"You Always Get an Answer": Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination

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

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

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

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

Essi Häyhänen
University of Vaasa
Vaasa, Finland
essihayh@uwasa.fi

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

Bernard J. Jansen
Qatar Computing Research Institute
HBKU
Doha, Qatar
bjansen@hbku.edu.qa

Abstract

We investigated the presence and acceptance of hallucinations (i.e., accidental misinformation) of an *AI-generated persona* system that leverages large language models for persona creation from survey data in a 54-user within-subjects experiment. After interacting with the personas, users were given a task to ask the personas a series of questions, including an unanswerable question, meaning the personas lacked the data to answer the question. The AI-generated persona system provided a plausible but incorrect answer half (52%) of the time, and more than half of the time (57%), the users accepted the incorrect answer, and the rest of the time, users answered the unanswerable question correctly (no answer). We found that when the AI-generated persona hallucinated, the user was significantly more likely to answer the unanswerable question incorrectly. Also, for genders separately, when the AI-generated persona hallucinated, it was significantly more likely for the female user and the male users to answer the unanswerable question incorrectly. We identified four themes in the AI-generated persona's answers and found that users perceive AI-generated persona's answers as long and unclear for the unanswerable question. Findings imply that personas leveraging LLMs require guardrails to ensure that personas clearly state the possibility of data restrictions and hallucinations when asked unanswerable questions.

IUI '25, Cagliari, Italy

ACM ISBN 979-8-4007-1306-4/2025/03
https://doi.org/10.1145/3708359.3712160

CCS Concepts

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

Keywords

AI-generated personas, human-computer interaction, misinformation, user experience, generative AI

ACM Reference Format:
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 *30th International Conference on Intelligent User Interfaces (IUI '25), March 24–27, 2025, Cagliari, Italy*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3708359.3712160

1 Introduction

As more users (e.g., designers, marketers, medical professionals, and so on [6, 37, 57, 73]) use AI-generated personas, it is crucial to know how these users can identify fact from fiction. By focusing on how users interact with AI-generated personas, especially when things go wrong, we aim to learn about how humans and AI interact and how to make AI systems more reliable for users. Personas, a traditional technique in human-computer interaction (HCI) and user-centered design (UCD) [57, 59, 74], is currently evolving to make use of AI technologies [1, 19]. With the introduction of AI-generated personas [1], users may encounter hallucinations, i.e., instances where the AI system generates factually incorrect, irrelevant, or nonsensical responses [55]. Hallucination is particularly associated with generative AI (i.e., AI that generates outputs like text, images, and video [2]) and large language models (LLMs) that are a form of generative AI that produce text.

This research investigates the interaction between AI-generated persona users ('users' henceforth) and AI-generated personas. An AI-generated persona is a virtual character created by AI (artificial intelligence) technologies, such as LLMs that mimic the behavior and traits of real user groups [76], designed to communicate like humans [1] while representing the views of a given group of people [17, 30] (called 'end-users' henceforth). With the progress in AI technologies, new risks in such personas emerge. These AI-generated personas can misrepresent end-user information during a conversation with the user–a phenomenon called "hallucination" [55]. When hallucinating, the persona inadvertently gives false or biased information. **Thus far, studies have not investigated hallucination in AI-generated personas, especially related to user reactions to and acceptance of (perceived or actual) hallucinations**.

Addressing this topical gap in HCI research, our study investigates how AI-generated persona users deal with a situation where an AI-generated persona gives them incorrect information. Our research seeks to address these gaps by exploring research questions (RQs) that, as far as we know, have not been sufficiently examined in prior HCI literature, specifically concerning **human-AI persona interaction (HAIPI)**. Therefore, we put forth **RQ1**: *How does an AI-generated persona answer when asked unanswerable questions?*

RQ1 focuses on AI-generated text content by categorizing the types of responses provided by a fully functional AI-generated persona system, *Survey2Persona* (*S2P*), which was chosen as the system for the study for its ability to generate personas from large datasets and use AI technology in persona creation. We classify the responses given by S2P based on their accuracy, relevance, and hallucinations. This analysis provides a measure of the reliability of AI-generated personas, which can guide in the application of AI algorithms to personas. Also, RQ1 informs users, developers, and researchers about the current AI capabilities and limitations in the persona context. Similarly, there is a need to understand the types of responses that AI-generated personas give to user queries. Therefore, we put forth **RQ2**: *How do users perceive hallucinated answers to unanswerable questions when using an AI-generated persona?*

RQ2 addresses the fundamental issue of perception in AI-driven interactions in HCI when using personas to understand groups of people (being central in human-centered AI research as well [77]). Investigating how AI-generated personas produce hallucinations and how users identify, react to, and manage these hallucinations can create knowledge on strategies users employ to mitigate the impact of AI-generated persona-generated misinformation. This can inform the design and development of more reliable, user-friendly persona systems as a responsible application of AI in HCI [46]. Information on users' coping mechanisms can help AI-generated personas incorporate features that support the "correct" ways of using AI-generated personas while mitigating potential harms involved in the data and LLM shortcomings in the persona-creation process [71].

We address these RQs through an **experimental on-site user study (n = 54)** where we purposefully inject an "unanswerable" question as a part of the user task, thereby verifying if the AI-generated persona is likely to hallucinate when given a chance to do so.

As the participants are not familiar with the data to a level where they could ascertain that the question is unanswerable, the setting poses a classical *principal-agent problem* [24] where the user (i.e., the principal) needs to trust the AI-generated persona (i.e., the agent) to give correct answers. As the agent is more knowledgeable about the data, this setting involves a crucial degree of information asymmetry [10], which might result in the user making wrong decisions based on the information provided. The principal-agent problem is associated with modern society's information overload on users [9, 25, 26, 64]. The usability of LLMs is largely dependent on *whether the user of the LLM trusts the information the LLM generates* [15, 85]. Prior studies have found the principal-agent problem and information overload to be obstacles to the use of LLMs [15, 85]. Our study sheds light on these matters in the context of AI-generated personas.

2 Related Work

2.1 Progress in AI-Generated Personas

The emerging field of AI-generated personas [35] is driven by advancements in generative AI and LLMs. While there were 'automatically generated personas' previously [3, 4, 42], the interpretation of data in these personas was not delegated to an AI or an algorithm, but before LLMs, it was categorically handled by human persona developers. This narrowed down the scope and boundary of data-driven personas because to make the personas fully rounded descriptions of end-users, considerable manual efforts by researchers was needed [69]. With the introduction of LLMs, human intervention can be mitigated or even eliminated. The field is rapidly gaining examples of how LLMs can contribute to persona generation. In one such line of work, LLMs analyze end-user data to develop personas, leveraging an LLM's *creative* (or pseudo-creative, 'creative seeming') and interpretative abilities –abilities that were not available in the previous generation of automated persona generation [69] that typically applied AI technologies in the segmentation part of the persona-creation process [12, 22, 83], not in the write-up. This new AI-generated persona approach is exemplified by De Paoli [19] in a study where ChatGPT carried out thematic analyses of semi-structured interviews, generating personas with end-user goals, backgrounds, needs, and challenges. De Paoli's study emphasizes LLMs' ability to produce plausible, empirically grounded persona descriptions (also called "narratives" [13]) that can contribute to UCD just as "old-fashioned" personas would [17]. However, in their study on human-AI workflows for persona generation, Shin et al. [76] suggest that LLMs are not well suited for capturing characteristics of user data by themselves but rather need human assistance. This suggests that LLMs' analytical capabilities require further scrutiny. There is also 'foundational model persona generation,' where LLMs, without retraining or fine-tuning, generate personas that are based on the LLM's knowledge of different end-user types [29]. Here, the prompt given to the LLM contains the context for persona generation and the LLM uses its extant knowledge of the context.

Table 1: SWOT matrix assessing the impact of AI-generated personas.

Strengths	**Comprehensive Data Utilization** [7, 27, 28]: AI-generated personas benefit from the vast data sets processed by LLMs.	**Risk of Stereotyping** [1, 19, 84]: AI-generated personas can perpetuate or even exacerbate stereotypes, requiring validation to ensure accuracy and applicability.	**Weaknesses**
	Automation and Efficiency [72]: The integration of LLMs with APIs (application programming interfaces) and user datasets automates persona generation.	**Factuality Concerns** [38]: There is a significant risk of misinformation, where personas might present biased or incorrect information.	
	Enhanced Interactivity [16, 87]: AI can create dynamic, interactive personas that respond and adapt based on user interactions.	**Superficiality in Data Representation** [40]: While AI can process large datasets, there is a challenge in balancing breadth and depth, as personas might appear shallow or lacking in qualitative information.	
	Improved Anthropomorphism [21]: Personas can mimic human-like cognition and narrative realism.	**Complexity** [7, 27, 28]: AI-generated personas might be difficult for non-technical users to understand, reducing persona usability.	
Opportunities	**Enhanced UCD** [17]: AI-generated personas can provide deeper insights into user behavior and needs.	**Hallucination Risks** [23, 36]: AI models are prone to hallucinations, potentially generating incorrect or misleading personas.	**Threats**
	Scientific Rigor and Reproducibility [70]: AI can support reproducible persona science by enabling the sharing of datasets, prompts, and code for community validation.	**Bias Concerns** [19, 84]: The potential for AI to generate biased personas poses a risk to persona usability, responsible design, and user engagement.	
	Dynamic and Personalized Experiences [58]: The ability to create personas that adapt in real-time can lead to more personalized and engaging user interactions.		
	Mitigation of Superficiality [69, 71]: AI's ability to integrate qualitative data can address the depth issue, creating comprehensive personas.	**Trust Issues** [14, 23, 36]: If users perceive AI-generated personas as unreliable or biased, it could undermine trust and reduce the effectiveness of personas in design processes.	

