Behaviour & Information Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tbit20

The 'fourth wall' and other usability issues in AI-generated personas: comparing chat-based and profile personas

Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, João M. Santos, Essi Häyhänen, Trang Xuan, Jinan Y. Azem & Bernard Jansen

To cite this article: Ilkka Kaate, Joni Salminen, Soon-Gyo Jung, João M. Santos, Essi Häyhänen, Trang Xuan, Jinan Y. Azem & Bernard Jansen (27 Feb 2025): The 'fourth wall' and other usability issues in AI-generated personas: comparing chat-based and profile personas, Behaviour & Information Technology, DOI: 10.1080/0144929X.2025.2469659

To link to this article: https://doi.org/10.1080/0144929X.2025.2469659

Published online: 27 Feb 2025.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2025.2469659

The 'fourth wall' and other usability issues in AI-generated personas: comparing chat-based and profile personas

Ilkka Kaate[a], Joni Salminen[b], Soon-Gyo Jung[c], João M. Santos[d], Essi Häyhänen[b], Trang Xuan[e], Jinan Y. Azem[f] and Bernard Jansen[f]

[a]Department of Marketing and International Business, University of Turku, Turun yliopisto, Finland; [b]University of Vaasa, Vaasa, Finland; [c]Qatar Computing Research Institute, Doha, Qatar; [d]Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal; [e]Swinburne University of Technology, Melbourne, Australia; [f]Qatar Computing Research Institute, Doha, Qatar

ABSTRACT

Large Language Models (LLMs) are emerging as a powerful tool for AI-generated personas. This study evaluates the usability of AI-generated personas, comparing chat and profile formats. The findings indicate chat personas tend to be perceived more favourably, and profile personas exhibit greater variability in user perception. The increased difficulty and longer dwell time experienced by users with the profile persona, despite negative usability metrics, paradoxically resulted in better task performance. Usability issues indicate that many current limitations of AI, including verbosity, hallucinations, and empty rhetoric which was described as the persona having 'no soul', are inherited in AI-generated chat personas. However, there are also new issues. For one, the risk of information overload in an AI-generated profile persona implies that the AI does not consider human users' cognitive limitations when designing the persona (but usability scores for profile personas increase with dwell time, implying that users get used to the longer format the more time they spend). Another is the 'fourth wall' effect of AI-generated chat personas in which the user feels they are talking to someone describing the persona rather than the persona itself. Future work could address the usability paradox and the fourth wall effect of using personas.

CCS CONCEPTS: Human-centered computing Human computer interaction (HCI)

ARTICLE HISTORY
Received 24 May 2024
Accepted 5 February 2025

KEYWORDS
AI-generated personas; Personas; user interaction

1. Introduction

Personas are a notable technique of *user-centred design* (UCD), providing a fictional representation of a target user group to guide design decisions and empathy building (Cooper 1999; Nielsen 2019) among designers and other persona users ('users' henceforth). Traditionally, researchers created personas manually through user interviews and surveys (Adlin and Pruitt 2010; Grudin and Pruitt 2002; Mulder and Yaar 2006). However, the emergence of Artificial Intelligence (AI) (Xu et al. 2024) has introduced automated persona generation techniques (Alessa and Al-Khalifa 2023; De Paoli 2023; Zhang et al. 2023a) that leverage user data and advanced natural language processing (NLP) to create AI-generated personas (Salminen et al. 2023) (see examples in Figure 1). While AI-generated personas offer advantages in terms of speed and efficiency (An et al. 2018b; Jansen, Jung, and Salminen 2020; Jung et al. 2018; Kaate et al. 2024), their usability compared to traditional, manually crafted personas remains an open question.

While AI-generated personas offer a promising avenue for UCD, limited research explores their usability. Existing studies primarily focus on the validity and representativeness of AI-generated personas, with less attention paid to user interaction and experience (Zhang et al. 2023a, 2023b). Particularly, research on comparing different AI-generated persona formats (e.g. chat vs. profile) is scarce and still needs to be determined. This gap in knowledge hinders a comprehensive understanding of how users engage with AI-generated personas and how different formats influence UX.

To address this critical knowledge gap, the current research investigates the usability of AI-generated personas in the context of user interaction. We focus specifically on comparing usability and user perceptions with two common persona formats: chat and profile personas. Chat personas simulate conversation with a person (Alessa and Al-Khalifa 2023; Jiang et al. 2023; Kim and Im 2023); profile personas present information in a static format. We explore how users interact with

CONTACT Ilkka Kaate iokaat@utu.fi University of Turku, Turun yliopisto FI-20014, Finland

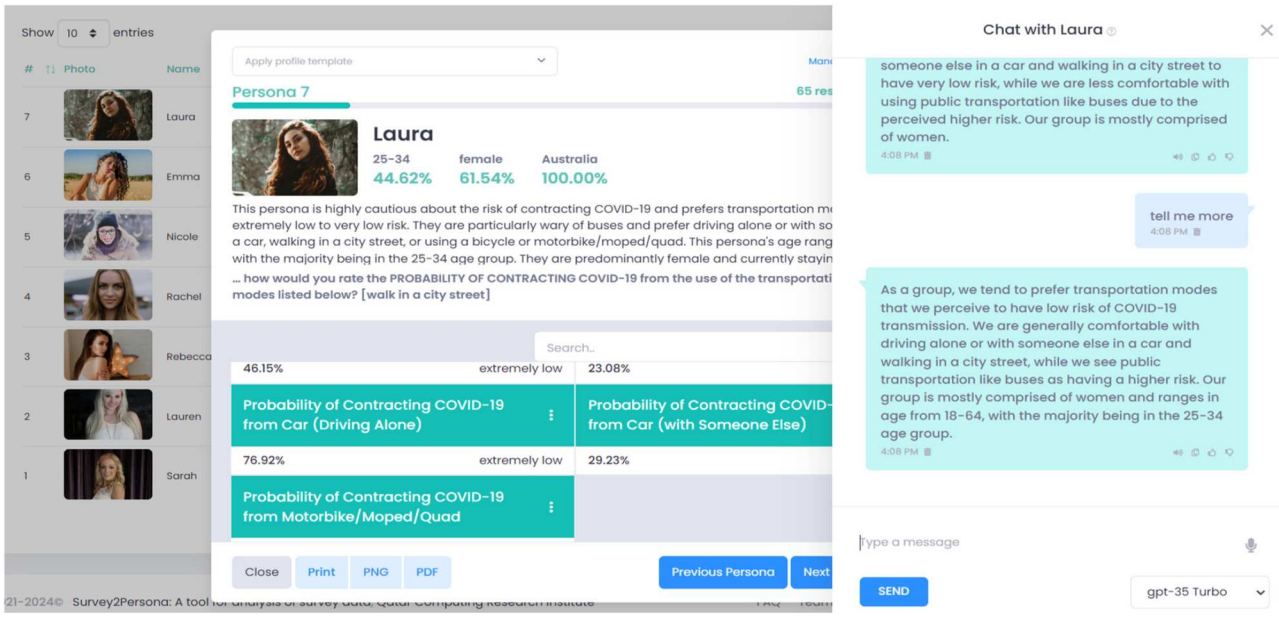


Figure 1. Chat-based (right-hand side) and profile personas (left-hand side). These two formats allow different interaction techniques The **chat** is, by nature, an interactive system where the user asks questions from the persona to gain knowledge about the persona. The user asks a question, the chat persona answers the question. For the **profile**, interaction is unidirectional. The user cannot discuss with the profile persona, only browse its information. Conventionally, the profile has the basic persona information (image, name, demographics, and description) visible while the chat persona has to be asked for the basic persona information.

these two important persona AI formats, dialogue and document. We address the following research questions (RQs):

> **RQ1:** Are there significant differences in user perceptions and interaction when using chat and profile personas generated by AI?
>
> **RQ2:** How is users' dwell time associated with their perception of AI-generated personas and their usability?
>
> **RQ3:** What are the central usability issues in AI-generated personas?

The division between chat and profile personas (see Figure 1 for illustration) allows us to investigate the impact of AI-generated persona approaches (conversation vs. static) on UX. These two formats representing two divergent approaches are particularly noteworthy – the profile format is a classic, standard way of presenting personas as templates, descriptions, or narratives (Nielsen et al. 2015; Salminen et al. 2020a). In contrast, the chat format is a slightly newer format, enabled specifically by large language models (LLMs) (Amin, Cambria, and Schuller 2023) and affording users direct, conversational access to the persona's information. Therefore, when investigating AI-generated personas, these two modalities are highly relevant. By understanding the usability strengths and weaknesses of different types of AI-generated personas, we can infer best practices for their integration into the UCD process. To this end, this research contributes to the ongoing academic research regarding the role of AI in design research, particularly in creating personas as a core UCD tool. By addressing the RQs, we shed light on the usability of AI-generated personas and identify potential benefits and drawbacks of different formats within the UCD process.

