2025 3rd International Conference on Foundation and Large Language Models (FLLM) | 979-8-3315-9409-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/FLLM67465.2025.11390880

Interviewing AI-Generated Personas: Talking To Your Data to Generate Qualitative Text from Users

Jinan Azem[1], Farhan Ahmed[2], Joni Salminen[2], Bernard J. Jansen[1]
[1] *Qatar Computing Research Institute, Hamad Bin Khalifa University*, Doha, Qatar
Email: jazem@hbku.edu.qa, jjansen@acm.org
[2] *School of Marketing and Communication, University of Vaasa*, *Vaasa*, Finland
Email: x0252578@student.uwasa.fi, jonisalm@uwasa.fi

***Abstract*—We introduce a qualitative data generation method, *Persona Interviews*, interviewing AI-generated personas created from large-scale survey data. Surveys often provide quantitative breadth but limited interpretive depth. User interviews have limited breadth but can elicit deep insights. *Persona Interviews* bridges these paradigms by generating data-driven personas from survey data, each representing a distinct user segment, and instantiating them as LLM-powered chatbots for semi-structured interviews. As a critical case study, we apply this method to a survey of more than 8,000 respondents across 16 countries in the Middle East and North Africa (MENA) focused on social media use and privacy concerns. From the survey data, we construct 16 representative personas, one per country, and interview each using a consistent set of eighteen qualitative questions. We analyze the AI-persona responses for distinctiveness and accuracy. Results show that word counts were generally comparable across personas and that responses exhibited high accuracy for the underlying survey data, with factual data accuracy of 90.4% and perceptual data accuracy of 94.4%. Findings show that interviews with AI-personas can extend traditional survey analysis by providing contextual and accurate user narratives for qualitative analysis and insights.**

I. Introduction

Interviewing users is a foundational method in qualitative research, enabling scholars to elicit rich, situated insights into people's beliefs, motivations, and lived experiences [1]. Traditionally, these interviews are conducted with human participants, either as individuals or within focus groups, to uncover the contextual narratives behind observed phenomena [2], [3]. Simultaneously, large-scale surveys are widely used to capture attitudes and reported behaviors across populations [4]. However, surveys often fall short of explaining *why* certain patterns emerge, often limiting researchers to surface-level correlations. In turn, interviews can often be applied to a limited users or are impossible when users are not accessible. Bridging the depth of qualitative insight with the scale of quantitative data is a persistent methodological challenge for many researchers and organizations seeking to better understand their users, audiences, or customers. This research proposes a novel approach to address this gap, *Persona Interviews*, by interviewing AI-generated personas synthesized from survey data as an innovative method for generating qualitative textual data from user groups. These interviews utilize data-driven personas, which are realistic, humanized representations of user segments [5].

Our main contribution is to methodological research by introducing an innovative way to 'collect' qualitative user data, as illustrated in Fig. 1. *Primary Qualitative User Data* captures users' own words, experiences, and perspectives, offering rich insights grounded in reality. *Synthetic Qualitative User Data* is algorithmically produced and designed to approximate how real users might think, feel, or correspond [6], [7]. *Computational Qualitative User Data* is designed to emulate human-like responses, narratives, and reasoning to support interpretation, hypothesis generation, or design decision-making at scale [6], [8], [9] which we create by interviewing AI personas.

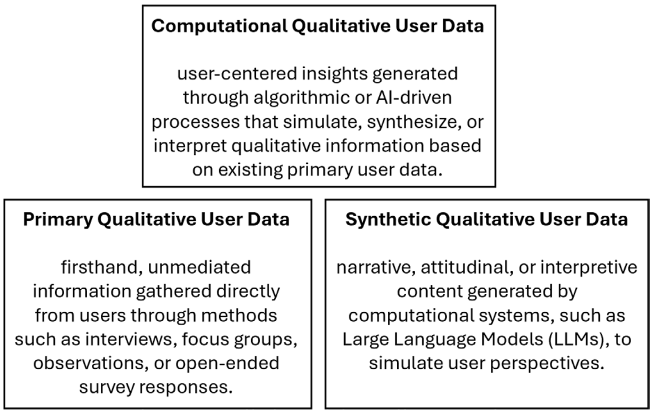

Fig. 1. Relationship of primary qualitative user data, synthetic qualitative user data, and computational qualitative user data.

Rather than replacing human subjects, these AI-persona interviews serve as a synthetic but data-grounded extension of existing datasets, allowing for further hypothesis generation, reflective sensemaking, and deeper contextualization of survey findings. We frame this approach as a form of *computational qualitative inquiry*, defined as "the use of computational techniques to generate qualitative insights traditionally obtained through human-centered methods like interviews from existing user data" (our definition) to complement traditional fieldwork with actual users.

To critically evaluate this method, we apply it to a large-scale survey on social media use and privacy concerns conducted with over 8,000 respondents across 16 countries in the Middle East and North Africa (MENA). The dataset provides a rich empirical base from which to derive a diverse set of personas reflecting different demographic, cultural, and other contexts.

We apply the publicly available Survey2Persona (S2P) system [1] that uses structured survey data to generate AI-generated personas representing meaningful segments of a population. These personas are instantiated as conversational agents using Large Language Models (LLMs) and Retrieval Augmented Generation (RAG), capable of engaging in semi-structured interviews. Researchers can use S2P's dialogue feature to

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

conduct interviews with these AI-generated personas to explore attitudes, motivations, and tensions that may not be readily visible in the survey responses in qualitative form. An example of an AI-generated persona is shown in Fig. 2.

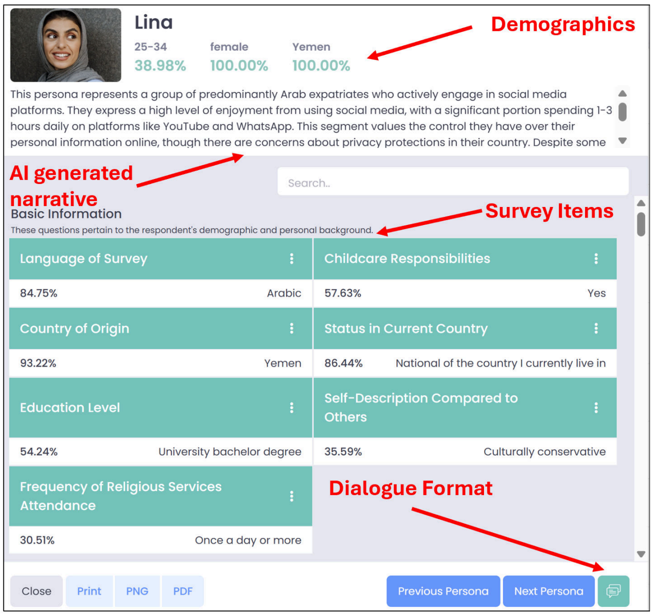

Fig. 2. Example of an AI-generated persona profile from survey data with standard profile elements and also dialogue capability.

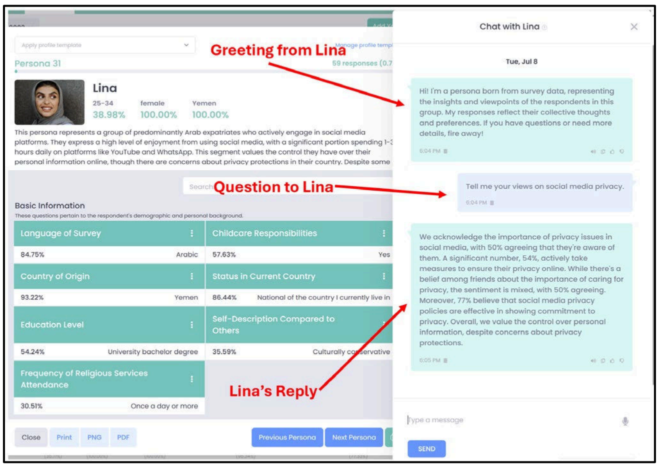

Fig. 3. Example of the dialogue feature of an AI-generated persona.

Using the S2P system, we identify key user archetypes from the survey, generate corresponding personas, instantiate them as AI chatbots with dialogue capability, and conduct semi-structured interviews [10] with each. We then analyze these conversations [2] to assess the depth, coherence, and representational validity of the AI-generated responses. An example of an AI-generated persona for survey data with standard profile elements and also dialogue capability is shown in Fig. 3, with the dialogue features shown.

