a deeper understanding of the users' preferred HIUP: the DP or the CPP. Each quote is marked with the participant ID and a reference to the HIUP on which the quote was obtained (C=chat persona, P= persona profile, m=male [Mark], f=female [Linda]). In the following results, quotes related to the DP are highlighted in light blue color to provide the reader with a visual way to rapidly distinguish between the quotes about each HIUP type. Quotes for the nine statistically significant SUS items and three PPS items are presented in Table 7.

4.1.2.1 SUS 1 I think that I would like to use this persona system for similar tasks. From the user quotes relating to SUS 1 (see Table 7), we identified evidence of why the users prefer the DP over the CPP for similar tasks. The quotes support the statistical results that the users found the DP useful for similar tasks due to the DP's ability to compress vast amounts of information (survey data) into a digestible package. Also, the user experience was seen as similar

[3]For full disclosure, we also ran a model with the following control variables: participant gender, persona gender, age, previous persona experience, and previous chatbot experience. This model computed the estimated marginal means – that is, the mean of the measurement after taking into account the controls. The results are consistent with what was observed without controls. The only difference was the effect of Completeness, which ceased to be significant in the presence of control variables (B = 0.296, p = .143).

Table 6: Comparison of the SUS items across groups with control variables. Significant results are bolded.

Measurement	DP EMM	CPP EMM	B	SE	p
I think that I would like to use this persona system for similar tasks.	**5.583**	**4.167**	**-1.450**	**0.313**	**<0.001**
I found the persona system unnecessarily complex.	**1.958**	**3.354**	**1.413**	**0.298**	**<0.001**
I found the persona system easy to use.	**6.271**	**4.479**	**-1.821**	**0.291**	**<0.001**
I think that I would need the support of a professional to be able to use the persona system.	**1.984**	**2.911**	**0.967**	**0.250**	**<0.001**
I found the various functions in the persona system were well integrated (are working well together to support the use of the system).	**5.500**	**4.583**	**-0.955**	**0.276**	**0.001**
I found there was too much inconsistency in the persona system.	2.187	2.333	0.150	0.212	0.480
I would imagine that most people would learn to use the persona system very quickly.	**5.973**	**4.869**	**-1.134**	**0.254**	**<0.001**
I found the persona system very difficult to use.	**1.736**	**3.154**	**1.450**	**0.271**	**<0.001**
I felt very confident using the persona system.	**5.792**	**4.750**	**-1.053**	**0.262**	**<0.001**
I needed to learn a lot of things before I could get going with the persona system.	**1.978**	**2.604**	**0.639**	**0.209**	**0.002**

Notes: A comparison of Estimated Marginal Means for the group variable in the context of a multi-level model is shown. EMM: Estimated Marginal Means. B: Unstandardized regression coefficient. SE: Standard Error. "Profile" is used as the reference category in the model estimation

to a familiar chatbot (ChatGPT). The CPP was seen not as favorable to perform similar tasks due to the vast amount of data that had to be browsed through while using the CPP to complete the task.

4.1.2.2 SUS 2 I found the persona system unnecessarily complex. From the user quotes relating to SUS 2 (see Table 7), we identified evidence of why the users find the DP less complex than the CPP. The quotes support the statistical results that the CPP was found more complex than the DP. DP was seen to save time when information was to be found quickly. CPP contained a lot of information presented as text and tables that were perceived as difficult to peruse and find the answers to complete the task, making the CPP more complex than DP. In other words, the complexity of the CPP was associated with the amount of information presented.

4.1.2.3 SUS 3 I found the persona system easy to use. From the user quotes relating to SUS 3 (see Table 7), we identified evidence of why the users find the DP easier to use than the CPP. The quotes support the statistical results that the DP was found easier to use than the CPP. The DPs functionality of *"just asking questions"* (P24) made the DP easy to use while the CPP's overwhelming information was seen as difficult to handle as well as navigating through the CPP (scrolling) and the need for looking for things. The CPP took more time to comprehend and delve into the vast amount of information than the DP.

4.1.2.4 SUS 4 I think that I would need the support of a professional to be able to use the persona system. From the user quotes relating to SUS 4 (see Table 7), we identified evidence of why the users find that the use of the DP needed less guidance from a professional than using the CPP. The quotes indicate that help was needed to use both persona systems. Interestingly, getting the hang of the DP was seen as difficult and the straightforwardness of the DP system was not uniformly perceived by the users. So, although chat has a general reputation of being an intuitive interface that requires little guidance, the results indicate that users can benefit from guidance especially when it comes to prompting the DP. On the other hand, the need for help to get going with the CPP was also needed by some users (P21), although others found it readily usable (P04).

4.1.2.5 SUS 5 I found the various functions in the persona system were well integrated. From the user quotes relating to SUS 5 (see Table 7), we identified evidence of why the users found that the DP had features better integrated than the CPP. The straightforwardness, and low number of functions, of the DP were noted by a participant (P07) and the question-response workflow of the DP was found easy (P35). For the CPP, it was also found that the functions of the persona system were well integrated, and the CPP was presented clearly but finding the information by scrolling through the CPP was seen as difficult. CPP scrolling made using the CPP system difficult (P48).

4.1.2.6 SUS 7 I would imagine that most people would learn to use the persona system very quickly. From the user quotes relating to SUS 7 (see Table 7), we identified evidence of why the users would think that most people would learn to use the DP system very quickly more probably than the CPP system. The DP was seen to have a low number of features, "just" the chatting, and chatting was also seen natural way of finding information (answering the task) which makes learning the system fast. DP interface was perceived as simple making learning the system fast and the chatting itself was perceived as natural. The CPP system was seen as crafty, hard to use, and having too much information to digest. The CPP system was not seen as impossible to use but more time-consuming to familiarize oneself with the system.

4.1.2.7 SUS 8 I found the persona system very difficult to use. From the user quotes relating to SUS 8 (see Table 7), we identified evidence

Table 7: User quotes from the session recording transcripts for statistically significant SUS items and PPS items.

Item	DP User Quotes	CPP User Quotes
SUS1	"I think that I would like to use it for similar tasks. Yeah, sure. Of course, I'm using ChatGPT." (P25, Cm) "Yes, if it's going to represent my survey in a different way, then yes [I would use it]." (P20, Cf)	"Somewhat agree [with using profile persona], I kind of don't like navigating through this much data." (P02, Pm) "I would not have used this persona if given the choice." (P22, Pm)
SUS2	"It's not about the mental effort. It's about the time consumed." (P26, Cf) "It [chat] was easier to use than the last one [profile]. The last one [profile] was very difficult to figure out. The answers were complicated" (P42, Cf	"The [profile] system, it's all tables and stuff. I mean, I could use the system, but it's difficult to use. It just was difficult to use." (P25, Pf) "I was supposed to find the information quickly and it wasn't that easy, but the search option helped a bit, but I'd rather use the first persona [chat]." (P02, Pm) "It is complex, but it's not unnecessarily complex like it makes sense to be complex. Having all this much information." (P53, Pf)
SUS3	"I found the persona system easy to use. Strongly agree. Yes, it was very easy. As I said, you just ask questions to a chat bot and get answers." (P24, Cf) "I think this [chat] is super easy to use. I mean it's just typing with the persona. Very natural." (P04, Cf) "There weren't many functions there, just scrolling, typing and submitting., Scrolling and answering. Yeah, it was very simple, very functional." (P25, Cm) "It [chat] is not very difficult. It is easy to use but it is difficult to get what you want." (P47, Cm)	"I feel like it was harder to use [than the chat persona]. It was overwhelming at first. I started kind of searching with certain keywords. It got easier to understand, but I do think it takes a learning curve." (P21, Pf) "It's kind of not easy to navigate through all this information. You need to keep looking for things." (P02, Pm)
SUS4	"I actually did require your [the admin's] help." (P20, Cf) "No, it was very straightforward." (P02, Cf) "Disagree because I might need some. I don't know. The system was simple, but you [the admin] had to tell me about the prompting." (P25, Cm)	"I would think that I would benefit from having a professional with me at the beginning to kind of teach me how to go through that." (P21, Pf) "No, I don't think so. No, no need [for professional help]." (P04, Pm)
SUS5	"They were sort of working well together. The functions and again there aren't many functions because the UI is kind of straightforward." (P07, Cm) "The various functions and the persona system were well integrated. It was, I strongly agree. Also, since when I send the message, I get feedback that he's responding, so very easy to use." (P35, Cm)	"I would agree. I feel like everything was well integrated and easy to look through and presented very well." (P21, Pf) "Functions were working well together support the uses." (P22, Pm) "I like the colors. I like the representation, but the scrolling, etc. finding the information was difficult." (P48, Pm
SUS7	"There was not much to learn" (P02, Cf) "The interface was not complex." (P05, Cm) "Mostly yes, because chatting is natural." (P22, Cf)	"Yeah, I imagine. It's just very detailed to the point that you get confused." (P22, Pm) "Maybe not most people. I felt very crafty using the system." (P24, Pm) "They [some systems] are always difficult for normal people. "You just need to have an hour. Then you will be familiar with it." (P52, Pm)
SUS8	"Not at all [difficult to use]." (P24, Cf) "It's not very difficult to use but very difficult to get knowledge from. I feel like you are putting questions in but my interpretation [of the chat answers] can vary." (P47, Cm) "Not at all, no struggles there." (P53, Cm)	"It was quite tense." (P12, Pm) "Yes, somewhat [difficult to use]." (P50, Pm) "You don't remember many difficulties." P22, Pm)
SUS9	"Yeah, I think it was straightforward." (P34, Cf) "I would say somewhat agree. I think it's easy." (P20, Cf)	"You know, I need to ask a lot of questions from you [moderator]." (P18, Pm) "I wasn't very confident using it at first. I think after kind of getting a hang of it and kind of understanding the tables and things and familiarizing myself with the information available, I thought it got easier. So, I would say I felt very confident." (P21, Pf)