2. Related literature

AI-generated personas (Holzinger et al. 2022) are enabled by generative AI and progress in NLP, specifically LLMs (Amin, Cambria, and Schuller 2023). However, 'data-driven personas' or even 'automatically generated personas' are not in fact novel, but there are numerous studies on both (An, Kwak, and Jansen 2016; An et al. 2018a, 2018b; McGinn and Kotamraju 2008; Zhang, Brown, and Shankar 2016; Zhu, Wang, and Carroll 2019). The key difference is that even though there were automatically generated personas previously (An et al. 2018a, 2018b; Jung et al. 2018), these personas were never created in a fashion in which the user data would be automatically *interpreted* to write a persona description. Before the surge of LLMs, this process was mainly handled by human persona creators. Simultaneously, this limited algorithms' scope of contribution in the persona creation process and made them secondary to human reasoning.

However, this boundary is rapidly shifting with the rapid pace of development that LLMs like ChatGPT currently exhibit. Consequently, new examples of using AI in the persona creation process are continuously emerging. Here, we review such work.

De Paoli (2023) exemplifies LLMs' ability to analyze user data and generate creative interpretations for generating personas. The researcher applied ChatGPT for thematic analysis of user interviews, resulting in personas with goals, backgrounds, needs, and challenges. These 'AI-generated narratives' (De Paoli 2023) offer a compelling alternative to traditional personas, though the researcher warns against using them without cross-validating the information manually.

Another approach to AI-generated personas is the 'foundational model persona generation' where LLMs, without additional training, generate personas based on their inherent knowledge (Cheng, Durmus, and Jurafsky 2023). These 'vanilla personas' benefit from LLMs' vast knowledge bases (Amin, Cambria, and Schuller 2023) but risk stereotyping, necessitating validation ('persona triangulation' (Jansen et al. 2022)) to ensure real-world applicability. Concerns regarding factuality and potential misinformation, including biased representations or inaccurate information, or using LLMs unethically, remain a major risk (Alessa and Al-Khalifa 2023; Cheng, Durmus, and Jurafsky 2023; De Paoli 2023; Koch, Romero, and Stachl 2022).

LLMs can also be tailored to assume user roles, create dynamic dialogue systems (Alessa and Al-Khalifa 2023; Zhou et al. 2019), or predict a user role (Chatterjee et al. 2019). This enhances interactivity and UX by providing personalised responses. This anthropomorphisation of dialogue systems as personas offers an advantage over static persona profiles that cannot 'talk back' (Deshpande et al. 2023). Here, LLMs contribute to the appearance of complex, human-like cognition that can increase the persona's narrative realism and credibility, even when the personas contain factual errors about the users they represent (De Angelis et al. 2023).

The ultimate potential lies in integrating LLMs with user interfaces, APIs, and real user datasets for fully automated persona generation, realising the field's long-standing vision (An, Kwak, and Jansen 2016). The use of LLMs in AI-generated personas could also address the 'superficiality problem' of data-driven personas, where quantitative data leads to shallow profiles lacking qualitative depth (Salminen, Jung, and Jansen 2021, 2023). *PersonaGen* (Zhang et al. 2023a, 2023b) exemplifies how LLMs can automate data cleaning, integration, prediction, and analysis, toward a full-fledged AI-persona system. Similarly, Yang et al. (2024) propose enhancing LLMs with knowledge graphs for improved factual reasoning. These advancements could generate personas that are both dynamic and factually accurate, overcoming the breadth–depth trade-off in current methods.

The broader application of LLMs in persona creation has the potential to enhance UCD by supporting more anthropomorphic information systems (Deshpande et al. 2023). This aligns with the vision of personas evolving from static profiles to dynamic entities engaging users in meaningful interactions (Jansen, Jung, and Salminen 2020). Additionally, AI-generated personas can support reproducible persona science (Salminen, Jung, and Jansen 2022c) by enabling the sharing of prompts, datasets, and code for community-wide validation, promoting scientific rigour in persona research.

In conclusion, AI integration in persona creation signifies a shift towards more interactive and scientifically rigorous practices. This holds promise for enhancing UCD research and practice, but only if the AI-generated personas are received well by users. To this end, usability and user perceptions related to AI-generated personas remain major research topics to address.

3. Method

3.1. Persona generation using Survey-to-Persona

The *Survey-to-Persona* (S2P) is a machine learning (ML) and AI methodology and a system for automatic persona generation from survey-based research data (Salminen, Jansen, and Jung 2022b). S2P differs from earlier automatic persona generation systems (An et al. 2018a, 2018b; Jansen, Jung, and Salminen 2020; Jung et al. 2018) in its compatibility specifically with survey data, being compatible with nearly all types of survey questions, including the Likert scale, Boolean, multiple-choice, and open-ended data (previous systems tend to employ online and social media analytics data for persona generation). The latest version of S2P, with a validation study currently in review, uses sophisticated NLP technologies, including LLMs and retrieval-augmented generation (RAG), to simulate realistic conversations with persona profiles that embody the collective characteristics and viewpoints of distinct user segments. The clustering of the data for persona generation itself is based on ML algorithms similar to those typically deployed in quantitative persona creation research (Brickey, Walczak, and Burgess 2012; Dupree et al. 2016; Tu et al. 2010). For persona generation, the end-user selects the survey question(s), which serve as the basis for the creation process, with no other input required. A more detailed, technical description of the S2P methodology is provided in the

supplementary online material. We omit details from here due to focus on the user study. The system is available online at [*link hidden for anonymous peer review*].

3.2. Experimental set-up

The S2P system was used to create two personas – Linda and Mark – both of which had chat and profile versions. The two personas were created for the usability study, but the system can generate any number of personas based on the underlying datasets. Screenshots of each generated persona are available in the online supplementary material, with an example provided in Figure 2. The personas were based on *Pew Research* survey research measuring US-based respondents' attitudes about AI, comprising around 10,000 completed surveys.[1] The two persona (Linda and Mark) modalities (chat and profile) were identical, being created from the dataset. Both persona modalities are created at the same time by S2P. Both personas had access to the same data. Limiting the number of personas was necessary for the practicality of the experiment. The experimental study followed the within-subject design, in which each participant interacted with both personas and both persona types (chat and profile), thus adhering to a 2 × 2 design. The order of the persona and type was randomly allocated by creating experimental flows to which the participants were randomly assigned. This mitigates any ordering effects. The profile persona could be interacted with using scrolling, searching, and clicking. The chat persona could be interacted with by typing and then reading the persona's answers.

3.3. Study procedure

Upon entering the user study room, participants were welcomed and provided with an explanation of personas, ensuring a common understanding of this key concept (regardless of their prior knowledge). This was followed by a review and signing of the informed consent form. To maintain consistency across sessions, the context of the study, including monitor setup, seating arrangements, and any provided guidance, remained constant. Next, participants were introduced to a fictitious scenario (referred to as the 'task') within an imaginary enterprise focused on AI solutions. Here, they assumed the roles of either a software engineer or a manager (most participants had experience with software engineering, though not all, which is why the role of the manager was also included to help participants identify with the scenario). Each participant then interacted with two unique personas, derived from the Pew Research dataset. Their task was to answer seven critical questions about these personas, as instructed by their (fictional) superiors. Essentially, participants used one persona (either chat or profile persona) to locate specific user information from the persona. This task involved a combination of

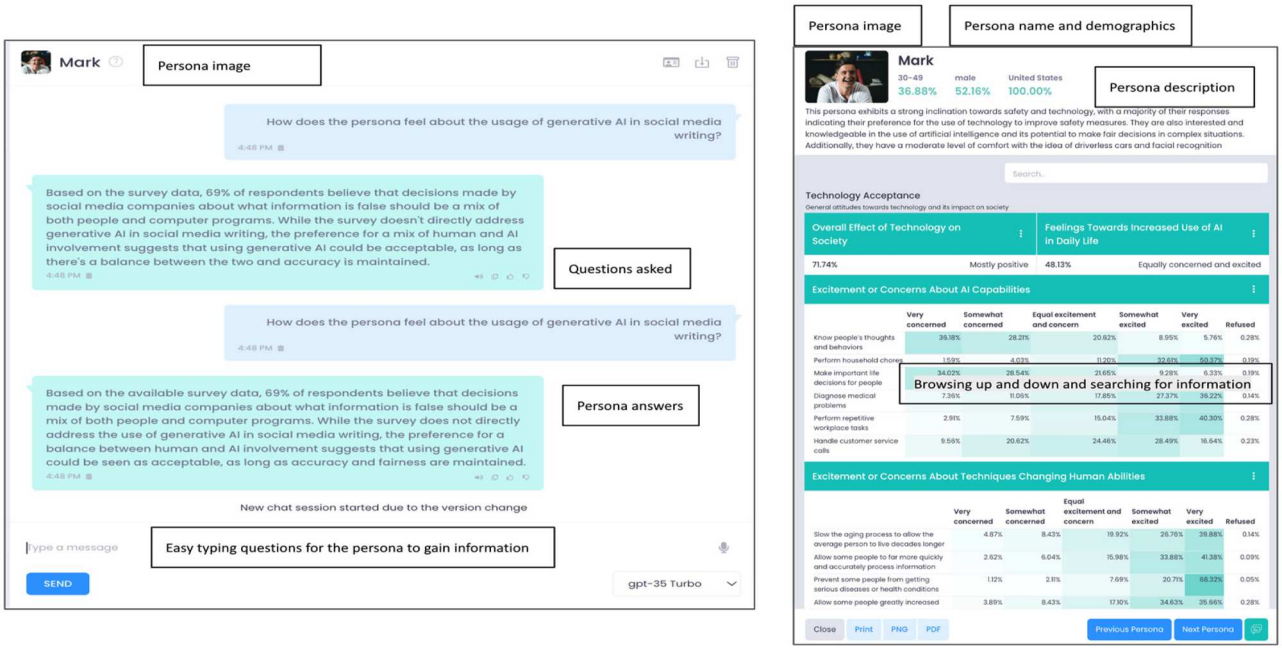


Figure 2. An example of an AI-generated persona, 'Mark', in two different formats: chat (left) and profile (right). Both have name and image. The interaction with chat version takes place by typing while the profile version can be interacted with by perusing the information and using the search function.

multiple-choice and open-ended questions designed to assess their understanding of the persona (details and answer keys are provided in the supplementary material). Participants could continuously view these questions on Monitor 1 while interacting with the persona displayed on Monitor 2. The participants were encouraged to think-aloud as they were using the system and completing, and these think-aloud records were later transcribed for manual analysis. After completing activities for both personas, participants were thanked for their time and presented with a gift card as a token of appreciation.