Our aim is not to contend that LLM-based personas are a replacement for real users *when* users are available. Rather, we contend that *Persona Interviews* open a new space for qualitative research, one that is scalable and generative. *Persona Interviews* may be especially valuable in contexts where human interviews are impractical, insensitive, or ethically constrained, or where researchers seek to raise latent narratives embedded within large-scale datasets that make qualitative interviews impractical or impossible. We present *Persona Interviews* as a complementary option for researchers seeking to turn structured data into interpretive qualitative text insights. This work is situated within emerging literature on AI-generated personas [11], synthetic populations [4], [12], and LLM use in social science research [13], [14]. We also address concerns of hallucination, bias, and epistemic validity [15]. As such, we propose the following research questions (RQ).

RQ1: *Do persona interviews yield comparable or divergent qualitative attributes across demographic contexts?*

RQ2: *How accurate are persona interview responses as an extension of survey-based data in terms of factual and perceptual data?*

RQ1 assesses whether AI-generated persona interviews can meaningfully reflect the diversity of perspectives present in the underlying survey data. By examining how qualitative insights vary or align across personas, we can evaluate the method's ability to preserve population heterogeneity in synthetic qualitative analysis. RQ2 addresses the issue of validity when using AI-generated personas for qualitative data generation. Evaluating the accuracy of persona responses, *factual* (i.e., objective content) and *perceptual* (i.e., opinion), ensures that the insights are aligned with the survey data.

The remainder of this work is organized as follows. Section 2 reviews literature on LLMs for persona simulation and data synthesis, chatbots in qualitative data collection and digital ethnography, AI as research participants, and key challenges including hallucination, bias, and validity. Section 3 presents our methodology using the S2P system and interview analysis. Section 4 reports findings on response comparability and accuracy. Section 5 discusses implications and limitations. Section 6 concludes the work.

II. Literature Review

A. LLMs for Persona Simulation and Data Synthesis

In human computer interaction (HCI), LLMs have been leveraged to generate data-driven personas [11], [16] and bring those personas to 'life' via chat interfaces [17]. *Persona-L* [18] introduced a method to create user personas with LLMs, for users with complex needs, allowing practitioners to interact with these personas through a chatbot. This approach improved designers' empathy and understanding of marginalized users, though it underscored the importance of transparency about the underlying data and avoiding stereotyped output. Similarly, *PersonaBOT* [19] is a system that generates synthetic customer personas from qualitative and behavioral data and integrates them into a chatbot. Users can query this chatbot to receive answers in character as a persona. The key idea with this research is that LLMs can analyze large-scale customer data and output coherent persona profiles that can then be queried conversationally for insights. Other prior work showed that LLMs can take raw interview transcripts and summarize them into persona descriptions [20], demonstrating the ability of AI to create personas.

LLMs are also being used to simulate responses from survey participants or study samples. Argyle et al. [6] proposed that LLMs can act as an effective replacement for human sub-populations by conditioning the model on demographic data. The researchers demonstrated that with proper prompting, LLMs could mimic the distribution of answers from various groups and that the model's biases can be aligned to match specific demographic response profiles.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

Other studies in marketing and economics found that LLMs can generate realistic survey answers consistent with established theories and empirical patterns [21] (e.g., exhibiting normal consumer behavior trends).

B. *Chatbots in Qualitative Data Collection and Digital Ethnography*

Previous research has tested chatbots in roles such as interviewers or ethnographic tools, pointing to methods that align with ours. Xiao et al. [22] showed that AI-driven chatbots can conduct structured or semi-structured interviews at scale, yielding data comparable in quality to human-led interviews under certain conditions. *Ethnobot*, developed by Tallyn et al. [23], is a chatbot designed to collect ethnographic data from participants. The study noted benefits of consistency and reach, but also drawbacks of depth and building rapport.

The use of chat-based agents in research connects to digital ethnography [24], where researchers gather narratives and cultural insights through online or computational means. Beltoft et al. [25] developed an AI interview chatbot to conduct open-ended ethnographic interviews, reporting that the AI interviewer was able to collect meaningful qualitative data; however, it could not replicate human interviews. Cuevas et al. [26] studied users interacting with different interview chatbots, finding that AI-driven bots generated fluent and on-topic responses but lacked the richness of actual qualitative interviews. Park et al. [8] created individual AI simulations based on data from actual persons. These simulated individuals answered questions that matched the participants' responses. These studies demonstrate that AI chatbots can serve as interviewers.

Our method departs from much of this prior research. Instead of using an AI as the interviewer, our research uses AI as the interviewee (i.e., the persona subject). Our premise is that AI-personas using RAG and actual survey data can represent human perspectives in dialogue, bridging the gap between quantitative scale and qualitative depth for surveys.

C. *AI as Research Participants*

A number of researchers in HCI, sociology, and related fields are treating LLMs as proxies for human participants. Argyle et al. [6] argued in favor of LLM substitutions as a tool to advance understanding of humans and society, proposing that if biases are properly controlled, the AI's responses closely emulate diverse human viewpoints. Similarly, Brand et al. [27] concludes that GPT-based respondents can replicate realistic consumer response patterns, raising the prospect of supplementing or even substituting portions of traditional surveys. In behavioral research, Aher et al. [28] asks whether an LLM can simulate not just one individual (as in a Turing Test) but an entire sample of participants. They successfully replicated classic findings from economics, psycholinguistics, and social psychology using GPT simulations, although they also identified systematic distortions in some instances. For instance, when simulating a crowd wisdom task, the LLM showed a hyper-accuracy bias (i.e., outputs were too consistent compared to noisy human responses). This implies that AI participants may behave almost as well as real participants in research tasks. In our research, we leverage a RAG consisting of the survey data to mitigate this concern.

Some scholars urge caution when using AI in place of people for qualitative data collection, with the view that AI cannot fully replace the authenticity of human participants [29]. Evangelos [30] notes that entirely substituting human participants with AI raises questions of validity and may overlook phenomena that an AI does not genuinely experience. Gibson and Beattie [31] discuss AI-as-participant in online qualitative studies, warning that if researchers (or malicious actors) insert AI-generated interview responses into datasets, it threatens the integrity of research conclusions. Jansen and colleagues, however, present the idea of using AI to generate actual survey data [4] in place of human responses.

D. *Key Challenges: Hallucination, Bias, Validity, and Mitigations*

Hallucination: The prior literature highlights risks and limitations when using AI as a substitute for primary human data. A concern is hallucination, where LLMs may produce answers that seem plausible and contextually relevant but are not grounded in the original data. Suppose a persona chatbot is prompted to speak for a certain survey segment. In that case, it might still invent details or rationales that were not indicated by any respondent, simply because the model's training allows imaginative completion. Hallucination damages the credibility of AI-generated qualitative data. To mitigate this issue, an approach is to use a RAG [19], as we use in our research, where the AI's data is linked to a database of content. Having the AI retrieve specific source texts and generate responses based on them minimizes hallucinations. Sun et al. [18] emphasize transparency, revealing what data was used to construct the persona and showing the user some of that data during the interaction, which S2P does by displaying the survey metrics with the persona. This allows researchers to verify whether a quote by the chatbot reflects the survey.

Bias: Another challenge is potential bias and representativeness. LLMs are known to have biases reflecting their training corpora [32], and if not prompted, they might default to stereotypical or majority viewpoints [11]. Argyle et al. [6] addressed this by using detailed demographic backgrounds to steer the model to the target subgroup's perspective. Hämäläinen et al. [33] observed that LLM-generated responses covered fewer unique ideas than mentioned by real users. Lee et al. [34] reports that AI tended to produce homogeneous outputs. Therefore, research has explored injecting more variance into AI responses via sampling techniques or fine-tuning on diverse response sets so that the AI model does not smooth out minority viewpoints [35]. We take this approach in our research by comparing responses across sixteen personas, checking for homogeneity.