		"It was not easy. The information was there for the persona but not confident to use." (P48, Pm)
SUS10	"I didn't have to learn anything." (P02, Cf) "You just have to know how to submit a piece of text." (P24, Cf) "Nothing complex about it. You just ask questions." (P27, Cm)	"I think it would take a learning curve to understand how to use the system effectively." (P21, Pf) "It was difficult to scan the information." (P28, Pm) "It was just tables with percentages and the tables were clear and I would imagine that most people would learn to use persona." (P33, Pf) "Not for me but non-tech people would definitely have a very, very hard time." (P34, Pm)
Completeness	"It gave me what I like. Whatever I asked, it gave me the response." (P02, Cf) "I just got the information that I asked for, but nothing extra." (P04, Cf) "I don't think it was detailed enough just because, I, for example, agree with some of the use cases personally. But like if you ask me whether it should be used in other use cases, not sure." (P12, Cf) "No, it didn't lack [any information]. It had more than enough. It didn't lack anything. Actually, I just had to ask it." (P20, Cf)	"Yeah, too much detail. I feel it was like a very long description. It was detailed. That was information that needs really time to analyze." (P22, Pm) "I was able to find everything except for the social AI generating content and social media, but for the rest, it was fine." (P36, Pm) "It didn't lack critical information. I don't know about the technology-related background, but it didn't affect the survey that I did. So, I had the percentages and therefore the needed information." (P35, Pf)
Transparency	"There was a survey, but I didn't understand how the persona was created. I could understand how the information of the persona was obtained." (P19, Cm) "I kind of understood that it was an amalgamation of a 10,000 people survey and then we some system created personas from that." (P21, Cm) "I couldn't understand how the information was obtained. Not really. It just says the information, but it doesn't say how it [the persona system] gets it." (P27, Cm) "I know the background [of how the personas were created], it's through surveys and responses. It was mentioned actually at the beginning of the study. It's more than 10,000 survey responses. So, I was provided with the information on how the Persona was created." (P20, Cf)	"Yeah, based on the survey that and I understand this". (P22, Pm) "Yeah, yeah. Things from the survey. So, but then how it [data] was processed to come up with the persona, no." (P48, Pm) "There was no information. I have no idea how it was created." (P04, Pm) "It's created based on the survey data." (P22, Pm) "I think I understood somehow the most important questions since I was going through the questions. It seems like you just somehow got their answers to these questions. Built the profile out of it." (P27, Pf)
Usability	"It didn't require any mental effort from me." (P02, Cf) "There was no extra distraction. There is no extra information, just the chat bot. This is like invisible persona, it's nice." (P04, Cf) "I didn't need to think that much. It's more like look up information to answer the questions. All I had to do was talk to it. Very, very intuitive." (P20, Cf) "It is similar to chatting on WhatsApp of ChatGPT." (P35, Cm) "I just have to complete the question and then it answers." (P24, Cf) "No mental effort, just ask the question and it gives you the answer. It is pretty easy to use, very clear and understandable." (P27, Cm)	"It's kind of not easy to navigate through all this information. You need to keep looking for things." (P02, Pm) "I think the more information you put in the survey, of course, it's going to be great for research. But for the average person looking at it. . . Maybe if information could be condensed just to make it easier to make decisions, then yes, I'd say it is easy to use." (P20, Pm) "In the previous one [chat], it gave out all the information, she gave me an answer of the group that she represents. Here, I have to do the work to search for the information." (P18, Pm) "I felt like it required more effort to get to an understanding with the persona just because of the amount of information that I was presented." (P21, Pf) "It [profile] needs more time and more mental effort than the chat." (P06, Pm)

of why users found the DP less difficult to use than the CPP. The DP was found easy to use but one user (P47) mentioned that it is difficult to gain knowledge from the DP. The answers DP gives leave too much room for interpretation. There is merely input and output in the DP making the system easy to use (P47). The CPP was found tense and difficult to use (P12) but for some users, the CPP was easy to use.

4.1.2.8 SUS 9 I felt very confident using the persona system. From the user quotes relating to SUS 9 (see Table 7), we identified evidence of why users felt more confident using the DP than while using the CPP. Overall, the DP was found straightforward and easy to use making the DP system's usage confident. The CPP was found hard to use due to the amount of data and its presentation structure (tables) in the system. The information was there to be found in the CPP to perform the task but was not easily available. The longer the CPP was used, the more confident the user was while using the CPP system (P21) which indicates a steeper learning curve for the CPP than for the DP. To assess the possible correlation between *system use time* and *confidence* using the persona system, we calculated Pearson's correlation coefficient between system use time and SUS9 (measuring confidence). The results indicate a weak positive, non-significant, correlation between system use time and SUS9 (r=0.104, p=.284).

4.1.2.9 SUS 10 I needed to learn a lot of things before I could get going with the persona system. From the user quotes relating to SUS 10 (see Table 7), we identified evidence of why users felt that they had to learn more things before getting going with the CPP than with the DP. The DP was found easy to use since the DP only requires the user to write text (questions) and the DP will give the user the answer. The CPP was found easy to use by some users because of the clear tables but there were notions about the CPP requiring the user to be tech savvy to use the CPP easily. One user (P53) liked it when the CPP had all the information there to be seen (P53) compared to the DP that only revealed information that was asked from it.

4.1.2.10 Completeness. From the user quotes relating to Completeness (see Table 7), we found evidence of why users found the CPP more complete. Survey items for Completeness are shown in Appendix 1. The DP was not seen as complete enough because it was difficult to get an overall look at the persona while the CPP offered all the information at once. It was easier for some users to get the overall perception of the CPP by scrolling. With the DP, the user would have had to ask for more information with both personas the task required only answering certain questions that left little room for additional questions to be asked from the DP while with the CPP the users had to scroll up and down the CPP to find the answers to the task and simultaneously, they could, almost by accident, form a general understanding of the persona. Then again, for some the length and volume of information in the CPP was too much, causing issues of information overload (P48).

4.1.2.11 Transparency. From the user quotes relating to Transparency (see Table 7), we found evidence of why users found the CPP more transparent. Survey items for Transparency are in Appendix 1. The user quotes on Transparency echo the statistical results. All the participants were provided with basic information on the data for the persona generation. For the CPP, the user quotes show that being forced to go through the visualization of the CPP (P27) helped to get the feeling of a deeper understanding of the CPP creation process than the DP creation process. Many users (P19, P02, P22) referred to data but still some users felt that no information on the persona generation data was given (P04). It is vital to note that we in fact did *not* give a detailed account of how the personas were created. So, the fact that some participants had no issues with transparency could imply the participants presumed to know how the personas were generated due to their general knowledge of computer science and AI. It was also the case that a select few participants did ask technical questions about the system details, which we answered. However, this was outlier behavior and the vast majority of the participants did not possess information on how the S2P system works. So, it is interesting to find that many of them nonetheless considered transparency to be reasonably high.

4.1.2.12 Usability. From the user quotes relating to Usability (see Table 7), we found evidence of why users found the DP more usable. Survey items for Usability are in Appendix 1. The user quotes on Usability reflect the statistical results. The intuitiveness, familiarity, and low need for mental effort in different forms were mentioned for the DP (P02, P04, P20, P35, P24, P27) concerning increased usability while the CPP was perceived as less usable with a lot of information in the CPP (P18, P20, P21). The sheer amount of data in the CPP was hard to digest and a user also felt pressure while using the CPP (P33).

4.2 RQ2: Under What Conditions Is *Document Modality* More Preferrable than *Dialogue Modality*, and Vice Versa?

The participants were asked, after using both HIUP modalities, which one they preferred. A third of the participants preferred the CPP (n = 18, 33.3%), while two-thirds preferred the DP (n = 36, 66.7%). Women were 25% more likely to prefer the CPP than men, although the gender difference is not statistically significant, X^2 (1, N = 54) = 0.34, p = .561. Our impressions of the conditions under which each persona modality is preferred can be seen in Table 8.

The participants who preferred the CPP shared how they liked the CPP's details as they were able to read and analyze information clearly. They considered that the CPP provided them with more information than they needed. Some participants also disliked the DP's tendency to show distributional variability in the HIUP group (in the style of: "There are X% said answer A is correct. However, Y% agreed with answer B."). This kind of response confused them, while CPP enabled them to independently draw conclusions from the available data about the HIUP. In contrast, the participants who preferred the DP often did so because they could save time: they mentioned how fast the DP would come up with answers related to the questions. The participants also appreciated the convenience afforded by the DP, as well as the sense of certainty that came with the DP's clear and confident responses – in contrast, some of the participants struggled to answer the questions using the CPP since using the CPP required them to engage in an analytical process, essentially doing their own analysis rather than digesting the ready answers given by the DP.

Each participant was also asked for their HIUP preference at the end of each session. Instead of pointing out a clear favorite, some participants indicated that they would prefer the combination

Table 8: Conditions for preferring the CPP or DP modality.