3.4. Participant demographics and recruitment

The study employed a direct interaction approach and was conducted at two locations: a research institute and a college both within the same university. A total of 54 participants were recruited with informed consent obtained prior to their involvement. Measures were taken to ensure participant privacy; no personally identifiable information was collected beyond voice recordings, which were accessible only to the research team. The participant pool comprised primarily computer science researchers (36 individuals, 66.67%) and engineers (9 individuals, 16.66%). The remaining participants (9 individuals, 16.67%) came from diverse professional backgrounds, including a lab coordinator, a director of security and health, a bioinformatics specialist, five graduate students, and one business development manager.

In terms of prior experience with personas, 22 participants (40.7%) were entirely unfamiliar with the concept, 20 participants (37.0%) had some prior knowledge, and 12 participants (22.3%) had used personas before. The average experience with personas was 1.92 years (N = 54, SD = 5.84). Regarding chatbots, two participants (3.7%) had no experience, six participants (11.1%) were familiar with the concept, and 46 participants (85.2%) had used chatbots previously. The average experience with chatbots was 2.09 years (N = 54, SD = 2.50). The average age of participants was 33 years (SD = 10.60), with a slight majority identifying as male (n = 30, 55.6%) and the remainder identifying as female (n = 24, 44.4%). No other gender identifications were reported.

3.5. Study variables and data analysis

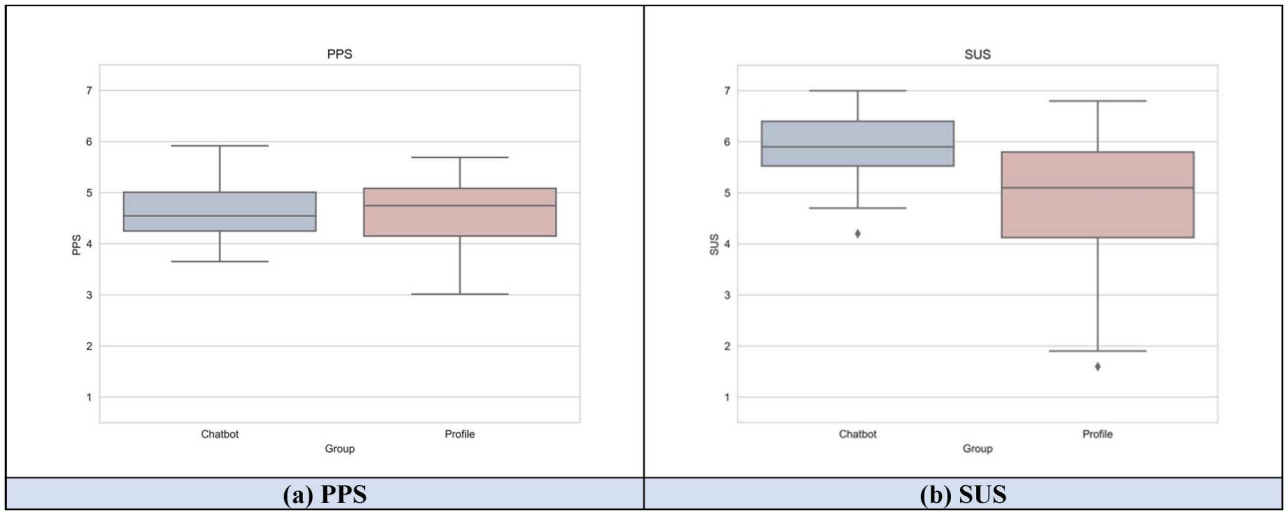
Following task completion, participants answered survey questions regarding their interaction with the persona (see the items in Appendix 1). These questions utilised two validated scales: (1) the *System Usability Scale* (SUS) (Lewis and Sauro 2009) which measures a user's overall perception of a system's usability, and (2) the *Persona Perception Scale* (PPS) (Salminen et al. 2020a) which specifically assesses a user's perception of a persona's attributes, including credibility, completeness, clarity, consistency, empathy, and others. In addition, to evaluate participant performance, we measured dwell time (i.e. how long they used each persona) and task success for each participant. The success rate involved comparing their responses to the seven questions with the personas' true attributes, detailed in the supplementary material. Participant responses to all seven questions were assessed for accuracy (correct or incorrect) and used to calculate a task success rate. For example, if three answers were correct, the task success rate would be 42.3% (3 correct / 7 questions). Finally, we also measured the perceived task difficulty, task enjoyment, and the participants' confidence in the correctness of their task completion, as these are central measures related to user study tasking (Doherty and Doherty 2019). In terms of data analysis, the results for RQ1 are visualised using box plots, accompanied by Mann–Whitney U tests for group comparisons. RQ2 is addressed using correlation analyses, and RQ3 is addressed through a qualitative analysis of transcribed user study sessions.

4. Findings

4.1. RQ1: are there significant differences in user perceptions and interaction when using chat-based and profile personas generated by AI?

Figure 3 compares the perceptions of the CP to the PP on the PPS (Figure 3a) and SUS (Figure 3b) scales. We can observe that the median value for the chatbot group is virtually identical to the median for the profile group. The Mann–Whitney U test indicated that there was no statistically significant difference in PPS between CP and PP, U = 1516.5, $p = 0.722$. However, the interquartile range (IQR), which is represented by the height of the boxes, appears to be higher for the PP. The IQR indicates that the middle 50% of the scores are spread over a higher range for the PP. Similarly, the range, indicated by the vertical lines (whiskers) from the boxes, is larger for the PP than for the CP. This suggests there is more variability in how users perceive profile personas, especially on the lower side of scores (whisker going toward zero). On the other hand, the CP has more outliers (indicated by the diamond shapes). This indicates that some users have a perception of the CP that is significantly different from the majority, either much higher or much lower.

Thus, while the CP is generally perceived more favourably than the PP (higher median), the PP has a



(a) PPS (b) SUS

Figure 3. Aggregate scores for chat and profile personas. **(a)** is the Persona Perception Scale and **(b)** is the System Usability Scale.

wider range of perceptions (greater variability around the median). This could imply that experiences with the PP are more polarising among users, with specific users finding them particularly favourable or unfavourable, while perceptions of the CP are more consistent. Whether variability in this regard is 'good' or 'bad' is not straightforward, because the persona design theory does not explicitly mention if the same personas should be perceived similarly by all users. Nonetheless, we surmise that the higher variability of perceptions among the PP stems from its richer information that allows for a higher degree of subjective assessment relative to the more clinical, factual, and clean layout provided by the CP. The same seems to repeat with usability (Figure 3b): the assessments for the PP are more varied and for the CP more centred.

There is a significant correlation between the SUS aggregate and the PPS aggregate scores, $r = .559$, $p < .001$. Recall that a high score in the PPS represents a favourable perception of the persona, whereas a high SUS score represents a favourable perception of the persona's usability. So, the results indicate that the usability of the persona is associated with the perception of its quality as a representation of end-user groups. The Mann–Whitney U test indicated that there was a statistically significant difference in SUS between CP and PP, $U = 1785.0$, $p = 0.0435$.

Another interesting finding is that there seems to be a usability paradox at play, by which we mean high useability and high effectiveness (or low usability and low effectiveness) do not necessarily correlate. On one hand, users find the CP to have better usability (see Figure 4a), $U = 2285.0$, $p < .001$. However, they tend to perform more poorly on the task when using the CP (Figure 4b), $U = 980.0$, $p = .0026$. Factoring in that the PP was found more difficult (Figure 4c), $U = 438.5$, $p < .001$ and users had longer dwell time with it (Figure 4d), $U = 1064.0$, $p = .0156$, we can surmise that users struggle with the profile persona more than with the chat persona. However, this 'hard labor' seems to bear fruit in terms of a higher success rate in actually using the persona. Therefore, the seemingly negative usability metrics could lead to a counterintuitive effect in which struggling more with the persona results in better recall and understanding of the persona details for the task at hand. The usability paradox deserves further consideration, but the paradox might be a result of usability being an umbrella term that includes, among other things, effectiveness, efficiency, and satisfaction. These components might be in tension with each other. Again, this usability paradox deserves further research.