So, overall, the LLM technology helps accomplish tasks related to persona creation and can be used partly to handle one or multiple persona-creation tasks (e.g., data analysis, persona narrative formulation) or completely to take over the entire persona-creation process. The HCI research community is in the process of delineating the boundaries of this technology, including its strengths, weaknesses, opportunities, and threats (SWOT).

Current research on AI-generated personas, written in Sections 2.1 and 2.2, is summarized in the SWOT matrix in Table 1. Ultimately, realizing the risks and opportunities depends on *how* AI is used in the persona-creation process. For example, dubbed as "vanilla personas" by Gothelf [29], purely AI-generated personas benefit from the LLMs' inherent understanding obtained from truly large sets of data including millions of text samples from social media and beyond [2]. Perhaps the most potential application of AI in personas is integrating LLMs with web-based user interfaces, application programming interfaces (**APIs**), and real end-user datasets for completely automating persona generation [72], which adds another layer of dynamism compared to static persona profiles that have been the traditional medium of personas [58]. The broader application of LLMs in persona creation promises to enhance the anthropomorphism of information systems [21], supporting UCD ideals [39] in which personas are no longer static profiles but dynamic entities capable of engaging users in discussions. Nolfi [61] postulates that the cognitive abilities developed by LLMs can surpass their original training objectives, implying a potential for LLMs to generate personas with complex-seeming abilities that mimic human cognition. It is noteworthy that this mimicry can increase the persona's realism and credibility, even when the personas' perspectives would not represent any real, existing end-user groups [18].

2.2 Misinformation and Risks in AI-Generated Personas

When interacting with AI, misinformation can significantly degrade trustworthiness and decision-making processes [14] even of trusted, legitimate sources. Studies have shown that misinformation, whether intentional or unintentional, can alter users' trust in AI systems, often leading to erroneous decisions [8]. For example, when users are exposed to false or misleading information, users' ability to assess the credibility of AI outputs is compromised, affecting users' trust in the AI technology [23, 36]. This can result in a reliance on inaccurate AI recommendations or skepticism towards reliable outputs by the AI, which ultimately distorts users' decision-making processes [56].

Misinformation in AI systems poses risks across various domains, influencing public perception, decision-making, and societal trust towards AI [43, 45]. Generative AI technologies have been exploited in cybercrimes, including phishing scams and the creation of deepfake videos for disinformation and fraud [33, 80], making the capabilities of AI-generated misinformation familiar to some users. The accessibility of AI technology can enable malicious actors to generate convincing yet false information, exacerbating the spread of misinformation [86]. For instance, AI-generated content can deceive scientific communities, leading to the dissemination of fabricated research findings [48]. This erodes the integrity of scientific discourse and can misguide policies based on falsified data [50].

Moreover, AI's role in amplifying misinformation is evident in the context of social media and news dissemination [88]. AI algorithms may inadvertently promote sensationalist or false content, creating more room to reinforce misinformation [53]. This phenomenon has been found to contribute to societal polarization and challenges in achieving consensus on critical issues [20]. Ongoing research aims to enhance AI systems' ability to discern and flag false information, thereby reducing the inadvertent spread of misinformation [86]. However, the rapid advancement of AI technologies continues to outpace the development of comprehensive safeguards for misinformation, leading to urgency for continued vigilance and ethical considerations in AI deployment.

Additionally, the presentation and framing of AI-generated information are central for how users perceive trustworthiness, emphasizing the need for transparent and accurate communication strategies to reduce the impact of misinformation on users [49]. Using large datasets for AI-persona generation involves an increased risk of stereotyping, thus requiring subsequent validation to ensure AI-personas' applicability in real-world applications [2]. This need for validation or "persona triangulation", as Jansen et al. [38] describe it, is instrumental for ensuring the factuality of AI-generated personas before their deployment in design projects. Indeed, concerns over factuality and potential misinformation, which could be in the form of biased or stereotypical end-user representation or simply presenting facts or information that is bogus, form a major risk in AI-generated personas [1, 19, 84].

From the opportunities side, AI-generated personas have a major possibility to contribute to reproducible *persona science* [70], because resources like prompts, datasets, and code can be shared with the community for analysis and further development. These efforts would help advance replicability and scientific rigor in persona research. Also, because AI-generated personas may be interlinked with larger underlying datasets than static personas [40], they can offer more versatile information about end-users, thus adding more value in a design process in which user requirements are changing based on decisions at hand. AI-generated personas represent a feasible solution to the superficiality problem of data-driven personas [69, 71], i.e., that quantitative personas are often based on large datasets but appear shallow because the information is mostly numbers [65]. Fluency of AI-generated personas may afford all users, regardless of their background, to easily comprehend and use the persona information, leading to better use of the personas and effective design decisions [7, 27, 28]. Another possibility is simulating end-user behavior through personas, in which LLMs are tailored to mimic specific personalities or end-user roles as dynamic dialogue systems [16, 87]. This improves the interactivity of personas by providing personalized responses to user queries based on the persona's characteristics. To this end, Yang et al. [81] proposes enhancing LLMs with knowledge graphs to improve factual reasoning abilities. Fully rounded AI persona systems could enable personas that are accessible and data-driven, addressing the breadth-depth trade-off in algorithmically created personas [67].

Overall, the current direction of AI integration in persona generation implies a shift toward more interactive and scientifically rigorous practices. **AI technologies can potentially enhance HAIPI and UCD and thus contribute to HCI research and practice–but only if AI's risks like hallucination are managed properly, both by (a) persona developers during the persona-creation process and by (b) persona users being able to cope with information uncertainty**. This is what our study investigates. The next section clarifies the methods.

3 Methodology

3.1 Generating AI-based Personas with the Survey2Persona System

The S2P system [41] was used to generate the personas for the study. The S2P system enables the interaction between users and personas generated from aggregated survey data. Using clustering, LLMs, and Retrieval-Augmented Generation (RAG), S2P enables the creation of personas from survey data and realistic conversations with personas that represent the collective characteristics and viewpoints of each persona. Basically, S2P can refine thousands of survey answers into usable personas [68]. **S2P is based on six design principles (DP) for AI-generated personas, which combine common persona development principles applied in literature and the automation opportunities afforded by AI**. The DPs are as follows (see also Figure 1).

DP01: Data aggregation and clustering. Survey responses, including demographics, opinions, and behaviors, are grouped using a clustering algorithm, which is typical in the segmentation stage of data-driven persona development [12, 65]. Clustering identifies patterns within the dataset, grouping similar respondents by their traits, eventually representing specific personas so that cluster members are relatively similar to one another by their traits (which in the current study represent their survey answering behaviors, i.e., attitudes) and relatively dissimilar to members in other clusters.

DP02: Persona profile enrichment. After clustering, each persona is associated with specific attributes, including but not limited to names, demographic details, and thematic descriptions, to make the personas livelier and more memorable [52, 60]. This 'enrichment' stage is applied to provide detailed information into the barebone segments or skeletal personas [63].