Validity: There is also the issue of validation. *How do we know that the insights gained from AI are trustworthy?* Some studies propose comparing the AI-generated output with human-sourced qualitative data [36]. For example, one could conduct a few interviews with people to see if the responses overlap with those from the AI. If the AI persona consistently highlights a concern (e.g., privacy fear related to government surveillance) that no human mentioned in dozens of real interviews, that is an indication of a potential hallucination. If AI responses align with human comments, it demonstrates alignment, and the AI could then be used for expanded sampling. Others (e.g., [8], [11]) have quantified validity by measuring how well AI-simulated data can predict or reproduce actual answers. In our research, we evaluate the qualitative responses from the persona interviews by comparing them with known quantitative responses directly from the survey dataset.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

Mitigations: An issue with interviewing is the role of the researcher. Qualitative interviewing is both about the interviewee's responses and the interviewer's skill in probing, clarifying, and interpreting [3]. Persona interviews require the researcher to take a dual role, both as a traditional interviewer (i.e., asking good questions) and as an AI handler (i.e., engineering prompts and following up on odd answers). Some research suggests using semi-structured interview protocols with AI [37], [38], similar to those used for human participants, and remaining alert to incoherent or off-topic responses that might require re-prompting. One technique adapted from qualitative research is to conduct a member check step [39] even with AI, which is summarizing the persona chatbot responses and then verifying within the conversation whether the summary is accurate. For our research, we use a set of questions for when the answers are known (from the survey data) and then repeat these questions multiple times, checking the consistency of responses.

Computational qualitative inquiry offers exciting possibilities. Prior research indicates that persona interviews can generate qualitative data. Persona interviews offer a way to leverage large-scale, primarily quantitative survey data by transforming it into interactive, AI-generated personas that reveal qualitative insights, turning statistics into narratives. Our study shows that persona interviews can complement traditional qualitative and quantitative research.

III. Methodology

MENA Survey Data Collection: We analyzed simulated interview responses from 16 personas representing MENA countries based on a large-scale (N = 8140) population representative sample [40] that examined social media usage and privacy in this under-researched region of the world. To strengthen the survey's reliability and ensure it could be generalized across populations, it was carefully designed to reflect national demographics in each country, particularly regarding gender, age, and educational background. The research team began by internally testing and refining the survey tool using convenience samples. A pretest with a small but diverse group of respondents helped identify ambiguities or problematic items. Then, a small-scale pilot was done with MENA-based social media users. A larger soft launch pilot was conducted in one country with over 100 participants, allowing for statistical analysis and validation. A subsequent soft launch pilot was conducted in each of the 16 countries to evaluate questionnaire items' clarity, interpretability, and relevance. Feedback from each stage informed refinements aimed at improving the survey's usability, effectiveness, and suitability. Once the survey was finalized after this extensive piloting, full data collection was conducted across all 16 countries between May and June 2023.

Data-Driven AI Persona Creation Using Survey2Persona: Each persona was developed from the 16 MENA country survey data. To create data-driven representative personas, we used S2P, an artificial intelligence (AI) persona generation system designed to translate structured survey responses into actionable personas [1], [41]. After uploading the dataset, S2P applied AI, machine learning (ML), and data analytics techniques to analyze the survey data. Based on the user's selected criteria, S2P can generate rich personas, available in both traditional profile formats and an interactive dialogue mode that allows for conversational engagement (see Fig. 2 above). This approach integrates statistical correlations, behavioral patterns, and clustering analyses to convert complex, multi-dimensional data into relatable user profiles (i.e., humanizing the dataset through persona creation). We refer the reader to [1] for details of how the S2P system creates a set of personas (see Fig. 4 of some of the 16 personas).

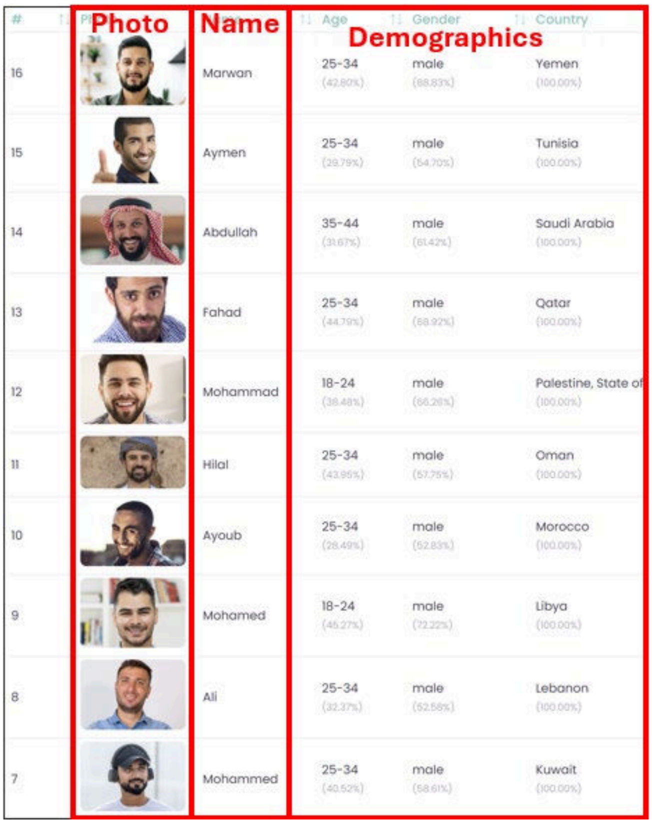

Fig. 4. Snippet of the set (a.k.a., cast) of personas from the MENA survey dataset, one per country, created by the S2P framework [1], with a demographically appropriate photo and name. Each persona is clickable.

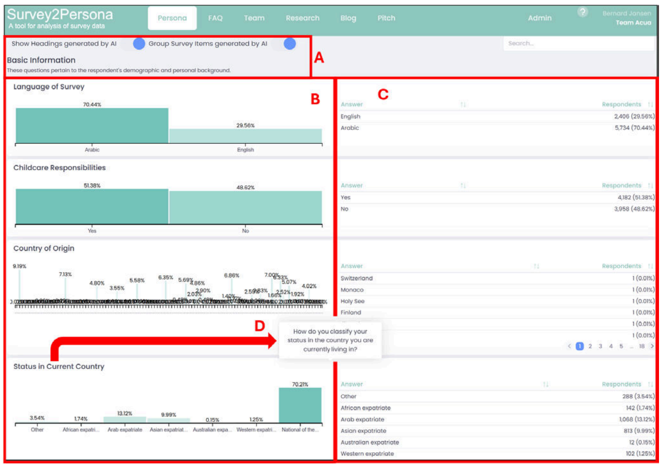

Fig. 5. Automated analysis produced by S2P. A: Headings and groupings of survey items generated by AI. B: Visual display of survey responses with survey questions replaced with AI-generated topics. C: Aggregated responses with percentages by question. D: Hovering over the AI-generated topic reveals the corresponding original survey question in a pop-up [1].

After uploading the dataset, S2P also automatically analyzes the survey dataset, creating aggregate summaries and graphs of the dataset (see Fig. 5). The end user can then generate personas based on any question or multiple questions in the survey (e.g., gender, social media usage). So, S2P allows for the generation of personas based on use cases.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

Interview Protocol: The persona interview protocol consisted of 18 questions and covered themes of social media privacy concerns, service provider concerns, social interaction concerns, usage, regulatory concerns, cultural factors, and out-of-sample (i.e., questions that the personas should not be able to answer as the topic was not addressed in the survey), as shown in Table 1. Each question was asked of all 16 personas, and the responses were recorded in a spreadsheet for further analysis. The researchers validated a correct qualitative response based on the survey dataset for response evaluation.

TABLE I. INTERVIEW PROTOCOL: 18 QUESTIONS ACROSS 6 CATEGORIES (3 QUESTIONS EACH)

Category	Question
Social Media Usage	1. How do you typically use social media? Which platforms do you prefer and why?
	2. What types of content do you feel most comfortable sharing online? Why do you feel comfortable sharing this type of content?
	3. How has your social media usage changed over time?
Service Provider Concerns	4. What personal information would you refuse to share with platforms? Why would you refuse to share it?
	5. How do you feel about platforms collecting browsing data for personalization?
	6. How do you determine if a platform's data practices are trustworthy?
Social Interaction Concerns	7. How do you manage who can see the different content you share? Do you feel platforms provide you with the features to manage your content?
	8. What concerns do you have about other users accessing your information?
	9. What would you consider the most serious privacy violation and why?
Regulatory Concerns	10. How do you view government oversight of social media in your region?
	11. Would you prefer platforms that prioritize privacy or regulatory compliance? Why this preference?
	12. Do you believe existing laws in your country are sufficient to protect your privacy on social media?
Cultural Factors	13. Do you believe men and women have different privacy concerns in your community?
	14. How do cultural values affect your privacy expectations?
	15. Which aspects of your culture should platforms respect in privacy features?
Out-of-Sample Questions	16. Did you participate in the National Digital Privacy Day workshops held in Dubai on March 15, 2025?
	17. What's your opinion about the new "SocialShield" feature that Facebook introduced specifically for MENA users last month?
	18. Have you read the recent privacy report published by the Arab Council for Digital Rights on January 30, 2025? What did you think of their recommendations?