When did people prefer using CPP? (For which purposes? Under which conditions?)	When did people prefer using DPs? (For which purposes? Under which conditions?)
When the participant was visually oriented. For example, the participant found the profile visually more appealing than the chat persona. When the participant wanted to have a quick broader look at the persona. It was possible easily with the profile persona. When they wanted to obtain more detailed information. When there were no specific questions to answer.	When the participant wanted to find answers to specific questions quickly. When the participant wanted to avoid going through a lot of information in the persona. When dialogue with the persona was wanted or needed. When there were specific questions to answer.

of both modalities. Excerpts from each preference are shown in Table 9. Overall, 33 participants preferred the DP (61.1%), while 15 participants preferred the CPP (27.8%), and 6 participants (11.1%) indicated a preference for using both. In many cases, the participants raised good points about the pros and cons of each modality, so we can say the "preference" often indicates "slight" preference, although there were clear-cut cases as well (e.g., P41 in Table 9).

From the user quotes on the HIUP modality preference, it can be seen that the DP was preferred for its need for low mental effort and easiness of searching for information; interestingly, even scrolling now appears to be a high-effort activity (P16)! On the other hand, the CPP was preferred for its ability to show all the HIUP information at one glance and for the CPP's structure where the user would unconsciously process information about the HIUP and form a thorough view of the HIUP while with DP such a thorough view formation was not seen possible. Overall, the DP was seen by many as easier to use compared to the CPP.

4.3 RQ3: How Does the HIUP Modality Affect Task Completion?

The second model represents a comparison of the CPP and DP groups for (a) task perceptions and (b) mouse clicks (no scrolling), and (c) Task Success Rate. The CPP was perceived as more difficult ($B = 1.712$, $p < .001$), and the success rate for the information retrieval task completion was higher ($B = 0.124$, $p = .002$). The DP was perceived as more enjoyable ($B = -1.519$, $p < .001$), and resulted in higher engagement as measured by the number of mouse clicks ($B = -6.866$, $p = .040$). The results are summarized in Table 10.

5 Discussion

5.1 Answers to Research Questions

In terms of RQ1 quantitative results, users find CPPs more complete and transparent than DPs while finding DPs more usable and easier to use than CPPs. CPPs were seen as more complete for their ability to show more information at one glance and CPPs were seen more transparent since the "back-end" of the HIUP was more visible in CPPs than in DPs. Similar results with conversational agents have been found, for example, by Schnabel et al. [75] and Schmitt et al. [74]. These were supported by the qualitative findings that mentioned user insights on CPP's transparency. Our qualitative results support the quantitative results through user comments on the lack of transparency in DPs. DP was just the chat interface making it hard for the user to explore the HIUP at one glance. That is, the CPPs had tables representing the survey items while the DPs had nothing. For usability, the DP was a clear winner over the CPP in our task of finding specific information about the HIUP. This result was supported by our qualitative results in which users mention the DP's ability to offer answers to questions with no need to browse through a CPP.

In terms of RQ2, users' visual orientation makes the user lean towards the visual HIUP, i.e., the CPP. For some users, a quick overview of the HIUP was easier to reach with the CPP. The usability of the DPs was lower when there were no specific questions to answer from the DP. With the CPP, it was easier for users to just browse through and "get to know" the HIUP. With DPs, this kind of browsing through would mean asking random questions from the DP. Results for RQ2 connect our study well to previous research on conversational agents and static persona presentations [42]. Surprisingly, it was difficult for some users to filter down the HIUP information from the DP's answers when the DP was giving specific distributions for the HIUP, i.e., percentages. With the CPPs, some users found it easier to gather information.

In terms of RQ3, on average, the DPs required more mouse clicks to use. This could indicate either positive or negative engagement. Positive in terms of more actively processing HIUP information, and negative in terms of being confused and having to click a lot to find the information. For task success, the CPPs resulted in better task performance than DPs. Perhaps surprisingly, not all users prefer the DP for the task. Some users prefer finding information from CPPs rather than using a chat to find the answer to the task. Also surprisingly, clicking was more frequent with the DPs than with the CPPs. Due to the ways the DPs are used, this could be natural. The chat's text bar has to be activated each time the user wants to write to the chatbot, and the message has to be sent by clicking or by pressing enter. Also, the CPPs used by browsing, not clicking. The DP is used by clicking, not browsing so the logic for use is different for the CPP and DP. On the other hand, the measurement did not capture the scrolling behavior in the CPP.

Another surprising finding is that despite the higher usability for the DP, the task success rate was lower for the DP than for the CPP. On average, users spent more time on the CPP than on the DP, the CPP was perceived as more difficult to use than the DP, but according to the qualitative data analysis, the DP was occasionally giving hard-to-interpret answers, and the DP was also hallucinating answers [43] which could have influenced users' task success rate.

Table 9: Excerpts from participants' answers when asked which of the HIUP modalities they preferred.

Preference for DP	Preference for CPP	Preference for both
"The first one [chat persona]. because I didn't need to put in any mental effort. I just asked and it gave me the answer. For the second one, I felt pressure to find the information quickly and it wasn't that easy, but the search option helped a bit. But I'd rather use the first persona [chat persona]." (P02) *"Chat was much easier to use and then the problem with the other one [persona profile] is that the data was kind of scattered all over the place. Whenever I need to look up something I need to scroll up and down and then even within them, there are subsections that I have to read."* (P16) *"The first one, definitely [chat persona]. Because it was easier to use. It seems very friendly and the last one [persona profile] is very difficult to find out the answers. It is complicated."* (P41) *"I like the first [chat persona] as it's easier to use and in providing information. However, what I liked about the second one [persona profile] is that it gives lots of information which actually helps me more to decide if I want different uses or different purposes, unlike the first one. The first one you just have a specific thing based on your question. However, you may miss information just because you did not ask the right question."* (P54)	*"I feel like there were some ups and downs. For example, for the profile, I felt more confident about my answers once I found them, but it was more tedious to find them than from the chatbot. I was always like a bit less confident with it [chat persona]."* (P17) *"I think it was the full picture, the other one [persona profile] had more information. I complained about having more information than this when you ask questions, like, what does this persona represent. Then I need more information because with my interaction with the chatbot, I just asked those seven questions and I only know the answers to those seven questions, although the answers become more comprehensive than needed to answer the question [task] that was helpful. The second one, chatbot, if you can chat more, ask more questions, and maybe have some free interactions around topics you can still pick the topic about facial recognition technology or driverless vehicles. But then I ask my own kind of questions. Then maybe I could have the chance to learn more about the general persona because of the survey set up. Then you have those questions you want to finish. You just ask those questions."* (P19] *"The profile was like the data. It gave me a better understanding of what the person I was dealing with is like. There was other information that I realized I didn't use. But in the back of my head, I think I did process the information. So, when I was answering the questions, I knew more about the persona. Whereas with the chatbot it was giving me information specific to that one question I asked and then I realized I could have probably asked any other question and if I hadn't asked more questions, I wouldn't get anything else. It was just telling me what I needed to know about a specific topic and everything else was irrelevant and there was some context. I think that I could have better shaped those opinions. Whereas when I was looking at the different data [persona profile], you know I was processing it all but really focusing on one thing. And so, it [chat persona] informed how I was viewing that one data point It just seemed it was very muddled. Yeah, so definitely the data [persona profile]. And if I were to go again, I would prefer the data [persona profile] over the chatbot [chat persona]."* (P39) *"I like detailed information about the issues. The first one [chat persona] was straightforward and they give you the percentages and without like very detailed information."* (P47) *"The traditional one, even though, like I feel comfortable because I can see all the information in one stretch. The chatbot interface, it felt difficult. Even for accessibility, while looking I want to scroll up to see the whole thing, so I feel difficult with the second interface [chat persona], but with the traditional [persona profile] one I felt more comfortable because I can use all the things which helped me to find out the right answers."* (P43)	*"I would like to have both of them, because for example, when I'm asking a question from the first system [chat persona] I was getting only a very specific answer, whereas in the second case, I was given a lot of information in which I might be able to correlate some of the factors. The second one [persona profile] was good for decision-making, where I was getting more information to make a better decision or maybe support my decision-making process. But for the first one [chat persona], the problem was that I was only getting a very tiny mass of information. The first one was better for what I was doing. It was making my life easy, but at the same time, by making my life easy, it was not giving me a lot of information."* (P05)

Also, stereotypicality was higher for the DP (not statistically significant). The higher sterotypicality for the DP could indicate that the DP's answers to user questions and overall appearance were seen more tied to a simplified idea of the HIUP whereas with the CPP the HIUP was seen as less stereotypical indicating universality. Credibility was higher for the CPP (not statistically significant) than for the DP. Low credibility can indicate a user's low trust towards the HIUP and also low credibility can indicate that the user's own experience with real people like the HIUP conflicts with the HIUP perception [14]. WTU was higher for DP than for the CPP, not a statistically significant result, but indicates user preference towards the DP for the task the users were given. From the qualitative analysis of transcribed session recordings, it was found that the DP usability for finding information from the HIUP was seen as easier than finding information from the CPP.

Table 10: Comparison of task measurements across groups without control variables. Significant results are bolded.