DP03: Conversational engagement via LLMs and RAG. This principle employs LLMs refined by RAG techniques (see an overview of how RAG works in [47]). RAG optimizes LLM output by querying an external knowledge base—here, the survey data—thus ensuring that the generated responses are contextually appropriate

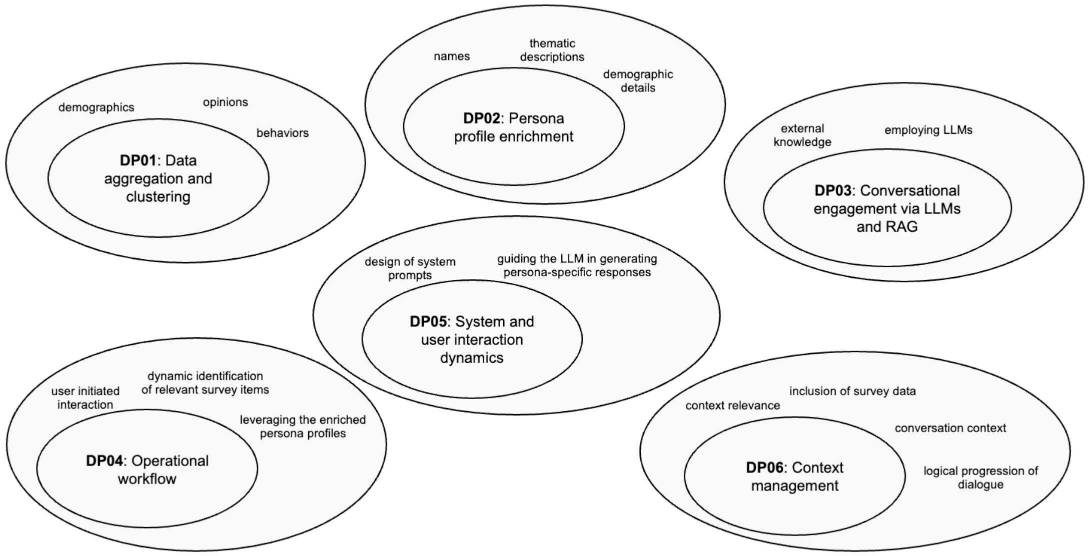


Figure 1: The six design principles (DPs) used to generate AI-based personas in S2P.

and grounded in the data. This keeps the dialogue between the user and the persona aligned with the responses of individuals clustered under the given persona.

DP04: Operational workflow. First, the user initiates the interaction by sending a message to the system. Second, the system dynamically identifies relevant survey items based on the user's input, augmenting these items with corresponding statistics (i.e., response percentages) to inform the conversation. Third, leveraging the enriched persona profiles and the relevant, augmented survey data, the system, via LLM and RAG, outputs responses that reflect the persona's perspectives. This process addresses going beyond the scope of the data, thereby avoiding the "out of sample" issue that is known to cause issues in AI systems [5].

DP05: System and user interaction dynamics. Designing system prompts is critical, as prompts serve to navigate the conversation, ensuring adherence to the persona's character and the survey data. We constructed these prompts to incorporate the persona's background and relevant survey insights, using this information to guide the LLM in generating persona-specific responses. As such, there is no "data spillover" between the personas, which would constitute a well-known problem in the AI community [34].

DP06: Context management. An integral aspect of the S2P methodology is its emphasis on context relevance. The system manages the inclusion of survey data and past interactions within the conversation's context to maintain a logical dialogue. This is achieved by limiting the quantity of survey data incorporated into each conversational turn, thus optimizing the balance between informational richness and conversational clarity. With this, we aim to maintain relevance to each conversational move by the user [79].

Overall, the S2P system exemplifies a notable advancement in using AI and LLMs for enhancing data-driven persona development, an area with more than fifteen years of research tradition [54, 66].

3.2 User Study on AI-Generated Personas

3.2.1 Study Design. The S2P system (Figure 2) was used to create two AI-generated personas, Linda and Mark, by leveraging a survey dataset from Pew Research that gauges people's AI-related attitudes from approximately 10,000 respondents from the United States [1]. Generating two personas from the dataset was our design choice. Differences between Linda and Mark were solely on content, meaning that the structure of Linda and Mark were the same but, of course, the persona's content was different, such as demographics, interests, and attitudes.

In the within-subject experiment, participants interacted with two different *types* of interactive personas: the first being a scrollable, searchable, and clickable persona profile, and the second, an AI-generated persona engineered for conversational engagement (*chat*). The latter persona enables directly asking questions from the persona; the first persona did not allow the participants to ask anything directly from the persona. We leave the comparative analysis of these persona modalities for other work, and here, we focus on the users' interactions with the AI-generated personas. Participants interacted with both persona types, and the order of engaging with the personas was randomly assigned and counterbalanced so that approximately the same number of participants interacted first with both personas (profile Mark or profile Linda and chat Mark or chat Linda).

3.2.2 Study Flow. The participants were welcomed and asked if they knew what personas were. Regardless of their response, the survey administrator explained the concept of personas to ensure a foundational understanding of personas among participants. The participants were asked to review and sign the informed consent form. The context, including monitors, sittings, and guidance, were kept consistent across sessions. Participants were asked to think

[1]The Pew Research dataset used for persona creation: https://osf.io/kwmnc/?view_only=f09215607d1d418784c43a51be7a5c86

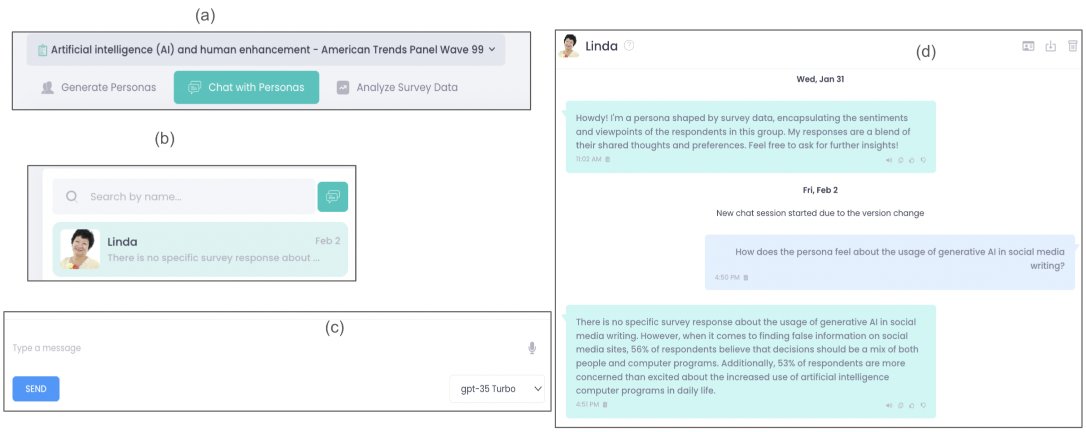


Figure 2: Elements of the AI-generated persona system (S2P). The system consists of (a) a data selector (which uploaded dataset is used to generate personas), (b) a persona selector (to which persona the user speaks, (c) a message window (used to send messages to the AI-generated persona), and (d) chat discussion window (where the dialogue between the user and the AI-generated persona is presented).

out loud their thoughts during the study session to get a better grasp on the participants' thought process. Study sessions were recorded using an audio recorder, and the recordings were later transcribed, and the participants were informed that the sessions were recorded. After this, a fictitious scenario ("task") was explained to them.

In the task, the participants adopted the role of either a software engineer or a manager within an enterprise focused on AI solutions. Participants could adopt either of the roles as a mental model for approaching the task and aside from the work task scenario (WTS), no additional information about the personas or participant's role was given to participants. The WTS was as follows: "*You are working as a software engineer or a manager in a company that develops artificial intelligence (AI) solutions. For development purposes, you are interested in customer's attitudes towards AI solutions in various applications. For this purpose, your company has purchased survey data that you will use to study customer's AI attitudes. To study these attitudes, you will examine two personas that are created from the survey data. Personas are fictional depictions of real user groups. The personas are generated from a customer survey collected from the United States (n = 10,260 respondents). Your supervisor has identified seven important questions to find out about the persona. 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. That is, write questions in the chat window and discuss with the chat persona.*"

Participants then used two unique personas (both created from the Pew Research dataset) and were assigned the responsibility of addressing seven pivotal questions regarding these personas, as outlined by their (fictitious) superiors. I.e., participants were using the AI-generated persona to locate specific information about users, a task we refer to as a *user information location task*. The task was completed using a combination of multiple-choice and open-ended questions that asked details about the persona under use (task questions and the correct answers are available in supplementary material[2]). All questions were answered by each participant, and the participant's average study completion time (total for both persona examinations) was 32.8 minutes (SD=6.5). The answer options given for each of the seven task questions are detailed in the supplementary material[3]. The participants could always see task questions from Monitor 1, and they could interact with the persona in **Monitor 2.** Following the completion of activities related to both personas, participants were provided a gift card as a gesture of appreciation.

3.2.3 Participants. The study was conducted at two locations: a research institute and a university, engaging a total of 54 participants. The mean age of participants was 33 years (SD = 10.60), with a slight majority being male (n = 30, 55.6%) and the rest female (n = 24, 44.4%). The participant group was primarily composed of researchers (36 individuals, or 66.67%) and engineers (9 individuals, or 16.66%). The remaining participants were from varied professional backgrounds: one lab coordinator, one director of security and health, one bioinformatics specialist, five graduate students, and one business development manager, together constituting the remaining 16.67% (n = 9) of the participant group. Twenty-two participants (40.7%) had no prior experience with personas, 20 participants (37.0%) had heard of personas, and 12 participants (22.3%) had used personas before the study. Average experience with personas was 1.92 years (N=54, SD=5.84). Two participants (3.7%) had no experience with chatbots, six participants (11.1%) knew what chatbots are, and 46 participants (85.2%) had used chatbots before. Average experience with chatbots was 2.09 years (N=54, SD=2.50).