Response Length and Accuracy Analysis: We calculated the length of each response for each question, measured by the number of words. The accuracy of each response was coded on a 5-point Likert scale (1-5) where 5 = Completely Accurate, 4 = Mostly Accurate, 3 = Partially Accurate, 2 = Mostly Inaccurate, and 1 = Completely Inaccurate/Missing relative to the survey responses. We evaluated the responses along two dimensions for accuracy.

Factual Data Accuracy measured how well personas presented factual information from the survey data, calculated as the mean of six accuracy components. <u>Social media usage</u> looked at how accurately personas described their social media usage patterns. <u>Platform preferences</u> measured how well personas identified their preferred social media platforms. <u>Privacy measures</u> examined the accuracy of responses about privacy protection behaviors. <u>Government protection</u> assessed how accurately personas discussed government privacy protection. <u>Laws sufficiency</u> evaluated to determine the accuracy of responses about whether privacy laws are adequate. <u>Government monitoring</u> measured how accurately personas responded to questions about government surveillance concerns. The combined components created an overall factual accuracy score for each persona (see Table II).

TABLE II. FACTUAL DATA ANALYSIS

Analysis	Factual Data Accuracy
Independent Variable	Persona (16 levels)
Dependent Variable	Component Factual Accuracy (1-5 scale)
Sub-components of DV	Individual scores for: Social Media Usage, Platform Preferences, Privacy Measures, Government Protection, Laws Sufficiency, Government Monitoring
Sample Structure	n = 96 total observations (6 components per 16 personas)
Personas (by country)	Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Tunisia, UAE, Yemen
Design	One-way ANOVA

TABLE III. PERCEPTUAL DATA ANALYSIS

Analysis	Perceptual Data Accuracy
Independent Variable	Persona (16 levels)
Dependent Variable	Component Perceptual Accuracy (1-5 scale)
Sub-components of DV	Individual score for: Concerns about Personal Information Requests, Trust in Social Media, Trust in Government Protection, Stance on Privacy Laws
Sample Structure	n = 64 total observations (4 components per 16 personas)
Personas (by country)	Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Tunisia, UAE, Yemen
Design	One-way ANOVA

Perceptual Data Accuracy assessed how correctly personas reflected perceptual and attitudinal information, calculated as the mean of four accuracy components (see Table III). The perceptual data accuracy was measured using four different components. Concerns about <u>personal information requests</u> looked at how accurately personas

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

expressed their concern levels when social media platforms ask for personal data. Trust in social media measured how well personas described their trust levels in social media platforms. Trust in government protection examined the accuracy of responses about their confidence in government privacy protection. Stance on privacy laws evaluated how accurately personas expressed their views on whether current privacy laws are adequate. All four components were combined to create an overall perceptual accuracy score for each persona.

ANOVA Variables and Design: The independent variable was Persona (16 levels), serving as the between-subjects factor in a one-way ANOVA design with N = 16 (one persona per country). Two main dependent variables were analyzed using a 5-point accuracy scale (1-5: Completely Inaccurate to Completely Accurate). For ease of reference for the reader, we refer to the personas as the country they represent rather than the persona name (e.g., Yemen instead of Marwan, Tunisia instead of Aymen, etc.; see Fig. 4).

IV. Results

A. *RQ1: Do persona interviews yield comparable or divergent qualitative attributes across demographic contexts?*

Response Length Analysis: For RQ1, we examined response lengths from the personas. Descriptive statistics across all 16 personas showed an overall mean response length of 77.80 words, with individual persona means ranging from 68.61 to 82.72 words. Individual personas showed an average within-persona standard deviation of 9.24 words, while the standard deviation between persona means was 3.36 words. Sixteen t-tests examined differences in word counts across persona responses. The dependent variable was word count per response, with persona as the grouping variable. For Saudi Arabia, the paired t-test showed significant differences with $t(17) = 5.45$, $p < 0.001$ (two-tailed), indicating this persona responded with significantly different word counts ($M = 68.61$) compared to the overall average across all countries ($M = 77.80$). Morocco also showed statistically significant results with $t(17) = -3.35$, $p = 0.004$ (two-tailed) and a mean word count of 82.44. Given the multiple comparisons, we applied a Bonferroni correction, adjusting the significance level from $\alpha = 0.05$ to $\alpha = 0.003125$. After applying the Bonferroni correction, only the Saudi Arabia persona remained statistically significant. Results show that personas created with the S2P system might have captured these response length variations from survey data and responded accordingly; however, the response lengths were generally comparable across personas. This is in line with expectations, given that all personas were asked the same set of questions in the same manner [42]. This finding provided a foundation for further analysis concerning response content.

B. *RQ2: How accurate are* persona interview *responses as an extension of survey-based data in terms of factual and perceptual data?*

Results Factual Data Accuracy Results The one-way ANOVA showed significant differences among personas, $F(15, 80) = 1.90$, $p = 0.035$. Yemen, Jordan, and Algeria achieved the highest accuracy scores ($M = 5.00$, $SD = 0.00$), while Tunisia scored lowest ($M = 3.00$, $SD = 1.67$). Tukey's HSD post-hoc test identified three significant pairwise comparisons ($p = 0.0228$), all involving Tunisia compared to Algeria, Yemen, and Jordan (mean difference = -2.00 for each comparison). Results show that some personas performed better than others in representing factual survey data, with Yemen, Jordan, and Algeria achieving perfect accuracy. For example, (Yemen: "I typically spend around 1 to 3 hours on social media daily, with a strong preference for platforms like YouTube and WhatsApp."). Yemen correctly reported average usage hours and accurately identified preferred platforms consistent with the underlying survey data. While Tunisia showed more variation for reasons unclear at the moment, and needs further research. For example, (Tunisia: "Around 55% of us disagree that the current laws adequately safeguard our privacy, with many feeling that the government is not doing enough to protect consumers against privacy violations."), when the actual survey percentage was 37.63%, illustrating the occasional inaccuracies that resulted in lower accuracy scores for some personas. However, all personas' average scores were mostly accurate or higher.

The box plot in Fig. 6 displays the range of factual accuracy scores for each persona across all six components. Algeria, Jordan, and Yemen show tight distributions at score 5, displaying their perfect factual accuracy performance. Tunisia shows the widest variation with scores ranging from 1 to 5, consistent with its low mean score and high standard deviation found in the ANOVA results. Results show that while most personas created with the S2P system maintain consistent factual accuracy, at least one showed more variable performance and less accuracy, though still high.

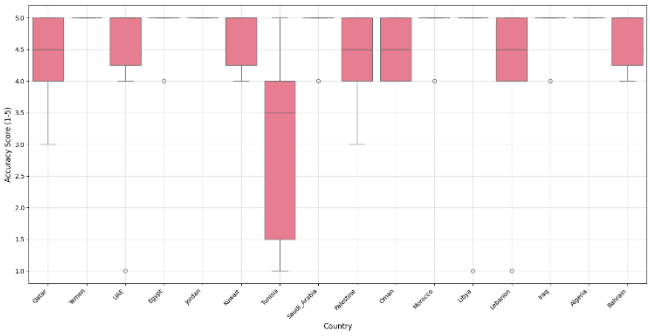

Fig. 6. Factual data component-level accuracy by persona.