Measurement	DP EMM	CPP EMM	B	SE	p
Task perceptions					
Enjoyment	**5.463**	**3.944**	**-1.519**	**0.292**	**<0.001**
Difficulty	**1.799**	**3.511**	**1.712**	**0.306**	**<0.001**
Confidence	5.721	5.544	-0.177	0.131	0.178
User behavior					
Mouse Clicks	**26.800**	**19.933**	**-6.866**	**3.341**	**0.040**
Task Success Rate	**0.711**	**0.835**	**0.124**	**0.041**	**0.002**
Dwell Time	721.527	828.604	107.078	68.257	0.117

Notes: A comparison of Estimated Marginal Means for the group variable in the context of a multi-level model is shown. EMM: Estimated Marginal Means. B: Unstandardized regression coefficient. SE: Standard Error. "Profile" is used as the reference category in the model estimation

5.2 Theoretical Implications

It is known from previous research that people have individual preferences when it comes to digesting HIUP information as numbers or text [70]. The current study further corroborates this state of affairs: *some users actually find tables in HIUPs as a useful form of communicating data*; these same users might find chat restrictive to their HIUP use because the DP does not "show itself"; it only shows fragments of the HIUP information based on specific prompts. On the other hand, for many users, this is enough, as they prefer the simple, naturalistic interaction brought about by the chat interface. Similar results have been found in previous studies on interactive personas [38–40]. No clear indication of anthropomorphism, i.e., humanlikeness, of the HIUPs was detected during the study but CPPs presented as documents were seen as more empathetic and credible than DPs, which makes us think that CPPs were more easily perceived as more human-like.

One important aspect from our findings is that, traditionally, a key risk in HIUP development is the question, "Are the personas informative enough?" (e.g., the PPS only involves items gauging the completeness of a persona from the perspective of having "enough" information, not "too much" of it [71]). Now, with the current study, and with the help of AI technology, it is easier to generate HIUPs with a lot of information because we do not need humans to curate the HIUP content. This now poses a new issue, which is that the HIUPs can become *too* informative. For the AI participating in the HIUP creation process, managing a large amount of information is not an issue, but several of our participants were struggling with cognitive overload (or, information overload) which was clearly observable from their think-aloud records as well as our observations during the study completion. In conclusion, the findings postulate a strong connection between the level of information (especially the excess of it) and HIUP usability, suggesting that too much information decreases usability, even when it is in a compact format with tables and statistical summaries.

The observed divergence in user preferences for HIUP presentation formats can be further explained by theoretical frameworks such as user control theory [30] and cognitive load theory [78]. User control theory posits that individuals prefer interfaces that allow them to manage and access information in ways that align with their cognitive styles and task requirements. This aligns with our finding that some users appreciated the structured and, presumably, exhaustive nature of tables, which provided users with a sense of control over the HIUP information. Conversely, others preferred the minimalist, conversational nature of DPs, as this interface reduced decision-making complexity, consistent with cognitive load theory. Cognitive load theory emphasizes the importance of balancing informational density to avoid overwhelming users, which is supported by our observation of cognitive overload in participants exposed to excessively detailed CPPs. Minimizing cognitive overload in interactive personas, especially those involving AI, can introduce problems with the AI hallucinating answers for the user [42, 43] which, again, can lead to trust problems yet not present in our study. These theoretical perspectives emphasize the importance of tailoring HIUP designs to user needs by optimizing the balance between informativeness and simplicity.

Yet another aspect of personas, in general, is the change of modality, i.e., the form of presenting the persona. Previous studies show different ways of presenting personas as narratives (text), conventional persona profiles, deepfake personas, and human personas [42–45]. These studies also present different levels of information-carrying capacity in the personas and different persona perceptions. Our study contributes to the persona discussion by offering insight into how CPPs and DPs differ in information retrieval tasks which is one of the uses of personas [17, 61]. Also, our findings resonate with research in CUIs, where user satisfaction is closely linked to perceptions of interaction quality. Even minor design choices in conversational agents influence how human-like they are perceived [16, 23]. In our study, while DPs offered dynamic interaction, CPPs were seen as more empathetic and credible likely due to CPPs' coherent, narrative structure.

5.3 Design Implications

There is no clear "winner" in pitting DPs against the classic CPPs. Rather, both have strengths and weaknesses. Therefore, their use is largely mandated by the specific use case (as well as individual differences in preferences, which should not be overlooked). In terms of using each HIUP interaction modality, we offer the following guidelines:

Dialogue Modality: The DP's strength is its ability to give answers to specific questions making the DP efficient for fast information search from the HIUP. The DP is preferred by users accustomed to fast information search with little attention to the graphical aspects of the HIUP in a task given to the users in this study.

Document Modality: The CPP's strength is its visual appeal, i.e., people preferring nice-to-look-at things will most likely prefer the CPP. The CPP is better for use cases when a quick overlook should be made of the HIUP. The CPP is also better for use cases where the HIUP user has large amounts of data to observe and when there are no specific questions to be asked about the HIUP. The CPP allows the user to browse through for information.

Document and Dialogue Modalities: Then again, the choice between the DP and the CPP is more of a false dichotomy since interactive persona systems, like the S2P, offer the deployment of both HIUP modalities within one system, affording users the ability to switch modalities. This allows the users to adjust their usage preferences for both HIUP modalities. In our study, the two HIUP modalities were kept separate, exposing the user to one HIUP modality at a time to maintain study controllability.

These findings have practical implications for the design of AI systems across various domains. For example, DPs could be particularly useful in customer service or educational contexts, where real-time, intuitive interaction enhances engagement and reduces cognitive load. The structured and comprehensive format of CPPs, on the other hand, may be better suited for tasks requiring detailed analysis, such as decision-making in knowledge management systems. Interestingly, the unexpectedly higher success rate for CPPs in task completion suggests that despite their perceived complexity, users may ultimately benefit from the additional depth of information CPPs provide. Designers could explore hybrid systems that combine the immediacy of DPs with the completeness of CPPs, potentially mitigating challenges such as cognitive overload and stereotypicality that could be observed in each modality individually. Future research could investigate adaptive persona systems that dynamically adjust interaction styles based on user preferences and task requirements.

5.4 Central Limitations and Avenues for Future Work

There are limitations and areas for further work. First, the type of task could limit the generalizability of the findings. The information retrieval task emphasized a specific use case in which the stakeholder has a given set of questions to which they want answers. This might always be the case; certainly, there are cases where the stakeholder wants to browse the HIUP's information freely without having a predefined list of questions. So, this must be considered as a limitation when applying our findings. However, information retrieval is a central task in HIUP usage, so the evaluation of this task is certainly a strength of our research. Future work can consider other use cases beyond information retrieval, including those where free browsing is enabled. The advantage, however, of using a predefined list of questions is that it gives a clear purpose and structure for the experiment. So, researchers need to consider this trade-off of framing versus non-framing the participant's interaction with the HIUP. There are grounds for both approaches.

Second, our sample of participants could be expanded to cover HIUP users of different professions and backgrounds. This could help expand the understanding of the wider user group beyond the users we were studying.

Third, the effects of DP hallucinating answers to users on task performance remains unexplored, leaving room for future research. Hallucinations could be a key reason why AI should not be relied upon as the primary source of information or information interpretation.

Fourth, the use of CPPs and DPs could vary when the HIUPs are used in collaborative or personal tasks. The effect of collaborative versus personal task use of different HIUP modalities is left for another research.

Fifth, the effects of user's gender on usability and perceptions of HIUPs were left outside of our study as a remaining avenue for future research. Similarly, studying difficulty of use, confidence with the system, or potential DP hallucinations effect on task performance could be further studied.

Sixth, future research could also study the information richness of different HIUP modalities which was left outside of our study as well as the very root of the differences in the manners of information retrieval from CPPs (search) and DP (asking questions). CPP and DP have different ways for information retrieval potentially altering their use for such tasks.

Seventh, another promising direction for future research involves integrating more closely the frameworks and methods from the field of CUI, particularly in evaluating how users perceive humanness, trust, and empathy across different HIUP modalities.

Eighth, future work could explore how varying levels of interactivity or proactivity in DPs affect user satisfaction and cognitive load, building on findings from CUI studies that examine proactive voice assistants [63, 81].

Lastly, future work could look into hybrid HIUP formats; e.g., combining CPP overviews with conversational agents to strike a balance between coherence, usability, and humanlike interaction.

6 Conclusion

In this study, we investigated two HIUP modalities for information retrieval: a CPP and a DP. With the advent of vast usage of AI, new HIUP modalities have emerged enabling variation to information search methods from and learning of HIUPs. Our results imply that both HIUP modalities tested have their advantages and the perception of the HIUP is largely dependent on (a) the use case of the HIUP and (b) the HIUP user. Modern interactive persona systems can combine the best properties of CPPs, DPs, and other HIUP modalities. More research could be put to study the need and level of interaction between the user and the HIUP in different use cases, tasks, and work scenarios.

References

[1] Abeer Alessa and Hend Al-Khalifa. 2023. Towards Designing a ChatGPT Conversational Companion for Elderly People. In *arXiv preprint arXiv:2304.09866*, 2023. .

[2] Bryce Allen. 1990. Knowledge organization in an information retrieval task. *Information processing & management* 26, 4 (1990), 535–542.

[3] Jisun An, Haewoon Kwak, Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2018. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining* 8, 1 (2018). https://doi.org/10.1007/s13278-018-0531-0

[4] Jisun An, Haewoon Kwak, Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2018. Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web (TWEB)* 12, 3 (2018).