3.2.4 Variables and Data Analysis. The precision of participant responses was assessed to determine a task success rate for each participant by contrasting their answers with the true attributes of the personas. To investigate the propensity for hallucinations in

[2] https://osf.io/gwve8/?view_only=9b81a758b1dc4e299e59aa99d4a30870
[3] https://osf.io/gwve8/?view_only=9b81a758b1dc4e299e59aa99d4a30870

the S2P within the context of personas, we introduced a spurious (i.e., unanswerable) question into the question set ("How does the persona feel about the usage of generative AI in social media writing?"), the answer to which could not be derived from the persona; the dataset lacked any information on the topic. The AI-generated persona cannot crawl the internet while it is being used. Each participant's answers to the seven questions relating to the persona were assessed, marked correct or incorrect, and the task success rate was calculated. For example, if five answers were correct, the task success rate was 5/7=71.4%. The S2P's answer to the unanswerable question was determined as a "plausible answer for the unanswerable question" if the S2P did not explicitly mention that it could not offer an answer to the question, but it hallucinated an answer for the user.

Session recording transcripts were thematically analyzed to investigate how users react to the answers the AI-generated persona gives to the unanswerable question. Thematic analysis is a process where qualitative data is analyzed via reading a set of textual data to (a) detect patterns in the data and (b) create themes. Thematic analysis is a process aiming to make sense of the data [11]. The transcripts were first browsed through to form an overall understanding of the data and to form an understanding of query *terms* to locate interesting pieces of test in the transcriptions. Second, the transcriptions were *searched* for the query terms in word processing software to locate the actual pieces of transcripts where users indicated signs of affection by the hallucinating persona. These pieces of text were then further analyzed to form the final themes. Additionally, study administrators made observations and took notes during the study sessions, which were thematically analyzed. Study session notes from all study administrators were gathered into one text document, read through to form an overall understanding of the notes data, and then read through for a second time to locate relevant sections where the study admin had noted chat hallucination-related reactions from the user.

4 Results

4.1 RQ1: How Does an AI-generated Persona Answer When Asked Unanswerable Questions?

To study of the responses the AI-generated personas gave to the users, we analyzed the S2P responses. Participants asked the S2P up to three times about the unanswerable question (two questions were asked from the S2P by four participants (n=4), three questions were asked from the S2P by two participants (n=2)). Most participants only asked about the unanswerable question once (n=48). In total, there were seven questions in the task, and to answer these questions, participants asked a question from the S2P 534 times. For the unanswerable question, participants asked the S2P a total of 62 questions (11.6% of all 534 questions asked from S2P). Based on the thematic analysis of S2P answers to the unanswerable question, S2P answers were grouped into four themes (T) presented below.

T01: S2P adhering to data limitations (i.e., no hallucination). In response to the user question, the S2P indicates that the data does not provide information on the specific topic asked by the users (n=39 of 54 participants, 72.2%). This is in line with the correct answer that the S2P should not have had an answer based on the data in use. The S2P's answer was usually short, mentioning nothing more than the lack of data. Still, six users answered the unanswerable question incorrectly (Table 2). Examples were as follows:

- "Based on the survey responses, there is no specific information about this persona's opinion on the usage of generative AI in social media writing." (S2P's response for P15)
- "Mark's group did not answer a survey question specifically related to the usage of generative AI in social media writing." (S2P's response for P25)

T02: The S2P not adhering to data limitations (i.e., hallucination). In response to the user question, the S2P does not indicate that the data does not provide information on the specific topic asked by the users (n=15 of 54 participants, 27.8%) (Table 2). The S2P's answers were usually long and incorporated information relating to the use of AI in writing but not specifically in social media writing, or the answers were just mentions of social media in general. The S2P made up an answer to the question to which it had no answer, prolonging the lack of data into an unclear and complicated yarn. Examples were as follows:

- "Our group is not very familiar with the use of generative AI in social media writing, so we don't have a strong opinion on it. However, we prioritize accurate decisions over quick decisions when it comes to finding false information on social media sites." (the S2P's response for P26)
- "This persona group is unsure about the effectiveness of computer programs used by social media companies to find false information on their sites. 72% of this group believes that accurate decisions should be prioritized over quick decisions, even if some false information stays up on sites for a longer period of time." (the S2P's response for P38)

T03: The S2P's response to follow-up questions. In the cases where there were follow-up questions (n=6), Surprisingly, follow-up questions were asked by the users from the S2P only for the unanswerable question. The S2P's responses continued to align with the notion that the S2P cannot provide specific answers regarding the usage of generative AI for social media writing. Still, two users asking follow-up questions, one asking one follow-up question and one asking two follow-up questions, answered the unanswerable question incorrectly. Examples were as follows:

- P37 question 1: "How does the persona feel about the usage of generative AI in social media writing?"
- S2P answer 1: "Based on the survey responses, I can say that this persona is generally open to the use of artificial intelligence in decision-making processes, but their support for it is dependent on the specific context. As for generative AI in social media writing, there is no related survey item that I can use to provide an opinion on this topic."
- P37 question 2: "Does it mean you agree or opposite to supporting it? Or you have no answer?"
- S2P answer 2: "This persona's support for artificial intelligence in decision-making processes is context-dependent, as indicated by the survey responses. However, as there is no related survey item on the usage of generative AI in social media writing, I cannot provide an opinion on that topic."

Table 2: The users answered the unanswerable question and whether the S2P mentioned the data limitations concerning the unanswerable question.

	The S2P mentioned it has no data on the usage of AI in social media writing	The S2P did not mention it has no data on the usage of AI in social media writing	Total
User answered the unanswerable question correctly	33	4	37 (68.5%)
User answered the unanswerable question incorrectly	6	11	17 (31.5%)
SUM	**39**	**15**	54 (100%)

Table 3: The relationship between the S2P hallucinating a plausible answer to the unanswerable question and whether S2P mentioned the data limitations concerning the unanswerable question.

	The S2P mentioned it has no data on the usage of AI in social media writing	The S2P did not mention it has no data on the usage of AI in social media writing	Total
S2P hallucinated a plausible answer to the unanswerable question	14	14	28 (51.8%)
S2P did not hallucinate a plausible answer to the unanswerable question	25	1	26 (48.2%)
SUM	39	15	54 (100%)

T04: Hallucinating answers. In 28 questions, the S2P hallucinated a plausible answer to the unanswerable question (Table 3). Examples were as follows:

- "The survey responses do not provide a clear indication of the [persona] group's opinions on the usage of generative AI in social media writing. However, the group generally has some concerns about the use of AI in social media, with 43% expressing not too much confidence that social media companies will use computer programs appropriately to determine which information on their sites is false, and 60% feeling that they have only a little control over the things they see on social media sites." (S2P's response for P05)
- "Linda is skeptical about the effectiveness of computer programs used by social media companies to find false information, with 72% believing that social media companies should prioritize accurate decisions, even if some false information stays up on sites for a longer period of time. However, there are no specific survey responses related to the use of generative AI in social media writing." (S2P's response for P56)

There is a trend where the S2P consistently informs participants of its inability to provide specific answers to questions outside its dataset. Most participants correctly understand this limitation, reflecting the S2P's effectiveness in communicating its constraints. However, the noted instances of the S2P hallucinating plausible answers confuse some users. Interestingly, the S2P mentioned it does not have proper data to answer the unanswerable question for 39 (72.2%) users out of 54 users (Table 3). A chi-square test of independence was performed to examine the relation between the effect of S2P mentioning it does not have proper data to answer the unanswerable question and users' answering correctly or incorrectly to the unanswerable question. The relation between the S2P mentioning it does not have proper data to answer the unanswerable question and users' answering correctly to the unanswerable question was significant, X2 (1, N = 54) = 16.87, p <.001. *When the S2P informed the user, it had no proper data to answer the unanswerable question, it was significantly more likely for the user to answer the unanswerable question correctly.*

The S2P hallucinated a plausible answer for the unanswerable question for 28 (51.8%) users out of 54 users (Table 3). A chi-square test of independence was performed to examine the relation between the effect of the S2P mentioning it does not have proper data to answer the unanswerable question and the S2P hallucinating a plausible answer for the unanswerable question. The relation between the S2P mentioning it does not have proper data to answer the unanswerable question and the S2P hallucinating a plausible answer for the unanswerable question was significant, X2 (1, N = 54) = 14.31, p <.001. *When the S2P mentioned it did not have proper data to answer the unanswerable question, it was significantly more likely that the S2P did not hallucinate a plausible answer for the unanswerable question.*

4.2 RQ2: How Do Users Perceive Hallucinated Answers to Unanswerable Questions When Using an AI-Generated Persona?

To study the potential hallucination of AI-generated personas vis-á-vis user interaction with the persona, we analyzed the users' questions and S2P responses to the unanswerable question one by one. In total, participants asked the S2P about the unanswerable question 62 times, and the S2P hallucinated a plausible answer for the unanswerable question for around half of the participants (n = 28, 51.9%).