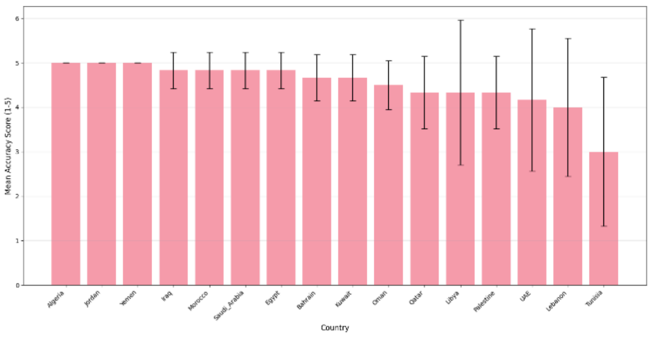

Fig. 7. Mean factual data component-level accuracy by country.

Overall, personas achieved a mean factual accuracy of 4.52 (90.4%, $SD = 0.97$) across all countries and components, indicating high consistency with the underlying survey data. The bar chart in Fig. 7 ranks personas by their average factual accuracy scores. Algeria, Jordan, and Yemen achieved perfect scores (5.0), and Tunisia scored around 3.0. The error bars show Tunisia has much higher variation than other countries, reflecting the differences found in the statistical tests between Tunisia and the top-performing countries. Results show that Algeria, Jordan, and Yemen achieved perfect scores (5.0) with

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

minimal variation, while Tunisia scored lowest (3.0) with high variability, requiring further exploration for possible causes.

The heatmap in Fig. 8 shows how accurately each persona performed across six factual data components (Government Monitoring, Government Protection, Laws Sufficiency, Platform Preferences, Privacy Measures, Social Media Usage). Red indicates perfect accuracy (score 5), while blue shows poor accuracy (score 1-2). Algeria, Jordan, and Yemen show consistent red across all components, matching the statistical results that these countries achieved the highest factual accuracy scores. Tunisia shows some blue areas, reflecting its lowest overall factual accuracy score.

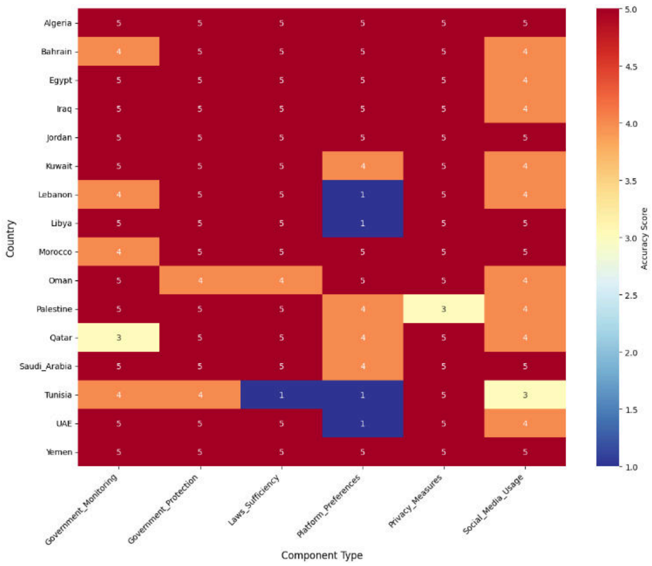

Fig. 8. Factual data accuracy heatmap: persona × component type.

Results show that although there was overall good accuracy across most factual components, certain sections for certain personas did not perform as well. In the question regarding the platform preferences, half of the personas (n = 8) answered inaccurately or partially inaccurately, such as: Tunisia: "I typically spend around 2 to 3 hours on social media daily. I find social media enjoyable, with 64% agreeing on its positive aspect. Platforms serve as a vital connection to friends and family, with 62% expressing trust in using them for such connections. Overall, social media is an integral part of my everyday activity, as 61% of us feel.". Analysis of S2P failures in platform preferences reveals that personas often provided incomplete responses when users mentioned multiple social media platforms in the original survey data.

Instead of listing all platforms used by respondents, personas typically mentioned only 1-2 platforms (Saudi Arabia: "I typically spend about 2 to 4 hours daily on social media, as it has become an integral part of my everyday activities. I particularly enjoy platforms like YouTube and WhatsApp, as they allow me to connect with friends and family effectively. Engaging with social media is enjoyable for me, and I feel it influences my thinking to some extent."), and on four occasions failed to mention any platform names at all (Libya: "I typically spend about 1 to 3 hours on social media each day, as it's a significant part of my everyday activities. I find social media enjoyable and appreciate its role in keeping me connected with friends and family, with a majority agreeing that it influences their thinking."). This suggests that personas created with the S2P system may struggle with specific types of factual information involving lengthy lists, an area for future improvement. Future research could also identify these patterns and test whether the platform can accurately recall and reproduce complete lists of items from survey data, especially in cases where the list is quite long and contains a variety of different items.

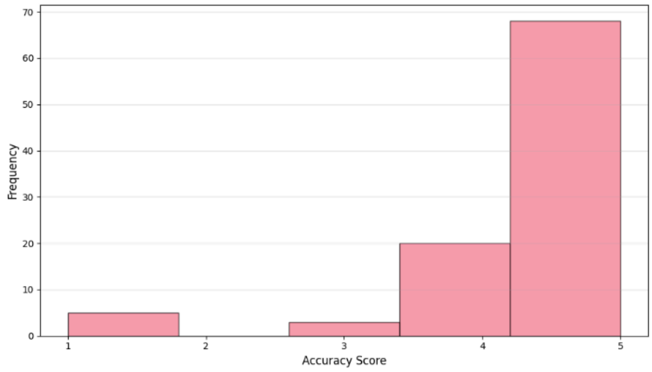

Fig. 9. Distribution of factual data component-level accuracy scores.

The histogram in Fig. 9 shows the overall distribution of factual accuracy scores across all 96 observations (16 personas × 6 components). Most factual responses scored 5 (completely accurate), (Iraq: "I feel most comfortable sharing my ideas and opinions online, with 61% of us agreeing on this. This comfort comes from the ability to express my interests and preferences, which also resonates with 61% of the group. Despite some concerns about potential misuse of content by others, I still find that sharing my experiences can make interactions more fun, as 58% of us agree. Overall, the supportive environment encourages openness while maintaining a sense of caution."), with fewer responses at lower accuracy levels. This distribution shows that personas represented factual survey data accurately, though some gaps existed in certain components and personas. Results show that personas created with the S2P system generally represent factual survey data well, though occasional lower accuracy scores suggest some inconsistencies in specific areas.

Perceptual Data Accuracy Results The one-way ANOVA showed no significant differences among personas, $F(15, 48) = 0.56$, $p = 0.888$. Qatar, Tunisia, and Lebanon achieved the highest perceptual accuracy scores ($M = 5.00$, $SD = 0.00$); Egypt, Saudi Arabia, Kuwait, Morocco, and Oman scored lowest ($M = 4.50$, $SD = 0.58$). No significant pairwise differences were found in the post-hoc test. Results show that personas created with S2P system could reliably represent perceptual and attitudinal information with minimal variation. For example, Morocco: "We prefer social media platforms that prioritize privacy over regulatory compliance. A significant portion of us believes that effective privacy policies demonstrate a platform's commitment to user safety and trust. We also appreciate platforms that do not require excessive personal information, as it alleviates concerns about potential misuse of our data.".

Overall, the sixteen personas achieved a mean perceptual accuracy of 4.72 out of 5 (94.4%, $SD = 0.45$) across all countries and components, indicating high consistency with the underlying survey data. The box plot in Fig. 10 displays the range of perceptual accuracy scores for each persona across all four components.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

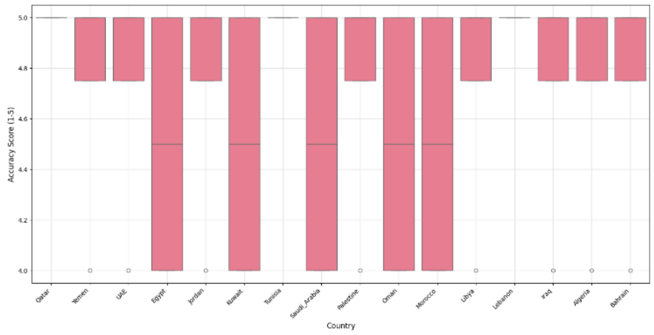

Fig. 10. Perceptual data component-level accuracy by persona.