[5] Farshid Anvari, Deborah Richards, Michael Hitchens, and Muhammad Ali Babar. 2015. Effectiveness of Persona with Personality Traits on Conceptual Design. In *Proceedings of the 37th International Conference on Software Engineering - Volume 2* (*ICSE '15*), 2015. IEEE Press, Piscataway, NJ, USA, 263–272. Retrieved July 11, 2018 from http://dl.acm.org/citation.cfm?id$=$2819009.2819048

[6] Mikio Aoyama. 2005. Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In *13th IEEE International Conference on Requirements Engineering (RE'05)*, 2005. IEEE, Paris, France, 85–94. https://doi.org/10.1109/RE.2005.50

[7] Theo Araujo. 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, (August 2018), 183–189. https://doi.org/10.1016/j.chb.2018.03.051

[8] A. Baki Kocaballi. 2023. Conversational AI-Powered Design: ChatGPT as Designer, User, and Product. *arXiv e-prints* (2023), arXiv-2302.

[9] David Bamman, Brendan O'Connor, and Noah A. Smith. 2013. Learning latent personas of film characters. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2013. 352–361.

[10] Susanne Bødker, Ellen Christiansen, Tom Nyvang, and Pär-Ola Zander. 2012. Personas, people and participation: challenges from the trenches of local government. In *Proceedings of the 12th Participatory Design Conference on Research Papers: Volume 1 - PDC '12*, 2012. ACM Press, Roskilde, Denmark, 91. https://doi.org/10.1145/2347635.2347649

[11] Jonalan Brickey, Steven Walczak, and Tony Burgess. 2012. Comparing semi-automated clustering methods for persona development. *IEEE Transactions on Software Engineering* 38, 3 (2012), 537–546.

[12] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189, (November 1995).

[13] Nancy Carter, Denise Bryant-Lukosius, Alba DiCenso, Jennifer Blythe, and Alan J. Neville. 2014. The Use of Triangulation in Qualitative Research. *Oncology Nursing Forum* 41, 5 (September 2014), 545–547. https://doi.org/10.1188/14.ONF.545-547

[14] Chris Chapman and Russell P. Milham. 2006. The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, October 01, 2006. 634–636. https://doi.org/10.1177/154193120605000503

[15] Herbert H. Clark. 2016. Depicting as a method of communication. *Psychological Review* 123, 3 (2016), 324–347. https://doi.org/10.1037/rev0000026

[16] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation?: Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, May 02, 2019. ACM, Glasgow Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300705

[17] Alan Cooper. 1999. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity* (1 edition ed.). Sams - Pearson Education, Indianapolis, IN.

[18] Amanda Cercas Curry, Judy Robertson, and Verena Rieser. 2020. Conversational assistants and gender stereotypes: Public perceptions and desiderata for voice personas. In *Proceedings of the second workshop on gender bias in natural language processing*, 2020. 72–78.

[19] Kieran Cutting and Erkki Hedenborg. 2019. Can Personas Speak? Biopolitics in Design Processes. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, 2019. 153–157.

[20] Stefano De Paoli. 2023. Writing user personas with Large Language Models: Testing phase 6 of a Thematic Analysis of semi-structured interviews. *arXiv preprint arXiv:2305.18099* (2023).

[21] Ewart J. De Visser, Samuel S. Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E. McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.

[22] Ameet Deshpande, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. 2023. Toxicity in ChatGPT: Analyzing Persona-assigned Language Models. Retrieved March 10, 2024 from http://arxiv.org/abs/2304.05335

[23] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, October 2019. ACM, Taipei Taiwan, 1–12. https://doi.org/10.1145/3338286.3340116

[24] Janna Lynn Dupree, Richard Devries, Daniel M. Berry, and Edward Lank. 2016. Privacy personas: Clustering users via attitudes and behaviors toward security practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. ACM, 5228–5239.

[25] Shamal Faily. 2015. Engaging stakeholders during late stage security design with assumption personas. *Information & Computer Security* 23, 4 (2015), 435–446.

[26] Shamal Faily and Ivan Flechais. 2011. Persona Cases: A Technique for Grounding Personas. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '11*), 2011. ACM, New York, NY, USA, 2267–2270. https://doi.org/10.1145/1978942.1979274

[27] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies* 132, (December 2019), 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

[28] Mauajama Firdaus, Naveen Thangavelu, Asif Ekbal, and Pushpak Bhattacharyya. 2022. I enjoy writing and playing, do you: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network. *IEEE Transactions on Affective Computing* (2022).

[29] Erin Friess. 2012. Personas and Decision Making in the Design Process: An Ethnographic Case Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '12*), 2012. ACM, New York, NY, USA, 1209–1218. https://doi.org/10.1145/2207676.2208572

[30] William Glasser. 1986. *Control theory in the classroom*. Perennial Library/Harper & Row Publishers, New York, NY, US.

[31] Chu Hiang Goh, Narayanan Kulathuramaiyer, and Tariq Zaman. 2017. Riding Waves of Change: A Review of Personas Research Landscape Based on the Three Waves of HCI. In *Information and Communication Technologies for Development* (*IFIP Advances in Information and Communication Technology*), 2017. Springer International Publishing, Cham, 605–616. https://doi.org/10.1007/978-3-319-59111-7_49

[32] Jonathan Grudin. 2006. Why personas work: The psychological evidence. *The Persona Lifecycle* (2006), 642–663.

[33] Jonathan Grudin and John Pruitt. 2002. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. In *Proceedings of Participation and Design Conference (PDC2002)*, 2002. Sweden, 8.

[34] Bernard J. Jansen, Soon-gyo Jung, and Joni Salminen. 2020. From flat file to interface: Synthesis of personas and analytics for enhanced user understanding. *Proceedings of the Association for Information Science and Technology* 57, 1 (October 2020). https://doi.org/10.1002/pra2.215

[35] Hang Jiang, Xiajie Zhang, Xubo Cao, Jad Kabbara, and Deb Roy. 2023. Personallm: Investigating the ability of gpt-3.5 to express personality traits and gender differences. *arXiv preprint arXiv:2305.02547* (2023).

[36] Soon-Gyo Jung, Jisun An, Haewoon Kwak, Moeed Ahmad, Lene Nielsen, and Bernard J. Jansen. 2017. Persona Generation from Aggregated Social Media Data. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (*CHI EA '17*), 2017. ACM, New York, NY, USA, 1748–1755.

[37] Soon-Gyo Jung, Joni Salminen, Kholoud Khalil Aldous, and Bernard J. Jansen. 2025. PersonaCraft: Leveraging language models for data-driven persona development. *International Journal of Human-Computer Studies* 197, (March 2025), 103445. https://doi.org/10.1016/j.ijhcs.2025.103445

[38] Soon-Gyo Jung, Joni Salminen, Jisun An, Haewoon Kwak, and B. J. Jansen. 2018. Automatically Conceptualizing Social Media Analytics Data via Personas. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2018)*, June 25, 2018. San Francisco, California, USA.

[39] Soon-Gyo Jung, Joni Salminen, and Bernard J. Jansen. 2020. Giving Faces to Data: Creating Data-Driven Personas from Personified Big Data. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (*IUI '20*), March 17, 2020. Association for Computing Machinery, Cagliari, Italy, 132–133. https://doi.org/10.1145/3379336.3381465

[40] Soon-Gyo Jung, Joni Salminen, Haewoon Kwak, Jisun An, and Bernard J. Jansen. 2018. Automatic Persona Generation (APG): A Rationale and Demonstration. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*, 2018. ACM, 321–324.

[41] Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, Hind Almerekhi, and Bernard J. Jansen. 2023. How Do Users Perceive Deepfake Personas? Investigating the Deepfake User Perception and Its Implications for Human-Computer Interaction. In *Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter* (*CHItaly '23*), September 20, 2023. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3605390.3605397

[42] Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, João M. Santos, Essi Häyhänen, Trang Xuan, Jinan Azem, and Bernard J. Jansen. 2024. Modeling the New Modalities of Personas: How Do Users' Attributes Influence Their Perceptions and Use of Interactive Personas? In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization* (*UMAP Adjunct '24*), June 28, 2024. Association for Computing Machinery, New York, NY, USA, 164–169. https://doi.org/10.1145/3631700.3664882

[43] Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, Trang Thi Thu Xuan, Essi Häyhänen, Jinan Y. Azem, and Bernard J. Jansen. 2025. "You Always Get an Answer": Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination. In *Proceedings of the 30th International*

Conference on Intelligent User Interfaces, March 24, 2025. ACM, Cagliari Italy, 1624–1638. https://doi.org/10.1145/3708359.3712160

[44] Ilkka Kaate, Joni Salminen, Joao Santos, Soon-Gyo Jung, Rami Olkkonen, and Bernard Jansen. 2023. The Realness of Fakes: Primary Evidence of the Effect of Deepfake Personas on User Perceptions in a Design Task. *International Journal of Human-Computer Studies* (June 2023), 103096. https://doi.org/10.1016/j.ijhcs.2023.103096

[45] Ilkka Kaate, Joni Salminen, João M. Santos, Soon-Gyo Jung, Hind Almerekhi, and Bernard J. Jansen. 2024. "There Is something Rotten in Denmark": Investigating the Deepfake persona perceptions and their Implications for human-centered AI. *Computers in Human Behavior: Artificial Humans* 2, 1 (January 2024), 100031. https://doi.org/10.1016/j.chbah.2023.100031

[46] Mandar Kulkarni, Praveen Tangarajan, Kyung Kim, and Anusua Trivedi. 2024. Reinforcement Learning for Optimizing RAG for Domain Chatbots. Retrieved March 14, 2024 from http://arxiv.org/abs/2401.06800

[47] James R. Lewis and Jeff Sauro. 2009. The Factor Structure of the System Usability Scale. In *Proceedings of the 1st International Conference on Human Centered Design: Held As Part of HCI International 2009* (*HCD 09*), 2009. Springer-Verlag, Berlin, Heidelberg, 94–103. https://doi.org/10.1007/978-3-642-02806-9_12

[48] James Lewis and Jeff Sauro. 2018. Item Benchmarks for the System Usability Scale. 13, (May 2018), 158–167.