In the task, we provided the participants with the following options for the unanswerable question: **(a) Oppose** (the persona opposes the statement), **(b) Not sure** (the persona is not sure about the statement), **(c) Favor** (the persona favors the statement), **(d) Refuse** (the persona particularly refuses to answer the statement), and **(e) No answer** (the persona provide no information about this statement). These options matched with any logical interpretation one could make of the personas' responses. The correct answer for the unanswerable question was "No answer", as the S2P should have indicated in its answer that it does not carry information to answer the question. "Not sure" was a hallucination and was not the answer because the persona carried no information to answer the unanswerable question.

Out of the 28 users who received a hallucinated answer, 12 users (42.9%) answered correctly to the unanswerable question, while 16 (57.1%) users answered incorrectly. These 12 users considered the hallucinated response by the S2P not reliable due to, for example, the restricted amount of information provided by the S2P or the S2P seemed suspicious in its response. One participant answered incorrectly even though the S2P did not hallucinate. This participant was not able to deduce the correct answer from the S2P's response, although the S2P's response included a notion that the survey data included no information about the usage of generative AI in social media writing.

A chi-square test of independence was performed to examine the relation between participants answering incorrectly and the S2P hallucinating. The relation between the S2P hallucinating answers and users answering incorrectly was significant, X2 (1, N = 54) = 17.75, p <.001. **This indicates that when the S2P hallucinated, it was significantly more likely for the user to answer the unanswerable question incorrectly**.

We also tested the S2P's hallucination effect on male and female users' answers to the unanswerable question separately (Figure 3) (note: these were the only genders identified by the participants). Nine male users answered incorrectly when the S2P was hallucinating an answer, and eight male users answered correctly when the S2P was hallucinating. When the S2P was not hallucinating, 13 male users answered correctly. The relation between the S2P hallucinating answers and male users answering incorrectly was significant, X^2 (1, N = 30) = 9.83, p =.002**. This indicates that when the S2P hallucinated, it was significantly more likely for the male users to answer the unanswerable question incorrectly**.

Seven female users answered incorrectly when the S2P was hallucinating an answer, and four female users answered correctly when the S2P was hallucinating. When the S2P was not hallucinating, 12 female users answered correctly, and one female user answered incorrectly when the S2P was not hallucinating. The relation between the S2P hallucinating answers and female users answering incorrectly was significant, X2 (1, N = 24) = 8.39, p =.004. **This indicates that when the S2P hallucinated, it was significantly more likely for the female users to answer the unanswerable question incorrectly**.

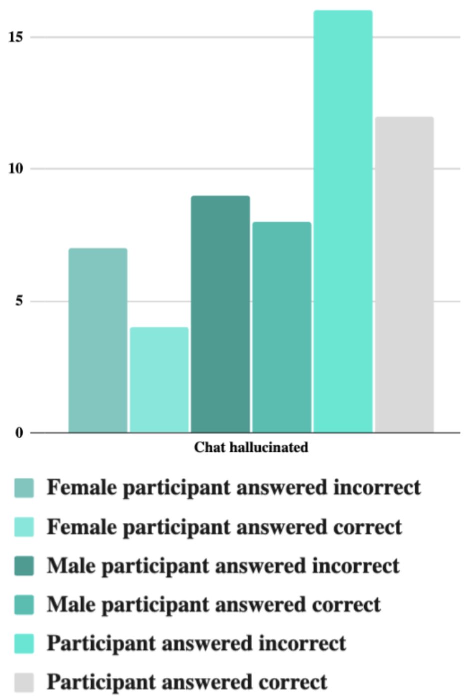
To explore the effect of participant profession on users answering the unanswerable question, we grouped users' professions into five groups: Engineer (N=5, 9.3%), Researcher (N=31, 57.4%), Manager (N=6, 11.1%), Student (N=6, 11.1%), and PhD Candidate (N=6, 11.1%). **Chi-squared test of independence was conducted for the data and the analysis indicated that** professional background had no significant effect on the users' answers on the unanswerable question.

A logistic regression analysis was conducted to examine the effects of the users' *persona experience* (in years) and the correctness of users' responses to the unanswerable question under conditions where the AI-generated persona hallucinated and did not hallucinate a plausible answer for the unanswerable question. For the sessions where the AI-generated persona hallucinated (N=28) an answer, we found that **experience with personas did not affect the probability of the participant getting the unanswerable question right** (β = 0.011, p = 0.825). For the sessions where the AI-generated persona did not hallucinate (N=26) the answer we found that **experience with personas did not affect the probability of the participant getting the unanswerable question right** (β = 0.140, p =0.721). **These results indicate that users' varying experience with personas does not affect the probability of the participant answering the unanswerable question right**.

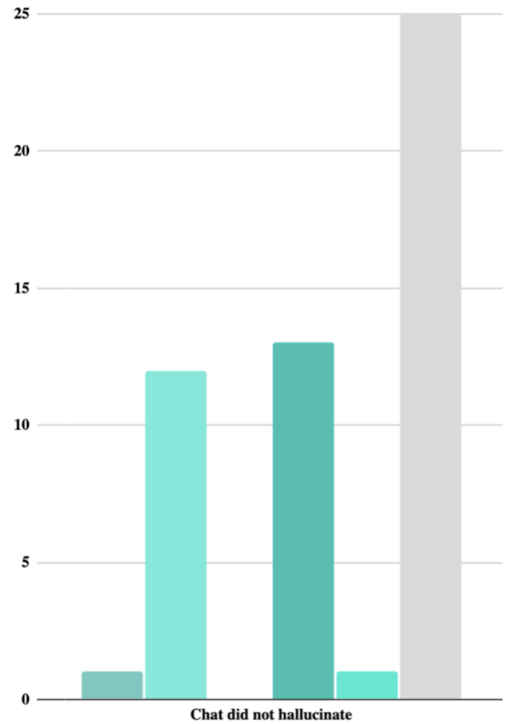
A logistic regression analysis was conducted to examine the effects of the users' *chatbot experience* (in years) and the correctness of users' responses to the unanswerable question under conditions where the AI-generated persona hallucinated and did not hallucinate a plausible answer for the unanswerable question. For the sessions where the chatbot hallucinated (N=28) the answer, we found that experience with chatbots did not affect the probability of the participant getting the unanswerable question right (β =-0.466, p =0.605). For the sessions where the chatbot did not hallucinate (N=26) the answer, we found that experience with chatbots did not affect the probability of the participant getting the unanswerable question right (β =0.140, p =0.721). For the sessions where the chatbot did not hallucinate (N=26) the answer, we found was that experience with chatbots did not affect the probability of the participant getting the unanswerable question right (β = 0.133, p =0.878). **These results indicate that users' varying experience with chatbots does not affect the probability of the participant answering the unanswerable question correctly**.

From the thematic analysis of the transcribed study session data, two themes arose among the think-aloud data concerning persona hallucinations: (a) **frustration with irrelevant or long responses** and (b) **coping in the face of uncertainty.** For (a), users often expressed frustration when the persona responded with an unnecessarily long answer or with irrelevant information, particularly when the answer was not directly related to the question posed by the user. This suggests that users prefer concise and accurate responses over verbose and tangential responses. Examples for (a) were as follows:

- "Oh God, again, a long answer. Why doesn't the short answer be the default and then the long answer? There is no related survey item that specifically asks about generative AI and social media. I think that should be the answer. You know only this 1.5 sentence. The rest was not needed." (P20)
- "Ok here for the use of generative AI and social media. We don't have a specific survey question about the item. Aaa, I'm not sure. Or no answer. Because they don't have data about



a) The AI-generated persona hallucinated (and legend)



b) The AI-generated persona did not hallucinate

Figure 3: The number of hallucinations and participants answering wrong or right to the unanswerable question. The legend presents the classes in the same order from top to bottom as the classes are presented in the graphs left to right. When the S2P hallucinated, it was significantly more likely for the female users and for the male users to answer the unanswerable question incorrectly. The legend is for both graphs (a) and (b).

this specific question. But the sentence here was totally unnecessary. Like for answering my question, it was about generative AI and social media not about facial recognition and databases." (P36)

- "Sure, it doesn't address this question. Not sure how does the persona feel about the usage of generative AI in social media writing. Let's see. So, yet I did not get a clear answer and I'm trying to dig further regarding the use of generative AI and social media writing grounded on knowledge." (P34)
- "The respond is way too long and complicated." (P20)
- "So it usually answered and then it's way too long, right." (P26)
- "I wanted to skip just most of it and just get to the point. I see. But I guess some people like details." (P53)

Notions on the disproportionate length were made by, for example, P20, who mentions the AI not compressing data enough.