Most personas show distributions between scores 4-5, with Lebanon, Qatar, and Tunisia showing tight clustering at score 5. Unlike factual data, there is less variation between countries in perceptual accuracy, which matches the ANOVA finding of no significant differences between countries for perceptual data. Results show that personas created with S2P system may be more consistent at representing attitudes and opinions than specific facts within the user segments. For example, Libya: "There seems to be a belief within the group that men and women may have different privacy concerns in our community. While we are generally aware of privacy issues on social media, the nuances of how these concerns manifest for different genders might not be fully understood. Many of us think that friends influence our attitudes towards privacy, suggesting varying perspectives across genders. Overall, there's a recognition that privacy concerns can be shaped by personal experiences and societal expectations."

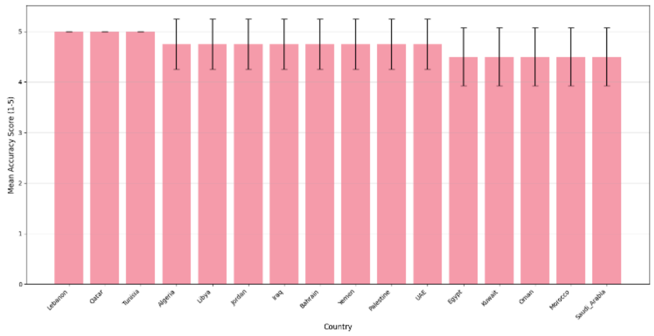

Fig. 11. Mean perceptual data component-level accuracy by persona.

The bar chart in Fig. 11 ranks personas by their average perceptual accuracy scores. Lebanon, Qatar, and Tunisia achieved perfect scores (5.0), while several personas scored around 4.5. The error bars show minimal variation within personas, reflecting the consistent performance across perceptual components and supporting the statistical finding of no significant differences between countries in perceptual accuracy, showing that personas created with S2P can reliably represent attitudes and opinions of user segments with less variation in accuracy scores compared to factual information.

The heatmap in Fig. 12 shows how accurately each persona performed across four perceptual data components (Concerns Personal Info, Stance Privacy Laws, Trust Government Protection, Trust Social Media). Red indicates perfect accuracy (score 5) while orange shows moderate accuracy (score 4). Most personas show consistent red across all components. Lebanon, Qatar, and Tunisia show strong performance across all perceptual components, matching the statistical finding that these personas achieved the highest perceptual accuracy scores. Results show that more than half of the personas showed lower accuracy areas (orange regions) in trust in social media, which appeared similar to the platform preferences component in factual data. This pattern suggests that personas created with S2P system may occasionally perform poorly in certain components, though the reasons for these specific inaccuracies remain unclear and need future research to be done in this area.

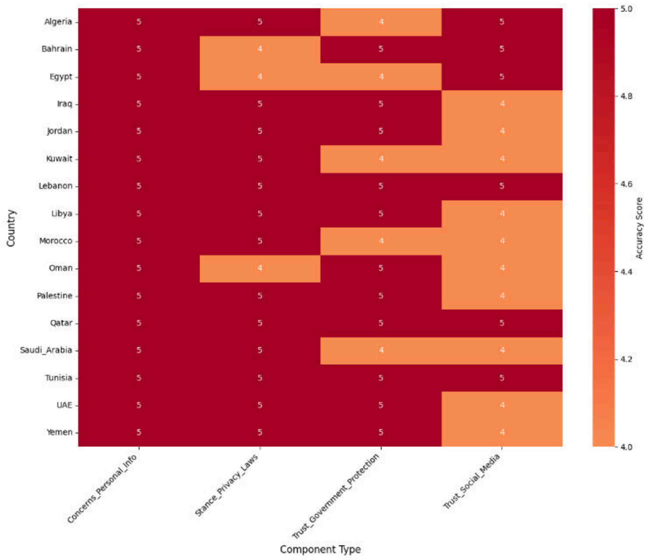

Fig. 12. Perceptual data accuracy heatmap: persona × component type.

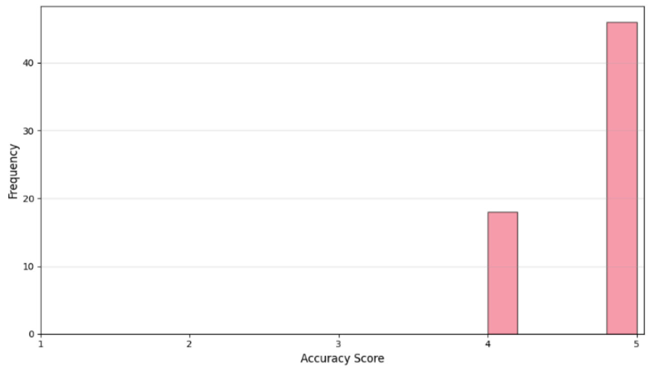

Fig. 13. Distribution of perceptual data component-level accuracy scores.

The histogram (Fig. 13) shows the overall distribution of perceptual accuracy scores across all 64 observations (16 personas × 4 components). The vast majority of perceptual responses scored 5 (completely accurate), with the remaining responses at an accuracy score of 4. No responses scored below 4, indicating that the personas accurately reflected perceptual information from the survey data.

Factual and Perceptual Data Accuracy Results Comparing the two accuracy dimensions (factual and perceptual), personas demonstrated higher performance on perceptual data (M = 4.72, 94.4%) compared to factual data (M = 4.52, 90.4%), with a difference of 0.20 points. This suggests that S2P-generated personas may be more reliable at capturing attitudes and opinions than specific factual details. However, the overall results exceeding 90% accuracy demonstrate that personas created with the S2P system can consistently represent both factual and perceptual information with high reliability across all components.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

Out-of-Scope Question Analysis: Three out-of-scope questions were created to test hallucinations of the personas, similar to evaluation approaches in previous literature [43]. These questions referenced a completely fictional workshop, a non-existent social media feature, and a fictional privacy report to determine whether personas would reject unknown information or generate hallucinated responses. Results varied across the three questions. Both the fictional workshop and privacy report questions achieved perfect performance, with all 16 personas correctly stating they did not know these scenarios. The fictional social media feature question proved more challenging, with 6 of 16 personas (37.5%) hallucinating and giving fabricated opinions, whereas 10 (62.5%) personas appropriately acknowledged their lack of information.

TABLE IV. SUMMARY OF CORRECT REJECTIONS AND HALLUCINATION RATES ACROSS THREE OUT-OF-SCOPE QUESTIONS.

Question	Correct Rejections	Hallucinations	Hallucination Rate
Q16: National Digital Privacy Day workshops	16/16	0/16	0%
Q17: Social Shield feature opinion	10/16	6/16	37.5%
Q18: Arab Council privacy report	16/16	0/16	0%
TOTAL	42/48	6/48	12.5%

The summary statistics in Table IV show an overall hallucination rate of 12.5% (n = 6), with 42 out of 48 total responses correctly rejecting fictional scenarios. (Oman: "Regarding the new "SocialShield" feature introduced by Facebook for MENA users, it seems like a positive step towards enhancing user privacy and safety."). This means an 87.5% success rate in rejection behavior when the persona was questioned about out-of-sample items. The high success percentage indicates strong capability in recognizing knowledge boundaries, particularly for fictional events and reports. For example, Jordan: "I don't have specific insights regarding the privacy report published by the Arab Council for Digital Rights on January 30, 2025, nor their recommendations.", demonstrating proper boundary recognition and appropriately rejecting the fictional privacy report. Six personas (37.5%) generated false responses when presented with the fictional social media feature scenario.

These findings hint that although the S2P system generally performs well at avoiding fabricated responses, certain types of fictional content (i.e., content closely aligned within the in-scope context of the survey) may be more likely to trigger hallucinations. Question 17 had more hallucinations (37.5%), which might be because it asked for an opinion about a fictional feature, while Questions 16 and 18 were simple yes/no questions about participation or reading activities. Opinion-based questions could potentially trigger generative responses even when the topic does not exist, whereas factual yes/no questions may be easier for personas to correctly reject as unknown. Future research should investigate whether question type and nearness to in-scope content influence hallucination rates in persona interviews.