Moreover, while the users enjoyed using the profile persona less than the easy chat interface (Figure 4e), $U = 2176.5$, $p < .001$, the profile persona may 'get the job done' more effectively when it comes to learning about the persona. Interestingly, there is a slight tendency, not statistically significant, for users to be more confident in their answers after using the CP (Figure 4f), $U = 1624.5$, $p = 0.2672$, even though these answers are more likely to be inaccurate. In other words, the chat persona's apparent easiness of use may lead to overconfidence among the users. In some cases, this can be problematic because the users would assume they have the right information about the people the personas represent while in fact they are mistaken. So, this overconfidence can pose a serious risk in the use of chat personas.

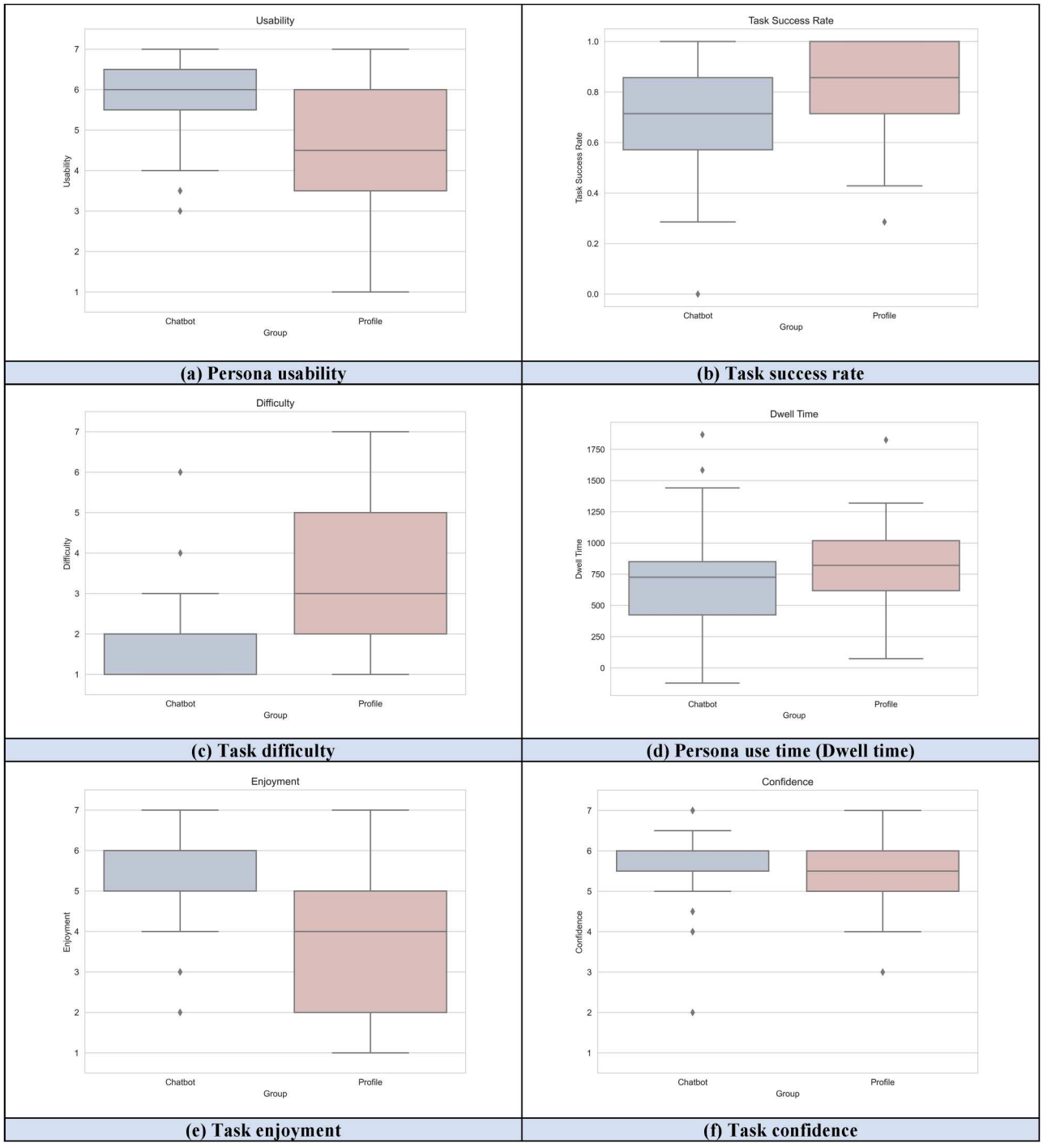


Figure 4. Usability paradox: people find the CP's usability better (a) but complete the task more poorly using it (b). As difficulty (c) and dwell time (d) are higher for the profile persona, this could indicate that users who find the profile persona harder to use need to use it more; therefore, they learn more about the persona in the process ('hard labor bears fruit' effect). The participants do not enjoy using the profile person as much as the chat (e), but it 'gets the job done' better. Interestingly, people have a slight tendency to be more confident in their answers given after using the CP (f), even though these answers are more likely to be wrong.

4.2. RQ2: how is users' dwell time associated with their perception of AI-generated Personas and their usability?

To address RQ2, we calculated Pearson's correlation coefficient between dwell time and each PPS variable and SUS aggregate variable. The results indicate a weak positive correlation between similarity (r = 0.333, p = .014) and system use time. The results also indicate a weak positive correlation between *WTU* (r = 323, p = .017) and dwell time. The results are summarised in

Table 1. The table presents Pearson's correlation coefficients (r) between dwell time and each PPS construct, as well as the SUS aggregate score. These coefficients indicate the strength and direction of the relationship between the time spent using a system and various user perceptions of persona characteristics and overall system usability.

Clarity shows a significant negative correlation with the CP dwell time (r = −0.347, p = .010), suggesting that longer user interaction with the chat persona is associated with decreased perception of its clarity. In contrast, there is a slight positive correlation with the PP that is not statistically significant and a negligible overall correlation.

Compassion has a positive and significant correlation with the CP dwell time (r = 0.277, p = .043), indicating that longer use is associated with higher compassion ratings for the chat persona. There is no significant correlation for the PP or overall.

Similarity exhibits a significant positive correlation with the PP dwell time (r = 0.333, p = .014), meaning that the longer users interact with the profile persona, the more they perceive it to be similar to themselves (or vice versa). The correlation is not significant for the CP or overall, but interestingly the directionality of the correlation between the chat persona's dwell time and similarity is negative.

Usability demonstrates a significant negative correlation with the CP dwell time (r = −0.316, p = .020), suggesting that longer interaction times with the chat persona are associated with perceptions of decreased usability. There is no significant correlation for the PP or overall.

Willingness to use shows a significant positive correlation with the PP dwell time (r = 0.323, p = .017), indicating that the more users interact with the profile persona, the more willing they are to use it going forward. There was no significant correlation for the CP or overall.

Completeness, *Consistency*, *Credibility*, *Empathy*, *Sterotypicality*, and *Transparency* show no significant correlations with system use time, suggesting that the duration of use does not significantly affect users' perceptions of these persona attributes. The SUS aggregate score indicated a significant positive correlation with the PP dwell time (r = 0.277, p = .043), that longer use is associated with better usability scores for profile personas.

Table 1. Pearson's correlation coefficient (r) between dwell time and each PPS construct and the SUS aggregate score. Significance levels indicate the likelihood that the observed correlation is not due to chance. In this context, significant negative correlations suggest that as dwell time increases, the perception of certain constructs (like clarity for the control persona) decreases. Conversely, significant positive correlations (e.g. similarity for the experimental persona) suggest that longer interactions lead to more favourable perceptions of those constructs. ▲ = indicates positive correlation, ▼ = indicates negative correlation.

Construct	r (CP)	r (PP)	r (total)
Clarity	▼ **−0.347****	▲ 0.262	▼ −0.074
Compassion	▲ **0.277***	▲ 0.091	▲ **0.189***
Completeness	▼ −0.109	▲ 0.129	▲ 0.024
Consistency	▼ −0.149	▲ 0.096	▼ −0.057
Credibility	▼ −0.136	▲ 0.115	▲ 0.021
Empathy	▲ 0.096	0	▲ 0.072
Similarity	▼ −0.193	▲ **0.333***	▲ 0.064
Stereotypicality	▲ 0.014	▼ −0.08	▼ −0.058
Transparency	▼ −0.018	▲ 0.059	▲ 0.041
Usability	▼ **−0.316***	▲ 0.128	▼ −0.146
WTU	▼ −0.058	▲ **0.323***	▲ 0.102
SUS	▲ −0.188	▲ **0.277***	▲ 0.092

Notes: *significant at a = .05; **significant at a = .01; ***significant at a = .001.

4.3. RQ3: what are the central usability issues in AI-generated Personas?

RQ3 was addressed by analyzing session recording transcripts. The transcripts were analyzed by session type (CP or PP) by the researcher with the assistance of ChatGPT. Themes recognised by ChatGPT were verified by the researcher via manual reading of the session transcripts after which participant quotes were assigned for each theme.