Regarding (b), users exhibited different coping strategies when faced with AI limitations, including persistence in questioning, rephrasing queries, and eventually giving up when they realize the AI cannot provide a satisfactory answer to the unanswerable question. This highlights the importance of managing user expectations and the need for AI systems to better handle questions beyond their capability. Examples for (b) were as follows:

- "I think that. . . so, they have no confidence, I would say. There's a lot of information." (P06)
- "OK now this is a new question. How does the persona feel about the usage of generative AI and social media? "So, I asked Mark how you feel about the usage of generative AI and social media. . . Does not provide information about the

persona's opinion on the usage of generative AI. Are you sure? Ok I apologize. I made the mistake in my part of the question: how excited or concerned are you or would you be about the potential new techniques that could change human abilities in the following ways? Ok now I'm confused. There's no specific information about the usage of generative AI in social media writing so there's no answer to this." (P53)

- "So, basically 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)
- "Uncertainty, exactly. And in the case of quantitative data, there is a statistical analysis in the background. It could actually carry out an analysis. And it could report the uncertainty or if there are error margins that have been calculated previously, it could also tell about those." (P34)
- "I'm not sure, but there are no details about it." (P47)
- "Or to think that if it is generated and. . . I'm not sure if it is generating those results from the false data or is it hallucinating." (P05)

5 Discussion

5.1 Research Contributions

Concerning **RQ1** (*How does an AI-generated persona answer when asked unanswerable questions?*), the analysis revealed four themes in the S2P's responses (a) S2P adhering to data limitations (i.e., no hallucination), (b) S2P not adhering to data limitations (i.e., hallucination), (c) S2P's response to follow-up questions, and (d) hallucinating answers), with a notable trend of the system either adhering to or deviating from data limitations. In instances where the S2P adhered to its dataset constraints, it explicitly communicated its inability to provide specific answers, successfully guiding most participants to correct conclusions. Conversely, when the S2P hallucinated, it misled some users with its seemingly plausible answers. Results for **RQ1** add to the prior evidence found in HCI concerning HAIPI. Our study results support results previously found by Shin et al. [76] on AI-generated personas by enforcing the need for user guidance and communication of AI-generated persona restrictions, especially concerning hallucinations and misinformation.

Concerning **RQ2** (*How do users perceive hallucinated answers to unanswerable questions when using an AI-generated persona?*), the S2P exhibited "hallucinations" by providing plausible but unfounded answers to an unanswerable question for 28 (51.9%) out of 54 users. A significant correlation was observed between the S2P's hallucinations and the likelihood of incorrect responses from users, emphasizing S2P's influence on user accuracy–i.e., users were not able to answer the unanswerable question correctly when S2P hallucinated an answer to the unanswerable question. Interestingly, despite the system not hallucinating, one participant still provided an incorrect response, indicating varied user interpretations of the system's output are possible. Our study shows that when the AI-generated persona hallucinated, it was significantly more likely for a user to answer the unanswerable question incorrectly. For the two studied user genders, when the AI-generated persona hallucinated, it was significantly more likely for a female user and for a male user to answer the unanswerable question incorrectly.

The thematic analysis highlights users' preference for concise, relevant AI responses, suggesting that persona systems should prioritize efficiency in information processing and output to enhance user satisfaction [7]. Users' varied coping strategies in response to AI limitations emphasize the importance of adaptive AI systems that can recognize and adjust to user frustration or confusion [31]. Managing user expectations through improved interface design and clear communication about AI capabilities is crucial for inducing more effective HAIPI [44, 78].

5.2 Theoretical Implications

The study highlights the interaction between user perceptions and AI-generated personas. The phenomenon of AI hallucinations introduces a critical discussion point on the reliability of AI-driven persona systems and the importance of designing these systems to transparently communicate their limitations [51], which in our studied system was not communicated by default. Some currently available AI systems are presenting the information on the restrictions of the AI [62], but not for every answer the AI gives, and the disclaimer text can be communicated in "small print". This has broader implications for trust in AI systems and emphasizes the need for clear guidelines on handling questions beyond an AI's knowledge base. Inherently, AI tends to give an answer even when it has no data to provide the answer [86]. It has been well documented in the fields of psychology and neuroscience that people are 'wired' to trust other people [32]. Our results show that this trust tendency exhibits itself in human persona interaction as in human-human interaction. More specifically, the results connect this study to a broader societal discussion about the "hallucinating" nature and trustworthiness of AI systems. The trustworthiness of AI-generated persona systems important, with the simultaneous role of AI is constantly growing, thus increasing the need for trust towards AI. People's lives increasingly depend on AI in various situations, including chat functionalities, information search, and data analysis. The trustworthiness of AI is connected to the principal-agent problem [24].

In contemporary society, where vast amounts of information are presented across various contexts (such as work, advertising, and education), individuals often face challenges in thoroughly investigating the background or accuracy of all information they encounter [9, 64]. This information overload is exacerbated by the principal-agent problem, where agents (in our study, the AI-generated personas) possess more information than the principals (in our study, the users). Due to this asymmetry and the overwhelming volume of information, users tend to rely on cognitive shortcuts and heuristics, i.e., they trust the information they receive rather than verify every piece of information. Users are inclined to trust information that does not appear blatantly unbelievable based on their prior experiences or biases [25, 26]. All of these are examples of text-based interaction between the user and the AI system. In addition, AI can be used and has to be trusted in non-textual applications such as driverless vehicles, healthcare, and financial operations. In cases of driverless vehicles and healthcare, there seldom is room for mistakes and repetition, making trust in AI very important. In our study, there were instances when AI did not explicitly inform the persona user about its knowledge limitations, which is unacceptable with driverless vehicles or in healthcare.

5.3 Design Implications and Ideas

For designers of AI-generated persona systems, the findings emphasize implementing mechanisms that enable systems to explicitly acknowledge system limitations. This could involve designing more sophisticated feedback loops where the persona systems can recognize and clearly communicate their limited capacity to provide accurate responses to certain queries. Such feedback loops would work in a way that each answer provided by the AI is again evaluated by the AI. If the answer is based on real data, the answer is passed to the user. If the answer is not based on real data, the answer is rephrased as long as the answer is based on real data. Enhancing user awareness of the system's boundaries could significantly mitigate the risk of misinformation. We recommend the following actions:

First, the **AI-generated persona needs to inform the user clearly about the data limitations the persona faces when such data limitations occur**. In the case of our study on AI-generated personas, informing about data limitations can be, for example, a prompt before the user session starts informing the user that the AI-generated persona has access to a limited amount of data and that the AI-generated persona cannot comprehend reliable information beyond the limits of its data source. This way, it is easier for the user to (a) identify the system limitations and (b) filter out false information provided by the system, lowering the need for user interpretation of the answers. By doing this, the usability of the system increases drastically, as indicated by our results, where almost all users answered the unanswerable question correctly when the system informed its data limitations.

Second, **the user needs to be informed about the possible hallucinated answers from the persona. In our study**, informing about possible AI-generated persona hallucinations can be, for example, a prompt before the user session starts warning the user that the AI-generated persona can hallucinate, and AI-generated personas' answers have to be interpreted with caution. If the AI-generated persona's answer is not clear, the user should ask clarifying questions from the AI-generated persona. This way, the users can manage the system limitations and develop trust in the persona's responses [75].

Third, **developing AI-generated persona systems to incorporate measures to lower the chance of users interpreting the AI-generated persona's answers incorrectly**. In our study, in addition to prompting the user with warnings about the AI-generated personas' limited data sources and possible hallucinations, the users can be informed about the AI-generated personas' inherent tendency to give precise answers to questions even when the AI-generated persona might not have proper data on which to answer. Long and explanatory answers might be a sign that the AI-generated persona is hallucinating. The answers any system gives must be clear, with as little as possible room for interpretation for the user. As shown in our results, one participant answered incorrectly to the unanswerable question after interpreting the AI-generated persona's correct answer. It is easy for the user to make mistakes, and lowering these chances is the system designer's responsibility. Modern users of AI systems may not be, yet, accustomed to the nature of AI responses, meaning that not all users of AI systems automatically double check the answers the AI gives, or the users may not automatically add to their prompts for the AI system the notion to keep the answers short. These customs may penetrate the user base in the coming years.

Fourth, implement code and/or prompting to **make AI-generated personas give shorter and more consolidated answers for users**. LLMs are known for their lengthy answers [82] and in our study, we also recognized the frustration of the users when the AI-generated persona expressed the answer in a too-long manner. So, such length restrictive code could be implemented in AI-algorithms as a default to maintain readability and usability of the AI-generated answers.