[49] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020. 1–15.

[50] Frank Long. 2009. Real or imaginary: The effectiveness of using personas in product design. In *Proceedings of the Irish Ergonomics Society Annual Conference*, 2009. Irish Ergonomics Society Dublin.

[51] Tara Matthews, Tejinder Judge, and Steve Whittaker. 2012. How Do Designers and User Experience Professionals Actually Perceive and Use Personas? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '12*), 2012. ACM, New York, NY, USA, 1219–1228. https://doi.org/10.1145/2207676.2208573

[52] Jennifer Jen McGinn and Nalini Kotamraju. 2008. Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008. ACM, 1521–1524.

[53] T. Mijač, M. Jadrić, and M. Ćukušić. 2018. The potential and issues in data-driven development of web personas. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018. 1237–1242. https://doi.org/10.23919/MIPRO.2018.8400224

[54] Jakob Nielsen. 2010. *Usability engineering* (3.Nachdr. ed.). Kaufmann, Amsterdam Heidelberg.

[55] Lene Nielsen. 2004. *Engaging personas and narrative scenarios*. Samfundslitteratur.

[56] Lene Nielsen. 2019. *Personas - User Focused Design* (2nd ed. 2019 edition ed.). Springer, New York, NY.

[57] Lene Nielsen, Kira Storgaard Hansen, Jan Stage, and Jane Billestrup. 2015. A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations. *International Journal of Sociotechnology and Knowledge Development* 7, 1 (2015), 45–61. https://doi.org/10.4018/ijskd.2015010104

[58] Lene Nielsen, Joni Salminen, Soon-Gyo Jung, and Bernard J. Jansen. 2021. Think-Aloud Surveys. In *IFIP Conference on Human-Computer Interaction*, 2021. Springer, Cham, 504–508.

[59] Lene Nielsen and Kira Storgaard Hansen. 2014. Personas is applicable: a study on the use of personas in Denmark. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014. ACM, 1665–1674.

[60] James E. Nieters, Subbarao Ivaturi, and Iftikhar Ahmed. 2007. Making Personas Memorable. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (*CHI EA '07*), 2007. ACM, New York, NY, USA, 1817–1824. https://doi.org/10.1145/1240866.1240905

[61] John Pruitt and Jonathan Grudin. 2003. Personas: Practice and Theory. In *Proceedings of the 2003 Conference on Designing for User Experiences* (*DUX '03*), 2003. ACM, San Francisco, California, USA, 1–15. https://doi.org/10.1145/997078.997089

[62] Lee Rainie, Cary Funk, Monica Anderson, and Alce Tyson. 2022. *AI and Human Enhancement: Americans' Openness Is Tempered by a Range of Concerns*. Pew Research Center. Retrieved from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf

[63] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions on Proactive Smart Speakers. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*, July 27, 2021. ACM, Bilbao (online) Spain, 1–10. https://doi.org/10.1145/3469595.3469629

[64] Joni Salminen, Kathleen Guan, Lene Nielsen, Soon-gyo Jung, Shammur Absar Chowdhury, and Bernard J. Jansen. 2020. A Template for Data-Driven Personas: Analyzing 31 Quantitatively Oriented Persona Profiles. In *Human Interface and the Management of Information. Designing Information. HCII 2020.*, S. Yamamoto and H. Mori (eds.). Springer, Copenhagen, Denmark, 125–144.

[65] Joni Salminen, Bernard Jansen, and Soon-Gyo Jung. 2022. Survey2Persona: Rendering Survey Responses as Personas. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (*UMAP '22 Adjunct*), July 04, 2022. Association for Computing Machinery, New York, NY, USA, 67–73. https://doi.org/10.1145/3511047.3536403

[66] Joni Salminen, Soon-gyo Jung, Hind Almerekhi, Erik Cambria, and Bernard Jansen. 2023. How Can Natural Language Processing and Generative AI Address Grand Challenges of Quantitative User Personas? In *HCI International 2023 – Late Breaking Papers* (*Lecture Notes in Computer Science*), 2023. Springer Nature Switzerland, Cham, 211–231. https://doi.org/10.1007/978-3-031-48057-7_14

[67] Joni Salminen, Soon-Gyo Jung, and Bernard Jansen. 2022. Developing Persona Analytics Towards Persona Science. In *27th International Conference on Intelligent User Interfaces* (*IUI '22*), March 22, 2022. Association for Computing Machinery, New York, NY, USA, 323–344. https://doi.org/10.1145/3490099.3511144

[68] Joni Salminen, Soon-Gyo Jung, João M. Santos, and Bernard J. Jansen. 2020. Does a Smile Matter if the Person Is Not Real?: The Effect of a Smile and Stock Photos on Persona Perceptions. *International Journal of Human–Computer Interaction* 36, 6 (April 2020), 568–590. https://doi.org/10.1080/10447318.2019.1664068

[69] Joni Salminen, Soon-gyo Jung, João M. Santos, Ahmed Mohamed Kamel, and Bernard J. Jansen. 2021. Picturing It!: The Effect of Image Styles on User Perceptions of Personas. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2021. ACM, Virtual conference.

[70] Joni Salminen, Ying-Hsang Liu, Sercan Sengun, João M. Santos, Soon-gyo Jung, and Bernard J. Jansen. 2020. The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces. In *IUI '20: Proceedings of the 25th International Conference on Intelligent User Interfaces*, March 17, 2020. ACM, Cagliary, Italy, 357–368. https://doi.org/10.1145/3377325.3377492

[71] Joni Salminen, Joao M. Santos, Haewoon Kwak, Jisun An, Soon-gyo Jung, and Bernard J. Jansen. 2020. Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas. *International Journal of Human-Computer Studies* 141, (April 2020), 102437. https://doi.org/10.1016/j.ijhcs.2020.102437

[72] Joni Salminen, Sercan Sengun, Soon-gyo Jung, and Bernard J Jansen. 2019. Design Issues in Automatically Generated Persona Profiles: A Qualitative Analysis from 38 Think-Aloud Transcripts. In *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, March 13, 2019. ACM, Glasgow, UK, 225–229. https://doi.org/10.1145/3295750.3298942

[73] Joni Salminen, Sercan Þengün, Soon-Gyo Jung, and Bernard J. Jansen. 2021. Comparing Persona Analytics and Social Media Analytics for a User-Centric Task Using Eye-Tracking and Think-Aloud. In *CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter* (*CHItaly '21*), July 11, 2021. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3464385.3464734

[74] Anuschka Schmitt, Thiemo Wambsganss, and Jan Marco Leimeister. 2022. Conversational Agents for Information Retrieval in the Education Domain: A User-Centered Design Investigation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (November 2022), 1–22. https://doi.org/10.1145/3555587

[75] Tobias Schnabel, Saleema Amershi, Paul N. Bennett, Peter Bailey, and Thorsten Joachims. 2020. The Impact of More Transparent Interfaces on Behavior in Personalized Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, July 25, 2020. ACM, Virtual Event China, 991–1000. https://doi.org/10.1145/3397271.3401117

[76] Sercan Şengün, Joni Salminen, Soon-Gyo Jung, Kholoud Khalil Aldous, and Bernard J. Jansen. 2024. "There's Something About Noura": Exploring Think-Aloud Reasonings for Users' Persona Choice in a Design Task. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (*DIS '24*), July 01, 2024. Association for Computing Machinery, New York, NY, USA, 1234–1247. https://doi.org/10.1145/3643834.3661512

[77] Michael Shumanov and Lester Johnson. 2021. Making conversations with chatbots more personalized. *Computers in Human Behavior* 117, (April 2021), 106627. https://doi.org/10.1016/j.chb.2020.106627

[78] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.

[79] N. Tu, X. Dong, P. P. Rau, and T. Zhang. 2010. Using cluster analysis in Persona development. In *Proceedings of the 8th International Conference on Supply Chain Management and Information*, October 2010. IEEE, 1–5.

[80] Aleksei Turobov, Diane Coyle, and Verity Harding. 2024. Using ChatGPT for Thematic Analysis. https://doi.org/10.48550/ARXIV.2405.08828

[81] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces*, July 26, 2022. ACM, Glasgow United Kingdom, 1–14. https://doi.org/10.1145/3543829.3543834

[82] Wenjia Zhang, Lin Gui, Rob Procter, and Yulan He. 2023. NewsQuote: A Dataset Built on Quote Extraction and Attribution for Expert Recommendation in Fact-Checking. https://doi.org/10.48550/ARXIV.2305.04825

[83] Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. 2016. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '16*), 2016. ACM, New York, NY, USA, 5350–5359.