4.3.1. Chat Persona usability issues

Difficulty Engaging with the Persona's Story (the Fourth Wall Effect): Some users expected responses to reflect a more personal tone, feeling a disconnect when responses referred to the persona in the third person. Instead, the chat persona sometimes lost the personal touch, replying as the 'persona group' rather than the individual persona. This is akin to the famous 'fourth wall' effect in cinema (Brown 2013) and here it is a side-effect of the prompting since the prompt does not ask the LLM to generate responses in first-person. Consequently, these users found it challenging to piece together the persona's narrative.

- *'I don't have any information. I feel very disengaged.' (P18, Cf)*
- *'I wasn't actually engaged or communicating with the persona.' (P04, Cf)*
- *'Here, it is talking about many things. Makes me feel distracted.' (P47, Cf)*
- *'Persona communicated a coherent story. Strongly disagree, it just provided data.' (P36, Cf)*

- *'I mean, you can feel that there is some personality behind it. But if you push the limits a little bit, there's no personality. There's no soul. (P14, Cf)*

Transparency: Because the chat persona relies on a 'hidden UI' (also called 'invisible UI' by some (Varaprasad and Mahalaxmi 2022)) that does not show the persona's information (like the profile interface does), users sometimes had trouble linking the persona presentation to the original survey data it was based on.

- *'Creation process [of the persona] appears transparent because I could think that ok, it's [the chat] like an interpreter between me and the persona. So, like it felt that ok, I can understand what is going on here. It was some kind of an agent.' (P04, Cf)*
- *'I felt again there was like this wall between us.' (P04, Cf)*
- *'But what things are happening behind the scenes, I don't know.' (P05, Cm)*
- *'I was more interested in getting a clear answer from it, whereas it was just trying to summarize things.' (P05, Cm)*

Inadequate Directness and Clarity: The chat sometimes failed to directly address questions, particularly unanswerable questions, leaving users confused. This might be due to a lack of understanding of the user's question. Users expected the chat to directly acknowledge if a question was not covered in the survey data. However, LLMs have a well-known property of not being able to tell when they are not able to answer a question in a factually correct way (Borji and Mohammadian 2023). The fact chat struggled with unanswerable questions not based on real data points to ChatGPT's desire to 'please' the user by giving answers to any questions, even when this forces it to make up things or 'hallucinate' (McIntosh et al. 2023). In this sense, the 'unanswerable question test' that we implemented is akin to the 'needle in the haystack test' used to test LLMs (Zhang et al. 2023a, 2023b).

- *'Oh God. Again, a long answer. I can tell you that it's always a little long. Why doesn't the short answer be the default and then the long one?' (P20, Cf)*
- *'So, yet I did not get a clear answer, I'm trying to dig it further regarding the use of generative and social media writing grounded on knowledge but there is no specific [survey] question that directly address this one.' (P34, Cf)*
- *'So, basically for the how does it feel about the usage of generative AI in social media writing, it has no specific information so I'm not sure.' (P54, Cf)*
- *'Oh, it's a long answer.' (P36, Cf)*

Overly Elaborate Responses: Some answers were too detailed, causing information overload or diverging from the specificity of the questions. This relates to ChatGPT's verbosity, meaning that they tend to generate longer answers rather than shorter ones (Zheng et al. 2024), possibly due to a strive for completeness. So, at times, the chat gave overly long and elaborate answers to specific questions (e.g. when asked 'What does the persona think about the use of AI to decide which patients should get medical treatment' (P54, Cf), the chat persona responded 'The persona's response to the survey question 'Would you favour or oppose the use of computer programs to make final decisions about which patients should get medical treatment?' indicates that they oppose the idea, with 64% opposing and only 18% in favour. However, their response to the survey question 'Do you think it is possible or not possible for people to design artificial intelligence computer programs that can consistently make fair decisions in complex situations?' suggests that they believe it is possible to design AI programs that can make fair decisions, which could potentially increase their comfort with the idea of using AI in medical treatment decisions.').

- *'I think it's good that it provides answers, although there is some extra information. In terms of identifying what a persona is, I don't think it's a good solution. Sometimes it provided information that I did not ask for.' (P36, Cf)*
- *'Yeah, like what I was reading, or I wanted to skip just most of it and just get to the point. I guess some people like details.' (P53, Cm)*
- *'The response is way too long and complicated.' (P02, Cf)*
- *'It was too long to absorb it all.' (P08, Cf)*

Confusing Interface Elements: Users confused the download button with the SEND button, suggesting a need for a clearer interface design (see Figure 5a).

- *'When I try to SEND, I keep navigating to as if it was 'download'.' (P53, Cm)*
- *'It's still not SEND. Yeah, it's a good point. I automatically went there for that [SEND]' (P56, Cf)*

Non-Responsive Periods: The chat occasionally stopped responding or gave errors, interrupting the user experience. This relates to server and API-related issues. At times, the chat persona stopped responding or gave errors, impacting the interaction. Based on our review of the logs, this happened in 0.56% (n = 3) of the 534 interactions (user messages to the chat). In other words, the system was highly stable.

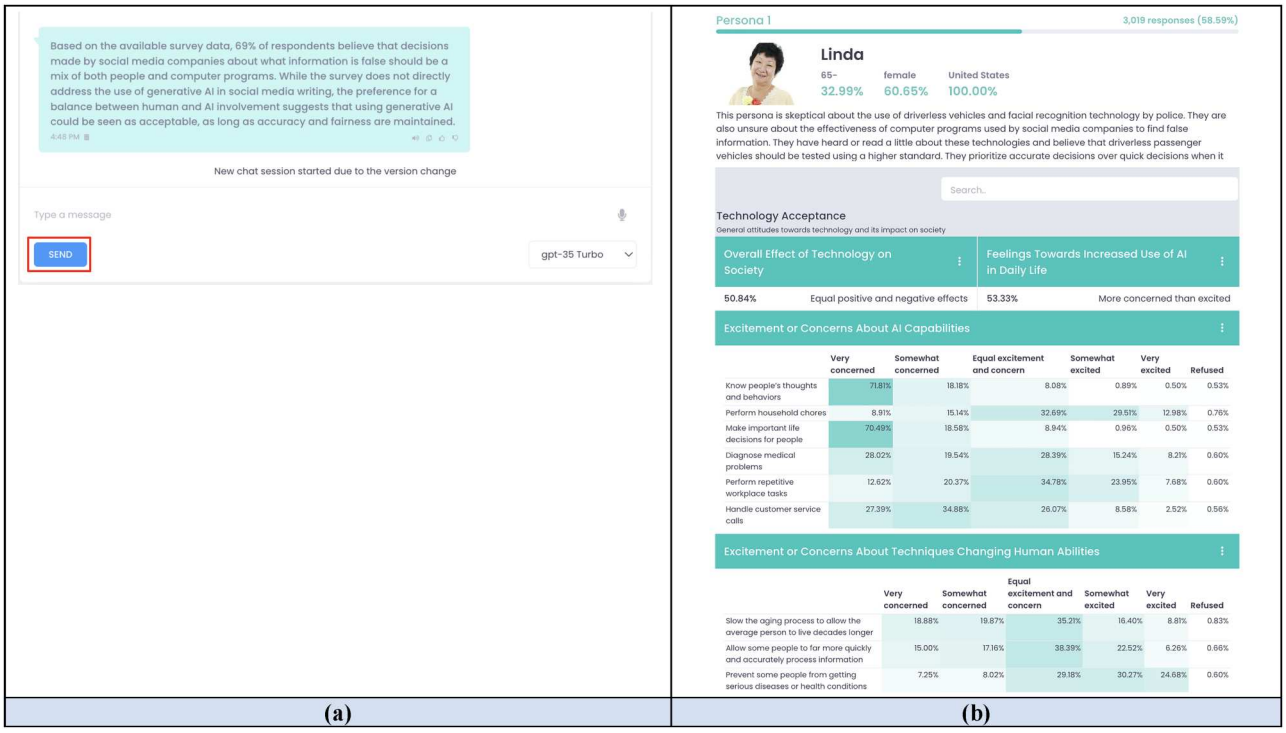


Figure 5. Examples of usability issues in AI-generated (a) chat and (b) profile personas. The former represents a simple UI problem of button placement while the latter represents a more complex problem of the AI not considering the cognitive limitations of human persona users and thus presenting an excessive amount of information.

- *'On page but error has occurred. It's not responding.' (P05, Cm)*
- *'We have an unexpected error occurred press ok.' (P06, Cf)*
- *'There's an error, I'll just go ahead and refresh.' (P48, Cf)*

4.3.2. Profile Persona usability issues

Ineffective Search Functionality: Interestingly, users relied more on browser find (Ctrl + F) than the built-in search. However, the search did not adequately navigate subsections or individual items, hindering efficient information retrieval. This observation denotes the importance of searching the data-driven persona's information thoroughly, including metadata. If the search function does not effectively search within subsections or individual survey items, the user might mistakenly assume that the information they are looking for is not to be found in the persona profile. On a related note, usability issues with web-based profile personas may inherit advantages or benefits from generic web interaction techniques like the use of keyboard shortcuts. This should be kept in mind when designing such personas.