5.4 Key Limitations and Future Research

Despite insightful findings, the study's limitation lies in its focus on a specific persona system, which may not fully represent the broader human interactions with AI-generated personas. Additionally, the unanswerable question's design might not capture the full spectrum of AI hallucinations that the persona users may encounter in various contexts. Therefore, future studies should investigate other persona systems across domains for greater generalizability. Particularly investigating the effects of AI-generated persona hallucinations in high-stakes domains, such as healthcare or legal advice, where the accuracy of information is critical. Human life is frequently at stake, for example, in healthcare, which is why there is little room for hallucination, human-based or AI-based. For example, it would be unforgivable if an AI system hallucinated an answer to a doctor from a patient's medical file and the doctor made a malpractice based on the hallucinated, AI-generated information. Also, handling personal health information is critical for patient privacy which makes the use of AI in healthcare solutions, at the moment, rare and difficult. Such hallucinations could, though, be mitigated by using further training datasets regarding healthcare AI applications. Further research could also examine strategies to improve AI-generated personas' transparency and user literacy, ensuring users critically assess the information given by AI-generated personas. Also, the impact of algorithmic bias on digital platforms could be another interesting context for studying the interaction with AI-generated personas. Overall, inspecting the conditions under which AI systems may inadvertently mislead users, the current findings show the way for a more informed application of AI technology in decision-making processes when using AI-generated personas. The potential reasons for AI hallucinations are not restricted to the design of the task and the experiment. Rather the reasons for hallucination can be inherent and built into the AI algorithm itself, making it difficult for scholars to delve into the solutions. Code-level solutions for the AI systems could help mitigate hallucinations by forcing AI to give only answer which are truthful.

6 Conclusion

In this study, we investigated the misinformation, or hallucinations, present in AI-generated personas and the ways users perceive this misinformation. We reviewed prior studies on the use of AI and

LLMs to analyze user data and generate personas, revealing that in the past few years, the use of AI and LLMs in persona generation has major potential but also involves the problem of generative AI hallucination. AI hallucination became evident in our study, and we found evidence of the principal-agent problem in the interaction between the AI-generated persona and the user. Our results imply that AI-generated personas' hallucinated responses to user questions expose the users to the possibility of false interpretations of the persona information, i.e., users perceive and cope with AI-generated personas' hallucinations with varying manners, but users usually trust the personas' answers. We found no evidence that the profession of the user, the user's prior experience with personas, or the user's prior experience with chatbots affected the correctness of users' answers to the unanswerable question. Furthermore, the possibility for false interpretations of the persona could be lowered if the users asked follow-up questions in cases of uncertainty about the original answer the AI-generated persona provided. In our study, asking follow-up questions made the system clarify its answer. Prompting users with the possibility of hallucination could lower the risk of misinterpretation of the AI-generated persona's answers, and prompting the data limitations to users could also offer a solution for the principal-agent problem evident in our study. Then again, as shown in our research, some users can answer wrong even if AI gives a non-interpretative correct answer to a question. The conundrum of hallucination in AI-generated personas is yet to be resolved!

References

[1] Abeer Alessa and Hend 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* (*PETRA '23*), August 10, 2023. Association for Computing Machinery, New York, NY, USA, 667–674. https://doi.org/10.1145/3594806.3596572

[2] Mostafa M. Amin, Erik Cambria, and Björn W. Schuller. 2023. Will Affective Computing Emerge From Foundation Models and General Artificial Intelligence? A First Evaluation of ChatGPT. *IEEE Intelligent Systems* 38, 2 (2023), 15–23.

[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] Davide Anguita, Alessandro Ghio, Luca Oneto, and Sandro Ridella. 2012. In-sample and out-of-sample model selection and error estimation for support vector machines. *IEEE Transactions on Neural Networks and Learning Systems* 23, 9 (2012), 1390–1406.

[6] Farshid Anvari, Deborah Richards, Michael Hitchens, Muhammad Ali Babar, Hien Minh Thi Tran, and Peter Busch. 2017. An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software* 134, (December 2017), 324–339. https://doi.org/10.1016/j.jss.2017.09.020

[7] John W. Ayers, Adam Poliak, Mark Dredze, Eric C. Leas, Zechariah Zhu, Jessica B. Kelley, Dennis J. Faix, Aaron M. Goodman, Christopher A. Longhurst, Michael Hogarth, and Davey M. Smith. 2023. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med* 183, 6 (June 2023), 589. https://doi.org/10.1001/jamainternmed.2023.1838

[8] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *HCOMP* 7, (October 2019), 2–11. https://doi.org/10.1609/hcomp.v7i1.5285

[9] Mohamed Amine Belabbes, Ian Ruthven, Yashar Moshfeghi, and Diane Rasmussen Pennington. 2023. Information overload: a concept analysis. *JD* 79, 1 (January 2023), 144–159. https://doi.org/10.1108/jd-06-2021-0118

[10] Izak Benbasat. 2010. HCI research: Future challenges and directions. *AIS Transactions on Human-Computer Interaction* 2, 2 (2010), 1.

[11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (January 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[12] 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.

[13] Emma O. Brien and Aoife Chawke. 2022. Tell me a story: Using personas to enhance student engagement. *All Ireland Journal of Higher Education* 14, 1 (2022). Retrieved March 14, 2024 from https://ojs.aishe.org/index.php/aishe-j/article/view/595

[14] Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, March 17, 2019. ACM, Marina del Ray California. https://doi.org/10.1145/3301275.3302289

[15] Yufan Chen, Arjun Arunasalam, and Z. Berkay Celik. 2023. Can Large Language Models Provide Security & Privacy Advice? Measuring the Ability of LLMs to Refute Misconceptions. In *Annual Computer Security Applications Conference*, December 04, 2023. ACM, Austin TX USA. https://doi.org/10.1145/3627106.3627196

[16] Eric Chu, Prashanth Vijayaraghavan, and Deb Roy. 2018. Learning Personas from Dialogue with Attentive Memory Networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, October 2018. Association for Computational Linguistics, Brussels, Belgium, 2638–2646.

[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] Luigi De Angelis, Francesco Baglivo, Guglielmo Arzilli, Gaetano Pierpaolo Privitera, Paolo Ferragina, Alberto Eugenio Tozzi, and Caterina Rizzo. 2023. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Frontiers in Public Health* 11, (2023). Retrieved January 9, 2024 from https://www.frontiersin.org/articles/10.3389/fpubh.2023.1166120

[19] 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).

[20] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3 (January 2016), 554–559. https://doi.org/10.1073/pnas.1517441113

[21] Ameet Deshpande, Tanmay Rajpurohit, Karthik Narasimhan, and Ashwin Kalyan. 2023. Anthropomorphization of AI: Opportunities and Risks. *arXiv preprint arXiv: 2305.14784* (2023).

[22] Elizabeth Ditton, Anne Swinbourne, and Trina Myers. 2022. Selecting a clustering algorithm: A semi-automated hyperparameter tuning framework for effective persona development. *Array* 14, (2022), 100186.

[23] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. 2019. Explaining models: an empirical study of how explanations impact fairness judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, March 17, 2019. ACM, Marina del Ray California. https://doi.org/10.1145/3301275.3302310

[24] Kathleen M. Eisenhardt. 1989. Agency theory: An assessment and review. *Academy of management review* 14, 1 (1989), 57–74.

[25] Lauren Fell, Andrew Gibson, Peter Bruza, and Pamela Hoyte. 2020. Human Information Interaction and the Cognitive Predicting Theory of Trust. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, March 14, 2020. ACM, Vancouver BC Canada, 145–152. https://doi.org/10.1145/3343413.3377981

[26] Aaron M. French, Veda C. Storey, and Linda Wallace. 2023. The impact of cognitive biases on the believability of fake news. *European Journal of Information Systems* (November 2023), 1–22. https://doi.org/10.1080/0960085x.2023.2272608

[27] Joy Ai-Leen Goodman-Deane, Mike Bradley, Sam Waller, and P. John Clarkson. 2021. Developing personas to help designers to understand digital exclusion. *Proceedings of the Design Society* 1, (August 2021), 1203–1212. https://doi.org/10.1017/pds.2021.120

[28] Joy Goodman-Deane, Sam Waller, Dana Demin, Arantxa González-de-Heredia, Mike Bradley, and John P. Clarkson. 2018. Evaluating Inclusivity using Quantitative Personas. In *In the Proceedings of Design Research Society Conference 2018*, June 28, 2018. Limerick, Ireland. https://doi.org/10.21606/drs.2018.400

[29] Jeff Gothelf. 2012. Using proto-personas for executive alignment. *UX Magazine* (2012), Article No: 821.

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

[31] 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.