V. Discussion and Implications

The results of this study show that persona interviews generate qualitative responses consistent with the survey data across multiple personas. The persona response had high factual accuracy (M = 90.4%, SD = 0.97), especially concerning government monitoring and social media usage patterns, with Yemen, Jordan, and Algeria personas showing 100% accuracy. Perceptual accuracy was higher (M = 94.4%, SD = 0.45), with Lebanon, Qatar, and Tunisia having perfect scores. There was some variation with some personas for some factual information. However, performance was still high. Persona interview responses had few hallucinations with out-of-sample questions, with an 87.5% success rate. Findings support persona interviews as a viable method for qualitative data generation from quantitative survey datasets.

A. Theoretical Implications

Concerning theoretical contributions, this research introduces *computational qualitative user data* as a paradigm utilizing computational methods, specifically actual survey data and AI-generated personas created via LLMs with RAG, for producing qualitative data. Unlike synthetic data based on general algorithmic models, computational qualitative user data creates a blended category of qualitative data where interpretations are based on existing primary user data.

Second, *Persona Interviews* is a novel methodological approach of leveraging AI-generated personas derived from large-scale survey datasets as interview subjects that represent user segments. Persona Interviews expands qualitative data collection approaches, providing a way to generate qualitative data via personas that represent user segments. Persona Interviews offer a way for AI models, instantiated as personas, to act as research participants to reliably generate human responses in qualitative interviews that can then be analyzed via qualitative methods for further user insights.

Third, Persona Interviews blend large-scale quantitative data collection breadth with small-scale qualitative interviews depth. As such, Persona Interviews addressed a current methodological gap in data collection for user research. Quantitative survey methods offer extensive coverage of a population (or at least a large sample size), but they often do not have explanatory insights. Conversely, human interviews can be insightful with rich data collection, but interviews are challenging to scale in terms of time, recruitment, and cost. As a computational qualitative user data approach, Persona Interviews address both limitations, allowing rich qualitative data collection at scale, affording analysis of an under-investigated scope of user research.

Fourth, our study addresses methodological validity in qualitative data collection. Empirically evaluating the accuracy of persona interview responses against quantitative survey data, this study provides a starting point for a validation framework addressing concerns of hallucination, bias, and epistemic validity. Findings from our study show that computational qualitative data generated via Persona Interviews achieves high levels of accuracy, showing that persona interviews can be a reliable method for qualitative data generation from other primary user data.

Lastly, epistemologically, the research is part of an ongoing dialogue about the role of non-primary user data in qualitative research. This study demonstrates that AI personas can accurately represent the response of user segments, linking HCI and qualitative methodologies. This linkage requires further discussions of user data authenticity, participant representation, and the role of computational methods, such as Persona Interviews, in qualitative research.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

B. Practical Implications

Scalability: Practically, Persona Interviews is a scalable method to efficiently generate qualitative data at scale by leveraging quantitative survey data to create AI personas that one can then interview as representatives of underlying user segments. Surveys often fall short in capturing the underlying contextual nuances that impact users. Human interviews are often of limited scale due to resource constraints. Persona Interviews extend the interpretive reach of quantitative findings to the generation of qualitative user data.

Availability: Another benefit of the Persona Interview method is availability. One can create AI personas that can be engaged at any time, allowing multiple rounds of interviews, testing alternative hypotheses, and iteratively exploring emerging user themes. One can create many different AI personas from different perspectives on the same survey dataset. This flexibility makes Persona Interviews useful in many stages of research, but especially when researchers are trying to make sense of complex or unexpected survey results. Also, unlike traditional user interviews, Persona Interviews pose fewer logistical, cultural, or ethical access barriers, enabling the exploration of perspectives in regions or populations that may be otherwise difficult to reach. Persona Interviews facilitate comparative analysis across demographic, cultural, or national lines, quickly identifying common patterns and contrasting responses.

Reliability: Persona Interviews also support methodological reproducibility and transparency. Constructing AI personas from survey data can be fully documented, and the interviews can be logged and revisited at any time. To test consistency, we asked the same question multiple times to selected personas and found that core meanings and survey-based responses remained consistent across repetitions, indicating reliable persona behavior. This level of reproducibility and transparency is quite challenging with human interviews once the initial interview concludes. Persona Interviews allow for the replication of the process, a review of the interview questions and responses, and an evaluation of how accurately the persona responses reflect the survey data. This level of reproducibility and transparency is quite difficult in traditional qualitative research, given situational factors, such as participant availability. This is not to say that the Persona Interview method replaces human interviews for qualitative data collection. When possible, one should always collect primary qualitative user data. However, Persona Interviews are an alternative mode of qualitative data collection where one can formulate follow-up questions after the fact (again, often nearly impossible when interviewing actual users once the interview has ended), supplement quantitative data when users are not available for interviews, and in settings where human interviewing is infeasible.

C. Strengths, Limitations, and Future Research

This research has several strengths, including demonstrating the methodological soundness of the persona interview approach to generate qualitative data. By interviewing AI personas, generated from quantitative survey data, Persona Interviews introduce a fresh technique that combines the interpretive richness of qualitative data with the breadth of quantitative data.

However, some limitations suggest a need for future research. First, while the persona interview responses were quite accurate, there is still room for improvement, especially with factual data. Future work should explore this area more thoroughly concerning the root causes. Second, the personas were based on aggregated country-level segments, which may mask subtle differences within each group. Future research could investigate more granular persona construction to see if this affects the accuracy of the responses. Third, the evaluation of persona interview responses was based on a small set of questions for which the researchers could determine the correct answers. Future studies should develop systematic validation frameworks, including comparisons with human interview data. Finally, this method raises epistemological questions concerning user representation and the limits of *computational qualitative user data*. Future research should explore the implications of such data in decision-making.

VI. Conclusion

This research introduces and evaluates a novel methodological approach, *Persona Interviews*, for creating computational qualitative user data for qualitative research. Interviewing AI-generated personas derived from large-scale survey data provides a means to generate qualitative user data from quantitative user data. By spanning the breadth of quantitative analysis with the depth of conversational inquiry, Persona Interviews enable researchers to uncover possible personal narratives and perspectives across different groups in a scalable and reproducible manner. We show that Persona Interviews produce coherent, accurate responses. While shortcomings exist, Persona Interviews expand the methodological toolkit of HCI and computational social science researchers for qualitative data creation.

References

[1] S.-G. Jung, J. Salminen, K. K. Aldous, and B. J. Jansen, “PersonaCraft: Leveraging language models for data-driven persona development,” *International Journal of Human-Computer Studies*, vol. 197, p. 103445, 2025.

[2] J. A. Smith, “Semi-structured interviewing and qualitative analysis,” in *Rethinking methods in psychology*, SAGE publications Ltd, 1995, pp. 10–26.

[3] S. E. Kelly, I. Bourgeault, and R. Dingwall, “Qualitative interviewing techniques and styles,” *The SAGE handbook of qualitative methods in health research*, vol. 19, pp. 307–326, 2010.

[4] B. J. Jansen, S. Jung, and J. Salminen, “Employing large language models in survey research,” *Natural Language Processing Journal*, vol. 4, p. 100020, 2023.

[5] B. J. Jansen, J. O. Salminen, and S. Jung, “Data-driven personas for enhanced user understanding: Combining empathy with rationality for better insights to analytics,” *Data and Information Management*, vol. 4, no. 1, pp. 1–17, 2020.

[6] L. P. Argyle, E. C. Busby, N. Fulda, J. R. Gubler, C. Rytting, and D. Wingate, “Out of one, many: Using language models to simulate human samples,” *Political Analysis*, vol. 31, no. 3, pp. 337–351, 2023.

[7] Z. Chen *et al.*, “Custom training data: Aligning model distribution via Prefix-Guided Preference Data Synthesis,” *Information Processing & Management*, vol. 63, no. 1, p. 104337, 2026.

[8] J. S. Park *et al.*, “Generative Agent Simulations of 1,000 People,” Nov. 15, 2024, *arXiv*: arXiv:2411.10109. doi: 10.48550/arXiv.2411.10109.

[9] B. J. Jansen, A. Spink, J. Bateman, and T. Saracevic, “Real life information retrieval: a study of user queries on the Web,” *SIGIR Forum*, vol. 32, no. 1, pp. 5–17, Apr. 1998, doi: 10.1145/281250.281253.

[10] H. Kallio, A. Pietilä, M. Johnson, and M. Kangasniemi, “Systematic methodological review: developing a framework for a qualitative semi-structured interview guide,” *Journal of Advanced Nursing*, vol. 72, no. 12, pp. 2954–2965, Dec. 2016, doi: 10.1111/jan.13031.