- *'I actually prefer scrolling to using the search function. The search function is making things hard.' (P17, Pf)*
- *'The search function can be improved.' (P18, Pm)*
- *'The search function only searches this top left part so there might be something under.' (P19, Pf)*
- *'I guess you're trying to measure this, but if the search function is not working well, I'm not sure if I did something wrong.' (P23, Pf)*
- *'Can I use control + F?' (P17, Pf)*
- *'I used control + F and then I found the one [answer] I needed and that was fine.' (P26, Pm)*

Information Presentation and Overload: The profile contained too much data, was hard to read due to narrow text formatting requiring scrolling, and lacked a clear hierarchy of information (see Figure 5b). This led to information overload and difficulty in finding specific information. This highlights the special case of a 'high information persona', meaning that the source data contains a lot of variables and information which thus transforms into an information-intensive persona profile. The crucial point here is that information volume directly affects usability: the persona can not only have too little information (Salminen et al., 2020b) but also too much. Information overload and difficulty finding specific details in the profile are related.

- *'It feels overwhelming. I feel like information overload.' (P04, Pm)*

- *'Ok, that's a huge [information] overload.' (P20, Pm)*
- *'I really feel it's too much information given.' (P22, Pm)*
- *'There is too much persona information for my task.' (P34, Pm)*
- *'A filtering system would have helped. I know we had the search options. Or maybe to be able to have the list of questions just instead of too much information being shown to me.' (P40, Pm)*

Difficulty with Data Interpretation: Users found interpreting certain items within the profile challenging, suggesting a need for clearer explanations or guidance. This highlights the need for algorithmic transparency in data-driven persona development (Salminen et al. 2019). It remains a challenging issue to simultaneously show the information and to describe it, explaining variables in a way that is perceived as adequate by users who are interested in technical details while not alienating those that do not necessarily require such information.

- *'There was a lot of information and it was just scrolling instead of maybe organizing the data. (P17, Pf)*
- *'I think it was very detailed. I think that's kind of the advantage over the chat because there was more detail. It was just hard to understand it for it has a learning curve.' (P21, Pf)*
- *'I think even if you give me time, it's hard to put multiple tables in my head and process them. My AI system cannot do that.' (P33, Pf)*
- *'I will not use it because it's hard to obtain information, hard to analyze. Maybe I missed something.' (P48, Pm)*

Visibility and Accessibility: The profile was considered too small, making it difficult to view and find information. This is an extension of the previous issue of information overload; with a lot of information, the user is forced to scroll down the persona profile. This makes it more demanding to view the persona, as one has to go up and down the profile in a repetitive manner. Furthermore, some users observed a lack of clear hierarchy or structure for the information in the profile. More specifically, the information was structured under headings; however, this appeared confusing due to the large number of headings. So, it remains a challenge to organize complex survey information for data-driven persona profiles; it appears a two-layer information hierarchy is not adequate, but three or more layers are needed when the information becomes more complex. Typical implementation on the web is to use pop-ups or dialogues to enable users to expand and collapse information as needed. Moreover, the UI of the data-driven persona needs to be designed such that it supports information-rich dataset representation. For example, narrow columns require excessive scrolling.

- *'The photo was not clear. It was too small.' (P18, Pm)*
- *'I'm fine with scrolling the information, but I don't like that it is hard to read because you have to scroll down and the lines are so low. The rows are quite lengthy and then you have to scroll. See, it's difficult to read this.' (P04, Pm)*
- *'It's so tricky with the features like search, scrolling, and clicking.' (P10, Pm)*
- *'Whenever I need to look up something, I need to scroll up and down and then even within them [tables] there are subsections that I have to read. There is a lot of unnecessary reading.' (P16, Pm)*

4.3.3. Common issues

Overall, there are both differences and intersections in the usability issues concerning AI personas across the interfaces. For example, transparency is an issue, but it manifests in distinct ways: for chat personas, the call for transparency stems from 'hiding' the persona's information (the user only sees the chat dialogue), whereas for the profile persona, there are questions about how the information is generated (the process). It is logical that aspects, such as transparency overlap, because it reflects a general point of improvement in data-driven personas (Salminen et al. 2019). It could be addressed with priming and better explanations are needed about what personas are and how they were created.

Similarly, the unanswerable questions confused users in both presentation types, but the confusion manifested differently. The confusion in the profile stemmed from the lookup process and the conclusion that they could not find the information. In turn, the chat interface did not enable skimming or scanning the persona's information but required specific questions. There was no way for users to corroborate the information given by the chat persona. In a way, the chat interface has more power over the user: if the chat persona says there is no information to answer a given question, the user has to believe this answer, whereas with the profile persona, the user has to make their own conclusion about whether or not such information exists for the persona.

5. Discussion

5.1. Discussion of findings

The present study sought to investigate the usability and user engagement with AI-generated personas, comparing chat and profile formats, and highlighting the fourth

wall effect (Shi et al. 2022; Stahlke, Robb, and Mirza-Babaei 2018), while contextualising these findings within cognitive load theory (Sweller 1988) to explore how different persona formats impact users' cognition and perception of AI-generated personas. These findings are general and transferable to another persona system, as the two approaches used in this research, dialogue, and document, are two major approaches for persona generation. Therefore, the findings highlight several key aspects of user interaction with these personas, offering insights into their utility and limitations within user-centred design. The findings also highlight the possible tension between usability and productivity, as discussed below.

Concerning **RQ1**, no significant end-user preferences were found between chat and profile personas. However, chat personas were generally perceived more favourably in terms of usability, while profile personas exhibited a broader variability in user perceptions. The differences between these two persona formats could be attributed to the interactive nature of chat personas, which may align more closely with users' everyday communication practices, such as messaging apps and social media. The wider variability in user perceptions of profile personas stems from the static nature of profile personas, which enables higher-density information consumption due to being readily visible (whereas information in the chat persona is only visible upon request and only pertaining to the specific request).

Interestingly, we found a *usability paradox* implying how users' perceived usability of a persona type does not always align with effective task completion. More specifically, the chat persona's greater usability creates an ease of use that could hinder deeper processing of the persona information. Conversely, the perceived difficulty of the profile persona may deepen the level of cognitive engagement (Li et al. 2025; Prestridge, Main, and Schmid 2024; Trudeau et al. 2023), aligning with cognitive load theory (Sweller 1988) where overcoming extraneous cognitive load can lead to enhanced schema construction and deeper processing. This may explain the improved persona comprehension despite the higher initial usability hurdle, suggesting that well-designed persona presentations can create more meaningful interactions and long-term usability gains. So, it is not immediately clear that we *should* aim to optimise usability, especially when it comes at the expense of not exposing the user to the persona's full information. Nonetheless, usability and a favourable overall persona perception were significantly correlated, which implies that usability still is a core component in persona UX.

Concerning **RQ2**, dwell time showed a weak positive correlation with the perception of similarity and willingness to use, particularly for profile personas. This suggests that longer interactions may lead to more favourable perceptions of these personas, which is also supported by the significant positive correlation between the SUS aggregate score and the PP dwell time. This supports the notion made in relation to RQ1, meaning that engagement in retrieving the persona's information, even though not immediately appreciated by the user, can drive successful task completions more than 'easy' retrieval of the information via the chat persona. Coincidentally, this finding is consistent with adaptive user behaviour with earlier, non-AI-generated personas, in which users adjusted their cognitive strategies based on the number of personas (Salminen et al. 2022c). A similar type of adaptation seems to be at play, so that even when users initially dislike the AI-generated information-laden personas, they then become accustomed to them over session usage time.

Concerning **RQ3**, there were several usability concerns. For the chat persona, there were four main issues. First, the *'fourth wall' effect* was a notable issue, with users feeling disengaged due to the lack of the persona's personal narrative. In this issue, users experienced a disconnect when the chat persona referred to itself in the third person, which makes it feel less personal and more robotic. Second, the chat personas sometimes *struggled to provide direct answers*, especially to questions not covered by their training data, leading to user confusion. Third, users faced a degree of *information overload from verbose responses*, which reduced the usefulness of the information provided. Fourth, UI elements, specifically the 'Send' button indicate that there is a *need for clear interface design* (and user testing) even when dealing with a seemingly 'simple' interface like a chat persona. Fifth, the users observed *occasional system errors or non-responsiveness*, stemming from LLM and server dependencies and affecting the UX negatively.

For profile personas, first, *information overload* was a significant challenge, implying that AI-generated personas can contain too much information. Second, users found the *search function inadequate*, often resorting to browser find (Ctrl + F), which indicates a need for improved navigation within the persona profile (possibly a deeper information hierarchy with multiple levels, where the first level is approachable and broad). Third, some users found it *challenging to interpret detailed information* within the profile, implying a need for clearer explanations or summaries to aid comprehension. Again, this relates to the previous point that for an LLM, it is extremely easy to interpret large datasets,

but for a human, understanding these interpretations in one view quickly becomes painstaking. Fourth, the profile layout also had some UI issues, including *small text size and extensive scrolling*, which hindered easy access to information.

Interestingly, it appears that 'too much information' in AI-generated personas can be a problem in both tested modalities, but for different reasons. In profile personas, the layout becomes cluttered with numerous information pieces about the persona. Even though web interaction techniques like search can, in theory, help people find a large number of information overwhelming. In chat personas, the length and irrelevance (off-topic) of the information provided can have the same effect: the user feels confused and overwhelmed.