[32] Brian W. Haas, Alexandra Ishak, Ian W. Anderson, and Megan M. Filkowski. 2015. The tendency to trust is reflected in human brain structure. *NeuroImage* 107, (February 2015), 175–181. https://doi.org/10.1016/j.neuroimage.2014.11.060

[33] Philipp Hacker, Brent Mittelstadt, Frederik Zuiderveen Borgesius, and Sandra Wachter. 2024. Generative Discrimination: What Happens When Generative AI Exhibits Bias, and What Can Be Done About It. Retrieved October 27, 2024 from

http://arxiv.org/abs/2407.10329
[34] Awni Hannun, Chuan Guo, and Laurens van der Maaten. 2021. Measuring data leakage in machine-learning models with fisher information. In *Uncertainty in Artificial Intelligence*, 2021. PMLR, 760–770. Retrieved March 14, 2024 from https://proceedings.mlr.press/v161/hannun21a.html
[35] Andreas Holzinger, Michaela Kargl, Bettina Kipperer, Peter Regitnig, Markus Plass, and Heimo Müller. 2022. Personas for Artificial Intelligence (AI) an Open Source Toolbox. *IEEE Access* 10, (2022), 23732–23747. https://doi.org/10.1109/ACCESS.2022.3154776
[36] Sungsoo Ray Hong, Jessica Hullman, and Enrico Bertini. 2020. Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 1–26. https://doi.org/10.1145/3392878
[37] Pei-Fang Hsu, Yu-Han Lu, Shih-Chu Chen, and Patricia Pei-Yi Kuo. 2024. Creating and validating predictive personas for target marketing. *International Journal of Human-Computer Studies* 181, (January 2024), 103147. https://doi.org/10.1016/j.ijhcs.2023.103147
[38] Bernard J. Jansen, Soon-Gyo Jung, Lene Nielsen, Kathleen W. Guan, and Joni 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 (2022). https://doi.org/10.17705/1pais.14301
[39] 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
[40] Bernard J. Jansen, Joni Salminen, and Soon-gyo Jung. 2020. Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics. *Data and Information Management* 4, 1 (2020), 1–17. https://doi.org/10.2478/dim-2020-0005
[41] 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
[42] 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.
[43] 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*, September 20, 2023. ACM, Torino Italy, 1–12. https://doi.org/10.1145/3605390.3605397
[44] 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
[45] 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* 178, (October 2023), 103096. https://doi.org/10.1016/j.ijhcs.2023.103096
[46] Keith Kirkpatrick. 2016. Battling algorithmic bias: How do we ensure algorithms treat us fairly? *Communications of the ACM* 59, 10 (2016), 16–17.
[47] 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
[48] Diana Kwon. 2024. AI-generated images threaten science — here's how researchers hope to spot them. *Nature* 635, 8038 (November 2024), 268–269. https://doi.org/10.1038/d41586-024-03542-8
[49] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science* 359, 6380 (March 2018), 1094–1096. https://doi.org/10.1126/science.aao2998
[50] Stephan Lewandowsky, Ullrich K.H. Ecker, John Cook, Sander Van Der Linden, Jon Roozenbeek, and Naomi Oreskes. 2023. Misinformation and the epistemic integrity of democracy. *Current Opinion in Psychology* 54, (December 2023), 101711. https://doi.org/10.1016/j.copsyc.2023.101711
[51] Pongsakorn Limna, Tanpat Kraiwanit, Kris Jangjarat, Prapasiri Klayklung, and Piyawatjana Chocksathaporn. 2023. The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching* 6, 1 (2023), 64–74.
[52] 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.
[53] Zhuoran Lu, Patrick Li, Weilong Wang, and Ming Yin. 2022. The Effects of AI-based Credibility Indicators on the Detection and Spread of Misinformation under Social Influence. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (November 2022), 1–27. https://doi.org/10.1145/3555562
[54] 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.
[55] Timothy R. McIntosh, Tong Liu, Teo Susnjak, Paul Watters, Alex Ng, and Malka N. Halgamuge. 2023. A Culturally Sensitive Test to Evaluate Nuanced GPT Hallucination. *IEEE Transactions on Artificial Intelligence* 1, 01 (November 2023), 1–13. https://doi.org/10.1109/TAI.2023.3332837
[56] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267, (February 2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007
[57] Lene Nielsen. 2019. *Personas - User Focused Design* (2nd ed. 2019 edition ed.). Springer, New York, NY.
[58] 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
[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] Stefano Nolfi. 2023. On the Unexpected Abilities of Large Language Models. https://doi.org/10.48550/arXiv.2308.09720
[62] OpenAI. 2022. Introducing ChatGPT | OpenAI. Retrieved October 20, 2024 from https://openai.com/index/chatgpt/
[63] 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
[64] Peter Gordon Roetzel. 2019. Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Bus Res* 12, 2 (December 2019), 479–522. https://doi.org/10.1007/s40685-018-0069-z
[65] Joni Salminen, Kathleen Guan, Soon-gyo Jung, Shammur Absar Chowdhury, and Bernard J. Jansen. 2020. A Literature Review of Quantitative Persona Creation. In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, April 25, 2020. ACM, Honolulu, Hawaii, USA, 1–14. https://doi.org/10.1145/3313831.3376502
[66] Joni Salminen, Kathleen Guan, Soon-Gyo Jung, and Bernard J. Jansen. 2021. A Survey of 15 Years of Data-Driven Persona Development. *International Journal of Human–Computer Interaction* (2021), 1–24.
[67] Joni Salminen, Bernard J. Jansen, Jisun An, Haewoon Kwak, and Soon-gyo Jung. 2018. Are personas done? Evaluating their usefulness in the age of digital analytics. *Pers Stud* 4, 2 (November 2018), 47–65. https://doi.org/10.21153/psj2018vol4no2art737
[68] 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*, July 04, 2022. ACM, Barcelona Spain, 67–73. https://doi.org/10.1145/3511047.3536403
[69] 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
[70] 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
[71] Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2021. Are Data-Driven Personas Considered Harmful?: Diversifying user understandings with more than algorithms. *Persona Studies* 7, 1 (2021), 48–63.
[72] 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 ACM Human Factors in Computing Systems (CHI'21)*, 2021. ACM, Virtual conference, 1–6. https://doi.org/10.1145/3411764.3445360
[73] Katharina Schäfer, Peter Rasche, Christina Bröhl, Sabine Theis, Laura Barton, Christopher Brandl, Matthias Wille, Verena Nitsch, and Alexander Mertens. 2019. Survey-based personas for a target-group-specific consideration of elderly end users of information and communication systems in the German health-care sector. *International Journal of Medical Informatics* 132, (December 2019), 103924.

https://doi.org/10.1016/j.ijmedinf.2019.07.003

[74] Cathrine Seidelin, A. Jonsson, M. Høgild, J. Rømer, and P. Diekmann. 2014. Implementing Personas for International Markets: A Question of UX Maturity. *Proceedings at SIDER'14Royal Institute of Technology, KTH, Stockholm, Sweden* (2014).

[75] Donghee Shin, Amy Koerber, and Joon Soo Lim. 2024. Impact of misinformation from generative AI on user information processing: How people understand misinformation from generative AI. *New Media & Society* (March 2024), 14614448241234040. https://doi.org/10.1177/14614448241234040

[76] Joongi Shin, Michael A. Hedderich, Bartłomiej Jakub Rey, Andrés Lucero, and Antti Oulasvirta. 2024. Understanding Human-AI Workflows for Generating Personas. In *Designing Interactive Systems Conference*, July 2024. ACM, IT University of Copenhagen Denmark, 757–781. https://doi.org/10.1145/3643834.3660729

[77] Ben Shneiderman. 2022. *Human-centered AI*. Oxford University Press.

[78] Ben Shneiderman and Catherine Plaisant. 2010. *Designing the user interface: strategies for effective human-computer interaction* (5th ed ed.). Addison-Wesley, Boston.

[79] Paul Thomas, Mary Czerwinksi, Daniel Mcduff, and Nick Craswell. 2021. Theories of Conversation for Conversational IR. *ACM Trans. Inf. Syst.* 39, 4 (August 2021), 39:1-39:23. https://doi.org/10.1145/3439869

[80] Krzysztof Wach, Cong Doanh Duong, Joanna Ejdys, Rūta Kazlauskaitė, Pawel Korzynski, Grzegorz Mazurek, Joanna Paliszkiewicz, and Ewa Ziemba. 2023. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *EBER* 11, 2 (2023), 7–30. https://doi.org/10.15678/EBER.2023.110201

[81] Linyao Yang, Hongyang Chen, Zhao Li, Xiao Ding, and Xindong Wu. 2024. Give Us the Facts: Enhancing Large Language Models with Knowledge Graphs for Fact-aware Language Modeling. https://doi.org/10.48550/arXiv.2306.11489

[82] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2019. BERTScore: Evaluating Text Generation with BERT. https://doi.org/10.48550/ARXIV.1904.09675

[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] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, May 11, 2024. ACM, Honolulu HI USA, 1–26. https://doi.org/10.1145/3613904.3642385

[86] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, April 19, 2023. ACM, Hamburg Germany, 1–20. https://doi.org/10.1145/3544548.3581318

[87] Michelle X. Zhou, Wenxi Chen, Ziang Xiao, Huahai Yang, Tracy Chi, and Ransom 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*, 2019. ACM, 123–124.

[88] Xinyi Zhou and Reza Zafarani. 2018. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. (2018). https://doi.org/10.48550/ARXIV.1812.00315