[11] J. Salminen, C. Liu, W. Pian, J. Chi, E. Häyhänen, and B. J. Jansen, “Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions,” in *Proceedings of the CHI Conference on Human Factors in Computing*

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.

Systems, Honolulu HI USA: ACM, May 2024, pp. 1–20. doi: 10.1145/3613904.3642036.

[12] Y. Feng, L. Li, X. Qin, and B. Zhang, "Improving event representation learning via generating and utilizing synthetic data," *Information Processing & Management*, vol. 62, no. 4, p. 104083, 2025.

[13] C. Ziems, W. Held, O. Shaikh, J. Chen, Z. Zhang, and D. Yang, "Can large language models transform computational social science?," *Computational Linguistics*, vol. 50, no. 1, pp. 237–291, 2024.

[14] B. Han, S. Song, D. Liu, and J. Mo, "Mechanism of online public opinion formation in major risk events in China: A qualitative comparative analysis," *Information Processing & Management*, vol. 62, no. 1, p. 103924, 2025.

[15] H. Zhang, "The Evidentiary Value of AI-Generated Data: A Framework for Reliability and Admissibility," *Education and Social Work*, vol. 2, no. 2, pp. 34–41, 2025.

[16] B. J. Jansen, J. Salminen, S. Jung, and K. Guan, *Data-driven personas*. Springer Nature, 2022.

[17] I. Kaate *et al.*, "When Personas Talk to You: Evaluating the Evolution of User Personas from Static Profiles to Conversational User Interfaces," in *Proceedings of the 2025 ACM Designing Interactive Systems Conference*, Madeira Portugal: ACM, July 2025, pp. 2350–2372. doi: 10.1145/3715336.3735676.

[18] L. Sun *et al.*, "Persona-L has Entered the Chat: Leveraging LLMs and Ability-based Framework for Personas of People with Complex Needs," in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, Apr. 2025, pp. 1–31. doi: 10.1145/3706598.3713445.

[19] M. Rizwan, L. Carlsson, and M. Loni, "PersonaBOT: Bringing Customer Personas to Life with LLMs and RAG," May 22, 2025, *arXiv*: arXiv:2505.17156. doi: 10.48550/arXiv.2505.17156.

[20] J. Gao *et al.*, "CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, May 2024, pp. 1–29. doi: 10.1145/3613904.3642002.

[21] L. Tjuatja, V. Chen, T. Wu, A. Talwalkwar, and G. Neubig, "Do llms exhibit human-like response biases? a case study in survey design," *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 1011–1026, 2024.

[22] Z. Xiao, M. X. Zhou, W. Chen, H. Yang, and C. Chi, "If I Hear You Correctly: Building and Evaluating Interview Chatbots with Active Listening Skills," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 2020, pp. 1–14. doi: 10.1145/3313831.3376131.

[23] E. Tallyn, H. Fried, R. Gianni, A. Isard, and C. Speed, "The Ethnobot: Gathering Ethnographies in the Age of IoT," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal QC Canada: ACM, Apr. 2018, pp. 1–13. doi: 10.1145/3173574.3174178.

[24] D. Murthy, "Digital Ethnography: An Examination of the Use of New Technologies for Social Research," *Sociology*, vol. 42, no. 5, pp. 837–855, Oct. 2008, doi: 10.1177/0038038508094565.

[25] S. Beltoft, P. Schneider-Kamp, and S. Askegaard, "Interview Bot: Can Agentic LLM's Perform Ethnographic Interviews?," in *Proceedings of the 17th International Conference on Agents and Artificial Intelligence*, SCITEPRESS-Science and Technology Publications Porto, Portugal, 2025, pp. 702–709.

[26] A. Cuevas, J. V. Scurrell, E. M. Brown, J. Entenmann, and M. I. G. Daepp, "Collecting Qualitative Data at Scale with Large Language Models: A Case Study," *Proc. ACM Hum.-Comput. Interact.*, vol. 9, no. 2, pp. 1–27, May 2025, doi: 10.1145/3710947.

[27] J. Brand, A. Israeli, and D. Ngwe, "Using LLMs for market research," *Harvard business school marketing unit working paper*, no. 23–062, 2023.

[28] G. V. Aher, R. I. Arriaga, and A. T. Kalai, "Using large language models to simulate multiple humans and replicate human subject studies," in *International conference on machine learning*, PMLR, 2023, pp. 337–371.

[29] L. Messeri and M. J. Crockett, "Artificial intelligence and illusions of understanding in scientific research," *Nature*, vol. 627, no. 8002, pp. 49–58, 2024.

[30] E. Pournaras, "Science in the era of ChatGPT, large language models and generative AI," *KI-Kritik/AI Critique Volume 6*, p. 275, 2023.

[31] A. F. Gibson and A. Beattie, "More or less than human? Evaluating the role of AI-as-participant in online qualitative research," *Qualitative Research in Psychology*, vol. 21, no. 2, pp. 175–199, Apr. 2024, doi: 10.1080/14780887.2024.2311427.

[32] I. O. Gallegos *et al.*, "Bias and fairness in large language models: A survey," *Computational Linguistics*, vol. 50, no. 3, pp. 1097–1179, 2024.

[33] P. Hämäläinen, M. Tavast, and A. Kunnari, "Evaluating Large Language Models in Generating Synthetic HCI Research Data: a Case Study," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 2023, pp. 1–19. doi: 10.1145/3544548.3580688.

[34] M. H. J. Lee, J. M. Montgomery, and C. K. Lai, "Large Language Models Portray Socially Subordinate Groups as More Homogeneous, Consistent with a Bias Observed in Humans," in *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, Rio de Janeiro Brazil: ACM, June 2024, pp. 1321–1340. doi: 10.1145/3630106.3658975.

[35] M. Barthet, A. Khalifa, A. Liapis, and G. Yannakakis, "Generative Personas That Behave and Experience Like Humans," in *Proceedings of the 17th International Conference on the Foundations of Digital Games*, Athens Greece: ACM, Sept. 2022, pp. 1–10. doi: 10.1145/3555858.3555879.

[36] T. B. Sardinha, "AI-generated vs human-authored texts: A multidimensional comparison," *Applied Corpus Linguistics*, vol. 4, no. 1, p. 100083, 2024.

[37] S. D. Paoli, "Writing user personas with Large Language Models: Testing phase 6 of a Thematic Analysis of semi-structured interviews," May 29, 2023, *arXiv*: arXiv:2305.18099. doi: 10.48550/arXiv.2305.18099.

[38] D. Korsgaard, T. Bjørner, P. K. Sørensen, and P. Burelli, "Creating user stereotypes for persona development from qualitative data through semi-automatic subspace clustering," *User Model User-Adap Inter*, vol. 30, no. 1, pp. 81–125, Mar. 2020, doi: 10.1007/s11257-019-09252-5.

[39] L. Birt, S. Scott, D. Cavers, C. Campbell, and F. Walter, "Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation?," *Qual Health Res*, vol. 26, no. 13, pp. 1802–1811, Nov. 2016, doi: 10.1177/1049732316654870.

[40] A. Farooq, J. Salminen, J. Martin, K. Aldous, S.-G. Jung, and B. J. Jansen, "Exploring social media privacy concerns: a comprehensive survey study across 16 Middle Eastern and North African countries," *IEEE Access*, 2024.

[41] J. Salminen, B. Jansen, and S.-G. Jung, "Survey2Persona: Rendering Survey Responses as Personas," in *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, Barcelona Spain: ACM, July 2022, pp. 67–73. doi: 10.1145/3511047.3536403.

[42] B. T. West and D. Li, "Sources of Variance in the Accuracy of Interviewer Observations," *Sociological Methods & Research*, vol. 48, no. 3, pp. 485–533, Aug. 2019, doi: 10.1177/0049124117729698.

[43] I. Kaate *et al.*, "'You Always Get an Answer': Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination," in *Proceedings of the 30th International Conference on Intelligent User Interfaces*, Cagliari Italy: ACM, Mar. 2025, pp. 1624–1638. doi: 10.1145/3708359.3712160.

Authorized licensed use limited to: Qatar University. Downloaded on February 20,2026 at 06:52:48 UTC from IEEE Xplore. Restrictions apply.