5.2. Design implications

Based on the findings, some guidelines for AI-generated data-driven personas can be proposed for the development of either dialogue or document personas.

For the AI-generated profile personas, a major implication is that a human designer would likely design clear and simple personas by default, but the AI needs specific instructions for that. Namely, a human designer would likely perceive that 'okay, this persona is becoming too complex, I need to drop some information, group it, or otherwise handle the complexity issue'. However, the AI-generated persona creation scales precisely with the dataset size. In this study, we had a multi-dimensional dataset with a lot of variables and respondents, so the outcome personas contained a lot of information. We can call these resulting personas 'high information personas', and such personas can easily be associated with lower usability (because it is more difficult to make sense of the information). In other cases, the opposite might happen: the underlying dataset is very scarce in information, and we would have 'low information personas' that would perhaps be usable but not useful (due to not containing enough information). So, it is important to acknowledge that when making AI-generated personas, the source dataset influences the outcomes a lot.

Actions can be taken to design more flexible layouts for the AI to use in the persona creation process. Human designers can implement better structuring and hierarchy of information to prevent cognitive overload. A human designer can assess whether the profile appears good for information scannability and if the information is logically organised. On a related note, the search functionality in AI-generated personas should effectively cover subsections and item-level data to allow for thorough information retrieval.

For chat personas, more work is needed to design the optimal system prompts for AI-generated personas. Our findings indicate that there is a clear need to adjust the prompting to make interactions (Kocur et al. 2021; Shi et al. 2022) feel more personal (avoiding the fourth wall effect), emphasising the importance of designing interactions that align with user expectations of conversational flow and relational dynamics (Kocur et al. 2021; Shi et al. 2022). Other guidelines include directly acknowledging unanswerable questions and adjusting verbosity to avoid off-topic and increase relevance. The latter can likely be manipulated easily but the acknowledgment of unanswerable questions appears like a more difficult problem. One way to circumvent this is by showing the user actual snippets or an overview of the data alongside the chat persona responses, thereby enabling comparison between the answer and the source material.

Overall, the format of AI-generated personas significantly influences the UX. Hence, designers should carefully choose between chat and profile personas when developing AI-generated personas. For optimal usability, it seems the chat personas are the best choice. For optimal task completion, profile personas seem a better choice.

5.3. Limitations, strengths, and future work

Although a rigorous study, our study is not without limitations. The participant sample was relatively small and predominantly comprised individuals with a background in computer science, which may limit the generalisability of our findings. Future research should aim to include a more diverse participant pool to capture a broader range of user experiences and perceptions. Additionally, our study examined two different persona systems, namely chat personas and persona profiles. The findings are applicable to similar systems, but wider applicability in persona systems would require more research on persona systems' usability and engagement. Addressing these limitations could provide further insights into optimising persona usability and engagement. An interesting study for future research could investigate end users employing both persona modalities under different task conditions.

Future research should explore strategies to mitigate the *fourth wall effect*, perhaps by increasing the AI-generated chat personas' ability to simulate first-person narratives more effectively. In turn, the design of AI-generated profile personas could benefit from employing data visualisation and information architecture principles to enhance user comprehension and prevent cognitive overload, because, while providing

comprehensive information is critical, it is equally important to present it in a digestible format. Future work calls for naturalistic studies that could provide validation of the findings in real-world settings.

This study contributes to the evolving discourse on the role of AI in UCD, particularly regarding the development of AI-generated personas. By highlighting the strengths and weaknesses of chat and profile personas, our research offers valuable directions for enhancing the usability and user engagement of AI-generated personas. As AI continues to advance, so too will its applications in design, emphasising the need for ongoing research on this topic.

6. Conclusion

AI-generated personas offer a promising avenue for UCD but have been understudied in terms of usability. Focusing on two common formats – conversational chat personas and static profile personas – the key findings indicate chat personas tend to be perceived more favourably, and profile personas exhibit greater variability in user perception. Different interaction modalities with AI-generated personas, the CP and PP, demonstrate distinct usability issues that influence user engagement and comprehension of the persona. Our results suggest that chat personas are generally perceived more favourably in terms of usability. The findings revealed a usability paradox where the CP, despite being perceived as easier to use, did not necessarily enhance task performance or understanding of the personas compared to the PP. This suggests that ease of use in the CP might detract the user from deeper cognitive processing necessary for effective learning and engagement with the persona information.

The study also shows the importance of interaction design in AI-generated personas. Users experienced different kinds of challenges with each type of persona interface. For instance, chat lacked personality and directness, which could lead to disengagement and confusion. On the other hand, the profile, while sometimes overwhelming due to the sheer amount of content and information, allowed users to engage more deeply with the persona, potentially leading to better retention and understanding, albeit at the cost of ease of use. The results indicate that the format of AI-generated personas can significantly impact UX in persona systems. Future developments could involve improving the first-person communication of chat personas, the information hierarchy and search functionality in the profile personas and providing support to help users navigate and understand complex ('high information') persona data.

While AI-generated personas hold great potential for enriching user interaction by embodying persona information, their effectiveness relies on a design that considers the inherent trade-offs between ease of use and the depth of information presentation. For future AI-generated persona usability research, our study offers a starting point for establishing chat persona and persona profile interaction differences.

Note

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

Acknowledgements

Open Access funding provided by the Qatar National Library.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Liikesivistysrahasto.

References

Adlin, T., and J. Pruitt. 2010. *The Essential Persona Lifecycle: Your Guide to Building and Using Personas* (1st ed.). San Francisco, CA: Morgan Kaufmann Publishers.

Alessa, A., and H. Al-Khalifa. 2023. "Towards Designing a ChatGPT Conversational Companion for Elderly People." In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, 667–674. https://doi.org/10.1145/3594806.3596572.

Amin, M. M., E. Cambria, and B. W. Schuller. 2023. "Will Affective Computing Emerge from Foundation Models and General Artificial Intelligence? A First Evaluation of ChatGPT." *IEEE Intelligent Systems* 38 (2): 15–23.

An, J., H. Kwak, and B. J. Jansen. 2016. "Towards Automatic Persona Generation Using Social Media." In *Proceedings of Third International Symposium on Social Networks Analysis, Management and Security (SNAMS 2016), The 4th International Conference on Future Internet of Things and Cloud. Third International Symposium on Social Networks Analysis, Management and Security (SNAMS 2016), The 4th International Conference on Future Internet of Things and Cloud*. 22–24 August, Vienna, Austria.

An, J., H. Kwak, S. Jung, J. Salminen, and B. J. Jansen. 2018a. "Customer Segmentation Using Online Platforms: Isolating Behavioral and Demographic Segments for Persona Creation via Aggregated User Data." *Social Network Analysis and Mining* 8 (1), https://doi.org/10.1007/s13278-018-0531-0.

An, J., H. Kwak, J. Salminen, S. Jung, and B. J. Jansen. 2018b. "Imaginary People Representing Real Numbers:

Generating Personas from Online Social Media Data." *ACM Transactions on the Web (TWEB)* 12 (3): 1–26.

Borji, A., and M. Mohammadian. 2023. *Battle of the Wordsmiths: Comparing ChatGPT, GPT-4, Claude, and Bard*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4476855.

Brickey, J., S. Walczak, and T. Burgess. 2012. "Comparing Semi-Automated Clustering Methods for Persona Development." *IEEE Transactions on Software Engineering* 38 (3): 537–546.

Brown, T. 2013. *Breaking the Fourth Wall*. Edinburg: Edinburgh University Press.

Chatterjee, A., U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal. 2019. "Understanding Emotions in Text Using Deep Learning and Big Data." *Computers in Human Behavior* 93:309–317. https://doi.org/10.1016/j.chb.2018.12.029.

Cheng, M., E. Durmus, and D. Jurafsky. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. *arXiv Preprint arXiv:2305.18189*.

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

De Angelis, L., F. Baglivo, G. Arzilli, G. P. Privitera, P. Ferragina, A. E. Tozzi, and C. Rizzo. 2023. "ChatGPT and the Rise of Large Language Models: The new AI-Driven Infodemic Threat in Public Health." *Frontiers in Public Health* 11), https://doi.org/10.3389/fpubh.2023.1166120.

De Paoli, S. 2023. Writing user personas with Large Language Models: Testing phase 6 of a Thematic Analysis of Semi-Structured Interviews. *arXiv Preprint arXiv:2305.18099*.

Deshpande, A., T. Rajpurohit, K. Narasimhan, and A. Kalyan. 2023. Anthropomorphization of AI: Opportunities and Risks. *arXiv Preprint arXiv:2305.14784*.

Doherty, K., and G. Doherty. 2019. "Engagement in HCI: Conception, Theory and Measurement." *ACM Computing Surveys* 51 (5): 1–39. https://doi.org/10.1145/3234149.

Dupree, J. L., R. Devries, D. M. Berry, and E. 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*, 5228–5239.

Grudin, J., and J. Pruitt. 2002. "Personas, Participatory Design and Product Development: An Infrastructure for Engagement." In *Proceedings of Participation and Design Conference (PDC2002)*, 8.