[84] Xishuo Zhang, Lin Liu, Yi Wang, Xiao Liu, Hailong Wang, Anqi Ren, and Chetan Arora. 2023. PersonaGen: A Tool for Generating Personas from User Feedback.

arXiv preprint arXiv:2307.00390 (2023).
[85] Haining Zhu, Hongjian Wang, and John M. Carroll. 2019. Creating Persona Skeletons from Imbalanced Datasets - A Case Study using U.S. Older Adults' Health Data. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (*DIS '19*), 2019. ACM, New York, NY, USA, 61–70. https://doi.org/10.1145/3322276.3322285

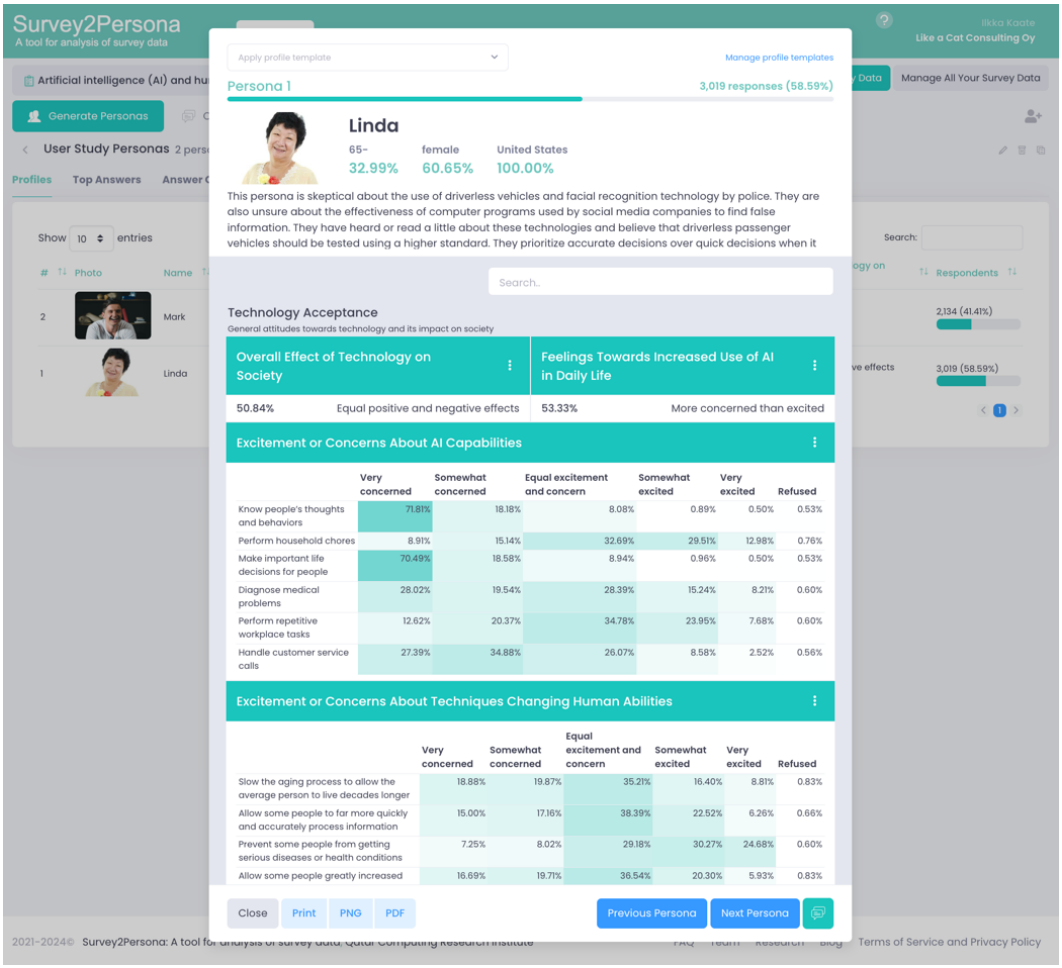
Appendix 1: Survey items

Table A1: Survey items with variables and scales.

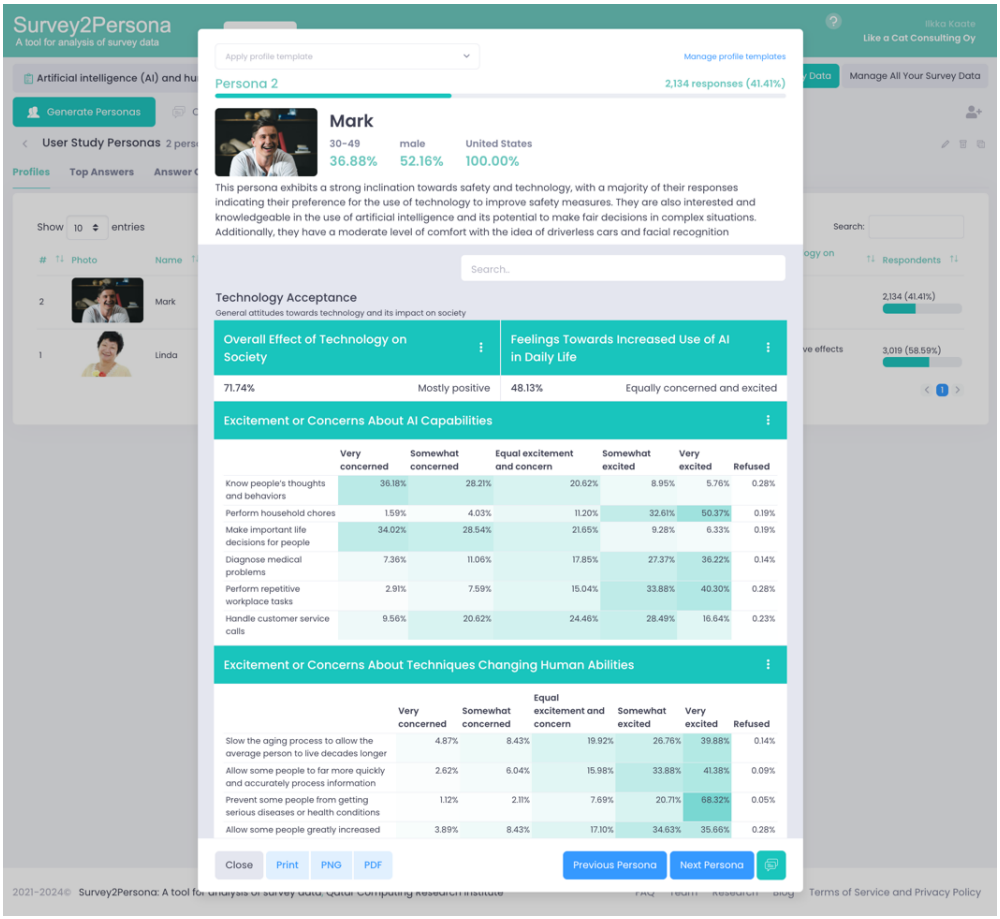
Scale	Item	Variable	Scale
PPS [71]	The information about the persona was presented clearly.	Clarity	7-point Likert
	I struggled to understand the information about the persona.	Clarity	7-point Likert
	The information about the persona was easy to understand.	Clarity	7-point Likert
	I experienced sympathetic concern for the sufferings or misfortunes of the persona.	Compassion	7-point Likert
	I did not really care about the persona.	Compassion	7-point Likert
	I had an urgent desire to aid the persona.	Compassion	7-point Likert
	The persona provided enough information to make decisions about the people it describes.	Completeness	7-point Likert
	The persona was detailed enough to understand the type of people it describes.	Completeness	7-point Likert
	The persona lacked critical information for my task.	Completeness	7-point Likert
	Some parts of the persona were contradicting each other.	Consistency	7-point Likert
	The persona communicated a coherent story.	Consistency	7-point Likert
	The persona was consistent.	Consistency	7-point Likert
	The persona could exist in real life.	Credibility	7-point Likert
	The persona had artifacts; i.e., something artificial, a distortion.	Credibility	7-point Likert
	The persona appeared natural.	Credibility	7-point Likert
	I felt like I understood the persona as a human being.	Empathy	7-point Likert
	I did not feel strong ties to the persona.	Empathy	7-point Likert
	I could imagine a day in the life of the persona.	Empathy	7-point Likert
	The persona felt similar to me.	Similarity	7-point Likert
	The persona and I think very differently.	Similarity	7-point Likert
	The persona and I share similar interests.	Similarity	7-point Likert
	The persona was stereotypical, i.e., it related to a widely held but fixed and oversimplified image or idea of a particular type of person.	Stereotypicality	7-point Likert
	The persona conformed to qualities that people usually expect of a particular type of person.	Stereotypicality	7-point Likert
	The persona contained surprising insights into the type of person it represents.	Stereotypicality	7-point Likert
	I was provided with information on how the persona was created.	Transparency	7-point Likert
	I did not understand how the persona was created.	Transparency	7-point Likert
	I could understand how the information about the persona was obtained.	Transparency	7-point Likert
	Using the persona required a lot of mental effort.	Usability	7-point Likert
	I found the persona easy to use.	Usability	7-point Likert
	Using the persona was clear and understandable.	Usability	7-point Likert
	If given the choice, I would not have used this persona for the task I was given.	Willingness to use	7-point Likert
	I can imagine multiple ways to make use of the persona's information in the task I was given.	Willingness to use	7-point Likert
	This persona improved my ability to make decisions about the people it describes.	Willingness to use	7-point Likert
	Do you have any other thoughts you would like to share about the persona?	PPS open-ended	7-point Likert
SUS [47]	I think that I would like to use this persona system for similar tasks.	System Usability Score (SUS) 1	7-point Likert