Holzinger, A., M. Kargl, B. Kipperer, P. Regitnig, M. Plass, and H. Müller. 2022. "Personas for Artificial Intelligence (AI) an Open Source Toolbox." *IEEE Access* 10:23732–23747. https://doi.org/10.1109/ACCESS.2022.3154776.

Jansen, B. J., S.-G. Jung, L. Nielsen, K. W. Guan, and J. Salminen. 2022. "How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment." *Pacific Asia Journal of the Association for Information Systems* 14 (3), https://doi.org/10.17705/1pais.14301.

Jansen, B. J., S. Jung, and J. 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), https://doi.org/10.1002/pra2.215.

Jiang, H., X. Zhang, X. Cao, J. Kabbara, and D. Roy. 2023. Personallm: Investigating the Ability of Gpt-3.5 to Express Personality Traits and Gender Differences. *arXiv Preprint arXiv:2305.02547*.

Jung, S.-G., J. Salminen, H. Kwak, J. An, and B. J. Jansen. 2018. "*Automatic Persona Generation (APG): A Rationale and Demonstration*." In Proceedings of the 2018 Conference on Human Information Interaction&Retrieval, 321–324.

Kaate, I., J. Salminen, S.-G. Jung, J. M. Santos, E. Häyhänen, T. Xuan, J. Azem, and B. 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*, 164–169. https://doi.org/10.1145/3631700.3664882.

Kim, J., and I. Im. 2023. "Anthropomorphic Response: Understanding Interactions Between Humans and Artificial Intelligence Agents." *Computers in Human Behavior* 139:107512. https://doi.org/10.1016/j.chb.2022.107512.

Koch, T. K., P. Romero, and C. Stachl. 2022. "Age and Gender in Language, Emoji, and Emoticon Usage in Instant Messages." *Computers in Human Behavior* 126:106990. https://doi.org/10.1016/j.chb.2021.106990.

Kocur, M., P. Lindemann, T. Pfeil, and M. Lankes. 2021. "The Door: Timing Effects of Fourth Wall Breaks on Immersion and Game Experience in an Adventure Game." *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*, 43–48. https://doi.org/10.1145/3450337.3483489.

Lewis, J. R., and J. 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*, 94–103. https://doi.org/10.1007/978-3-642-02806-9_12.

Li, Q., Y. Li, S. Zhang, X. Zhou, and Z. Pan. 2025. "A Theoretical Framework for Human-Centered Intelligent Information Services: A Systematic Review." *Information Processing & Management* 62 (1): 103891. https://doi.org/10.1016/j.ipm.2024.103891.

McGinn, J. J., and N. Kotamraju. 2008. "Data-Driven Persona Development." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1521–1524.

McIntosh, T. R., T. Liu, T. Susnjak, P. Watters, A. Ng, and M. N. Halgamuge. 2023. "A Culturally Sensitive Test to Evaluate Nuanced GPT Hallucination." *IEEE Transactions on Artificial Intelligence* 1 (01): 1–13. https://doi.org/10.1109/TAI.2023.3332837.

Mulder, S., and Z. Yaar. 2006. *The User is Always Right: A Practical Guide to Creating and Using Personas for the Web*. Berkeley: New Riders.

Nielsen, L. 2019. *Personas—User Focused Design* (2nd ed. 2019 edition. London: Springer.

Nielsen, L., K. S. Hansen, J. Stage, and J. 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): 45–61. https://doi.org/10.4018/ijskd.2015010104.

Prestridge, S., K. Main, and M. Schmid. 2024. "Identifying how Classroom Teachers Develop Presence Online:

Breaking the Fourth Wall in Online Learning." *Education and Information Technologies* 29 (2): 1357–1377. https://doi.org/10.1007/s10639-023-11714-8.

Salminen, J., K. Guan, L. Nielsen, S. Jung, S. A. Chowdhury, and B. J. Jansen. 2020a. "A Template for Data-Driven Personas: Analyzing 31 Quantitatively Oriented Persona Profiles." In *The Proceedings of the 22nd International Conference on Human-Computer Interaction (HCII'20)*.

Salminen, J., B. Jansen, and S.-G. Jung. 2022b. "Survey2Persona: Rendering Survey Responses as Personas." In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 67–73. https://doi.org/10.1145/3511047.3536403.

Salminen, J., S. Jung, H. Almerekhi, E. Cambria, and B. Jansen. 2023. "How Can Natural Language Processing and Generative AI Address Grand Challenges of Quantitative User Personas?" In *HCI International 2023 – Late Breaking Papers*, edited by H. Degen, S. Ntoa, and A. Moallem, 211–231. Copenhagen: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-48057-7_14.

Salminen, J., S. Jung, and B. J. Jansen. 2021. "Are Data-Driven Personas Considered Harmful?: Diversifying User Understandings with More Than Algorithms." *Persona Studies* 7 (1): 48–63.

Salminen, J., S.-G. Jung, and B. Jansen. 2022c. "Developing Persona Analytics Towards Persona Science." In *27th International Conference on Intelligent User Interfaces*, 323–344. https://doi.org/10.1145/3490099.3511144.

Salminen, J., S. Jung, L. Nielsen, S. Şengün, and B. J. Jansen. 2022c. "How Does Varying the Number of Personas Affect User Perceptions and Behavior? Challenging the 'Small Personas' Hypothesis!." *International Journal of Human-Computer Studies* 168:102915. https://doi.org/10.1016/j.ijhcs.2022.102915.

Salminen, J., J. M. Santos, S. Jung, M. Eslami, and B. J. Jansen. 2019. "Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas." *International Journal of Human–Computer Interaction* 0 (0): 1–13. https://doi.org/10.1080/10447318.2019.1688946.

Salminen, J., J. M. Santos, H. Kwak, J. An, S. Jung, and B. J. Jansen. 2020d. "Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas." *International Journal of Human-Computer Studies* 141:102437. https://doi.org/10.1016/j.ijhcs.2020.102437.

Shi, Y., T. Gao, X. Jiao, and N. Cao. 2022. "Breaking the Fourth Wall of Data Stories Through Interaction." *IEEE Transactions on Visualization and Computer Graphics* 29 (1): 972–982.

Stahlke, S., J. Robb, and P. Mirza-Babaei. 2018. "The Fall of the Fourth Wall: Designing and Evaluating Interactive Spectator Experiences." *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)* 10 (1): 42–62.

Sweller, J. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257–285.

Trudeau, A., Y. Xie, O. Ketsman, and F. Demir. 2023. "Breaking the Fourth Wall": The Effects of Cinematic Virtual Reality Film-Viewing on Adolescent Students' Empathic Responses." *Computers & Education: X Reality* 2:100009.

Tu, N., 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*, 1–5.

Varaprasad, R., and G. Mahalaxmi. 2022. "Applications and Techniques of Natural Language Processing: An Overview." *IUP Journal of Computer Sciences* 16 (3), https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2583441X&AN=159387678&h=icI2HkoiwiPGsucfFR5VmwW1rk5Nvr8rdhsfuMQ0EyVCZftDixJec4rQRiOSVJnK%2FTeMW0IrCqpjS4vjYYLLOA%3D%3D&crl=c.

Xu, R., Y. Sun, M. Ren, S. Guo, R. Pan, H. Lin, L. Sun, and X. Han. 2024. "AI for Social Science and Social Science of AI: A Survey." *Information Processing & Management* 61 (3): 103665. https://doi.org/10.1016/j.ipm.2024.103665.

Yang, L., H. Chen, Z. Li, X. Ding, and X. Wu. 2024. "Give Us the Facts: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling (arXiv:2306.11489)." *arXiv*. https://doi.org/10.48550/arXiv.2306.11489.

Zhang, X., H.-F. Brown, and A. 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*, 5350–5359.

Zhang, X., L. Liu, Y. Wang, X. Liu, H. Wang, A. Ren, and C. Arora. 2023a. "PersonaGen: A Tool for Generating Personas from User Feedback." *arXiv Preprint arXiv:2307.00390*.

Zhang, C., C. Zhang, C. Li, Y. Qiao, S. Zheng, S. K. Dam, M. Zhang, et al. 2023b. "One Small Step for Generative AI, One Giant Leap for AGI: A Complete Survey on ChatGPT in AIGC Era (arXiv:2304.06488)." *arXiv*. http://arxiv.org/abs/2304.06488.

Zheng, L., W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, and E. Xing. 2024. "Judging llm-as-a-Judge with mt-Bench and Chatbot Arena." *Advances in Neural Information Processing Systems* 36. https://proceedings.neurips.cc/paper_files/paper/2023/hash/91f18a1287b398d378ef22505bf41832-Abstract-Datasets_and_Benchmarks.html.

Zhou, M. X., W. Chen, Z. Xiao, H. Yang, T. Chi, and R. Williams. 2019. "Getting virtually personal: Chatbots who actively listen to you and infer your personality." In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, 123–124.

Zhu, H., H. Wang, and J. 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*, 61–70. https://doi.org/10.1145/3322276.3322285.

Appendix 1. Survey items.

Scale	Item	Variable	Scale
PPS (Salminen et al. 2020d)	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 artefacts; 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 (Lewis and Sauro 2009)	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