I found the persona system unnecessarily complex.	System Usability Score (SUS) 2	7-point Likert
I found the persona system easy to use.	System Usability Score (SUS) 3	7-point Likert
I think that I would need the support of a professional to be able to use the persona system.	System Usability Score (SUS) 4	7-point Likert
I found the various functions in the persona system were well integrated.	System Usability Score (SUS) 5	7-point Likert
I found there was too much inconsistency in the persona system.	System Usability Score (SUS) 6	7-point Likert
I would imagine that most people would learn to use the persona system very quickly.	System Usability Score (SUS) 7	7-point Likert
I found the persona system very difficult to use.	System Usability Score (SUS) 8	7-point Likert
I felt very confident using the persona system.	System Usability Score (SUS) 9	7-point Likert
I needed to learn a lot of things before I could get going with the persona system.	System Usability Score (SUS) 10	7-point Likert
The persona system was fun to use.	System Usability Score (SUS) 11 (own question)	7-point Likert
The task was difficult to complete using the persona system.	System Usability Score (SUS) task difficulty	7-point Likert
Do you have any other thoughts you would like to share about the persona system?	System Usability Score (SUS) open-ended	open-ended

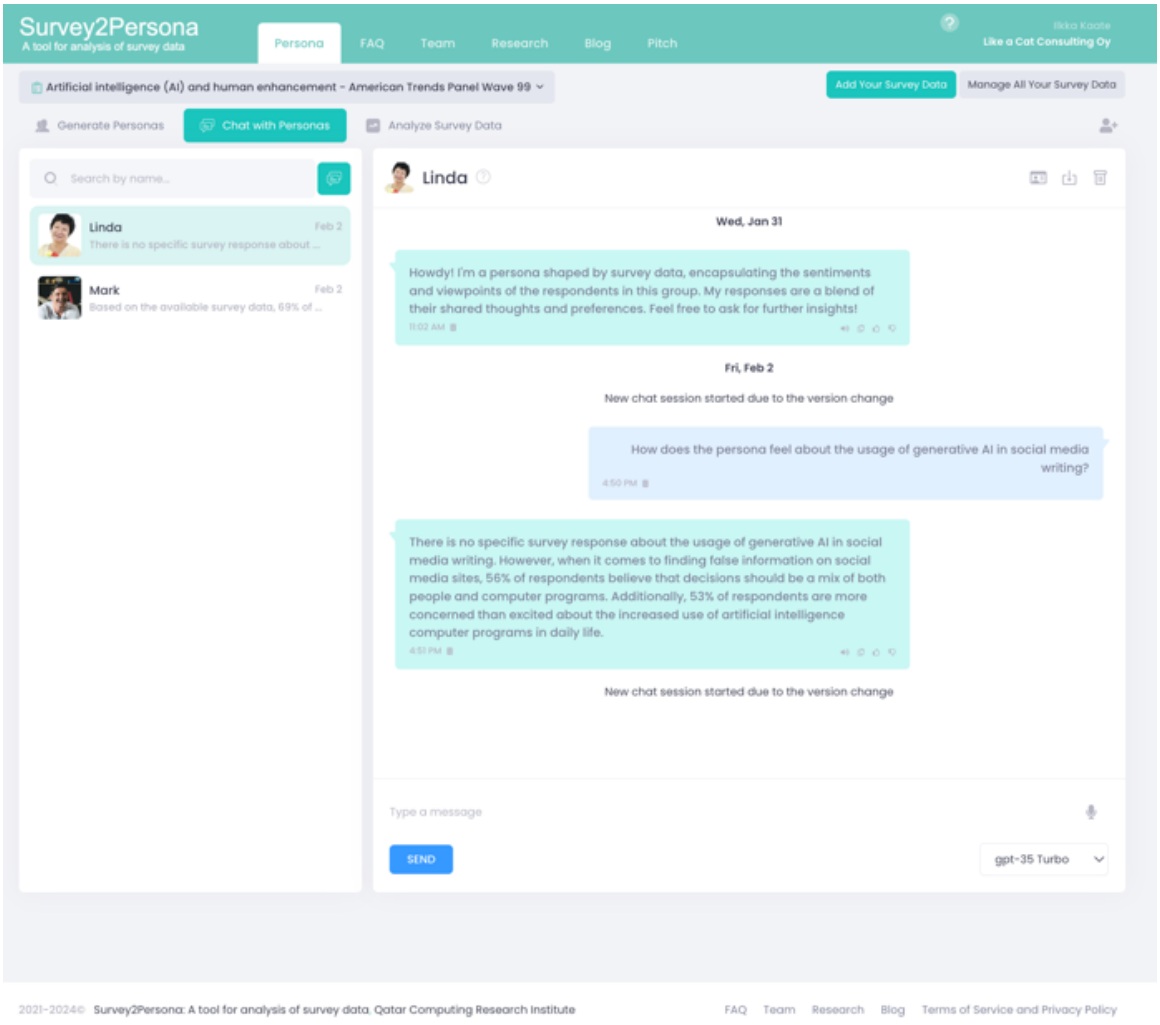
Appendix 2: The personas Used in the Study



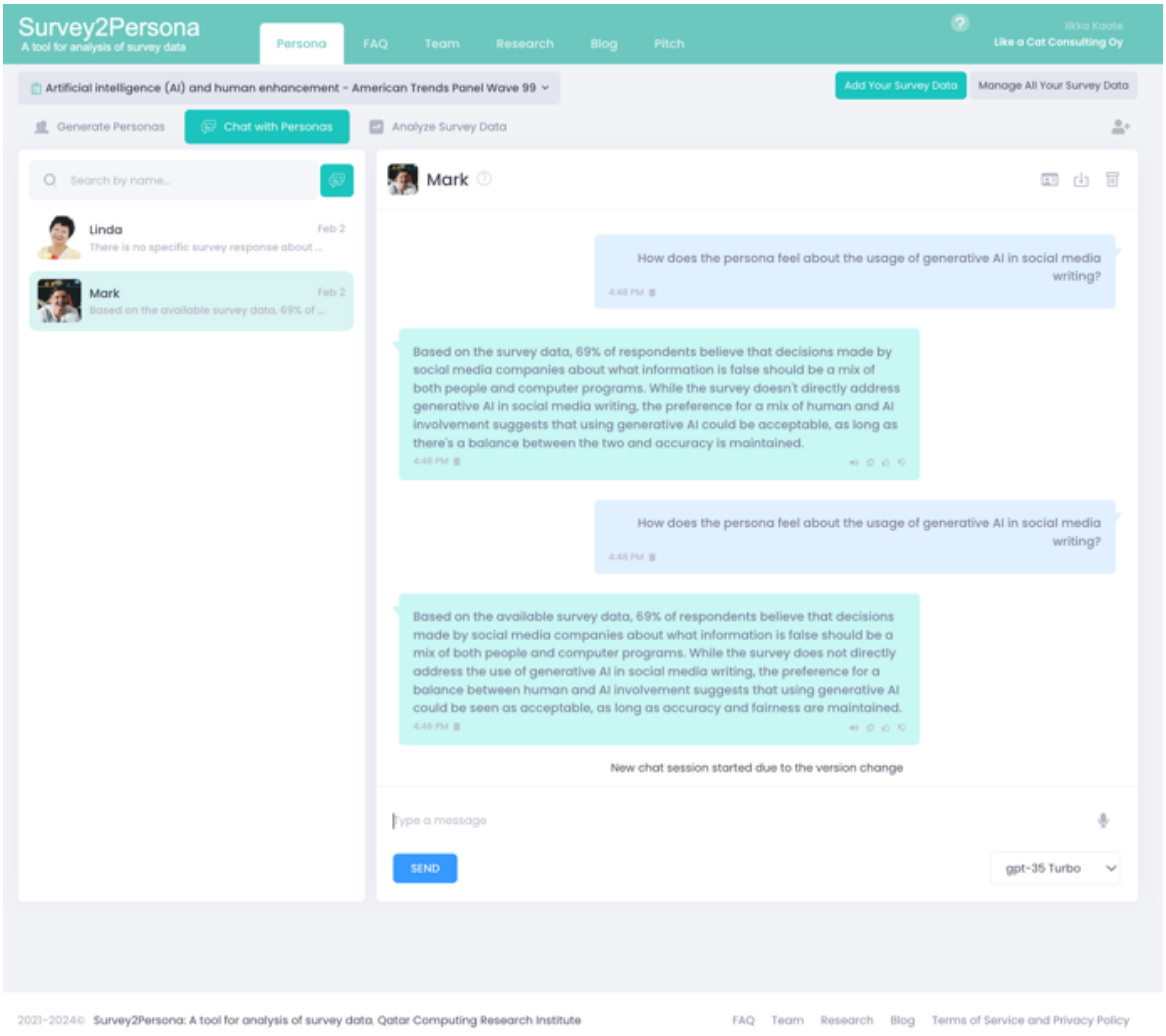
Appendix 2A: Persona profile Linda



Appendix 2B: Persona profile Mark



Appendix 2C: Chat persona Linda



Appendix 2D: Chat persona Mark