

## Research Article

**Cite this article:** Gu H, Chandrasegaran S and Lloyd P (2025). Synthetic users: insights from designers' interactions with persona-based chatbots. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **39**, e2, 1–17

<https://doi.org/10.1017/S0890060424000283>

Received: 29 February 2024

Revised: 09 October 2024

Accepted: 27 October 2024

### Keywords:


chatbots; design process; personas; large language models

### Corresponding author:

(Eric) Heng Gu;

Email: [h.gu@tudelft.nl](mailto:h.gu@tudelft.nl)

# Synthetic users: insights from designers' interactions with persona-based chatbots

(Eric) Heng Gu , Senthil Chandrasegaran and Peter Lloyd

Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands

## Abstract

Personas are hypothetical representations of real-world people used as storytelling tools to help designers identify the goals, constraints, and scenarios of particular user groups. A well-constructed persona can provide enough detail to trigger recognition and empathy while leaving room for varying interpretations of users. While a traditional persona is a static representation of a potential user group, a chatbot representation of a persona is dynamic, in that it allows designers to “converse with” the representation. Such representations are further augmented by the use of large language models (LLMs), displaying more human-like characteristics such as emotions, priorities, and values. In this paper, we introduce the term “Synthetic User” to describe such representations of personas that are informed by traditional data and augmented by synthetic data. We study the effect of one example of such a Synthetic User – embodied as a chatbot – on the designers' process, outcome, and their perception of the persona using a between-subjects study comparing it to a traditional persona summary. While designers showed comparable diversity in the ideas that emerged from both conditions, we find in the Synthetic User condition a greater variation in how designers perceive the persona's attributes. We also find that the Synthetic User allows novel interactions such as seeking feedback and testing assumptions. We make suggestions for balancing consistency and variation in Synthetic User performance and propose guidelines for future development.

## Introduction

The chief tasks of a designer include determining, addressing, and representing the experiences of user groups in an informative manner to support the creative design process of solution finding (Daalhuizen et al., 2019; Stolterman, 2021). Personas – profiles of individuals with specific and diverse goals and needs – are one such means of representation, used as a tangible means of condensing and communicating what would otherwise be abstract qualitative and quantitative data on target users, customers, or stakeholders (Pruitt and Grudin, 2003). They are applied as a method chiefly in the early stages of designing to stimulate and direct ideation but also at the end stages of the process of communicating important information (Salminen et al., 2022). Personas have become a mainstay in design research as a popular approach for representing different types of user needs (Salminen et al., 2020b), having been shown to strengthen user-centeredness in designers' decision-making (Heck et al., 2018), improve communication between designers (Grudin and Pruitt, 2002), and result in designs that better address usability requirements (Schneidewind et al., 2012).

Personas are traditionally synthesized manually through qualitative research techniques such as surveys, interviews, and ethnographic studies (Pruitt and Grudin, 2003; Jansen et al., 2021). These methods can be resource- and time-intensive (Cooper, 1999; Mulder and Yaar, 2006; Nielsen, 2019) and often not actually based on first-hand data (McGinn and Kotamraju, 2008). Even when personas are rigorously defined, they face decreasing relevance as user behaviors change over time (Salminen et al., 2020a; Jansen et al., 2019; Jung et al., 2019). As a result, quantitative and algorithmic approaches for generating personas have been a growing field in the last 15 years (Salminen et al., 2020a). These approaches have the benefit of scalability (An et al., 2018) and potential objectivity due to their data-driven nature, though issues of algorithmic bias remain (Salminen et al., 2021a). Advancements in natural language processing (NLP) have been foundational to these automated approaches.

The rise of chatbots – conversational interfaces that simulate social interaction with human interlocutors – in replacing online versions of frequently-asked questions or how-to documents, as well as in online domains for business, health, and entertainment (Adamopoulou and Moussiades, 2020). Pivotal work by Li et al. (2016) introduced persona-dense conversational agents by imbuing chatbots with a *speaker model* that encodes character, using specific background information, speaking style, and other elements of identity. Large language models (LLMs), having been trained on extensive datasets that include persona-dense information, demonstrate a remarkable ability to mimic various personalities, such as those of historical figures (Haller and Rebedea, 2013; Adamopoulou and Moussiades, 2020), fictional characters

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

(Schmitt and Buschek, 2021), and modern celebrities (Lee and Na, 2018). Recent advancements in these models enable them to accurately replicate the responses of actual user groups, enhancing their realism and applicability (Hämäläinen et al., 2023). However, studies have also illuminated the inherent biases captured within these models, necessitating cautious and critical application of the technology in practice. Notably, studies have identified significant political (Motoki et al., 2024; Rozado, 2023) and social biases (Zack et al., 2024; Howard and Borenstein, 2018) in models accessible to the public, potentially reflecting and amplifying existing user prejudices in a manner similar to existing media channels. For Synthetic User creation, the biases of LLMs require an examination of their influence on the performance and realism of these simulations.

The suggestion to employ persona-dense chatbots as stand-ins for real-world users in design interviews was made with the intent of streamlining the design process (Vandenberghe, 2017). While it is debatable whether chatbots can fully replace human participants, we posit that they can serve similarly to traditional persona summaries, offering a condensed representation of user groups to foster empathy and understanding in human-centered design endeavors. Furthermore, the similarity in data slots (Name, Age, Gender, Occupation, and so on; Song et al., 2019) highlights the potential relatedness and interoperability between Synthetic Users animated by persona data and traditional persona summaries. Their effectiveness could be tremendously beneficial for design research. If ineffective, it warrants investigation into design interventions necessary to leverage this technology optimally.

Motivated by this, our study investigates the application of persona-based chatbots as Synthetic Users – digital stand-ins for realistic user groups during design sessions. Our primary objective is to examine how they can preserve the advantages of traditional personas, such as eliciting empathy and increasing user-centeredness in designing products and services.

Notably, this paper contributes to the current fundamental discussion of design research about how generative artificial intelligence (AI) techniques can be implemented to improve the design process and its outcomes, while managing unintended outcomes (Calabretta and Kleinsmann, 2017; Stoimenova and Kleinsmann, 2020). Dortheimer et al. (2024) explored the experimental use of a chatbot as an interviewer to investigate interviewee engagement with the technology and its implications for stakeholder engagement. Similarly, our study examines the other aspects of human–AI interaction in stakeholder engagement within the design context. We present an experimental study on persona chatbots as Synthetic Users, focusing on how designers perceive and interact with this technology.

Specifically, our approach provides a generic Synthetic User template for animating any designerly made persona, from which additional studies can be explored to enhance the Synthetic User performance and capabilities. In the following sections, we will first discuss the evolution of persona characteristics and usage, followed by the development of our Synthetic User web app and corresponding study, analysis, and results using qualitative and quantitative methods. Finally, we discuss the challenges and opportunities presented by Synthetic Users in the design process.

## Background

### *Evolution of personas in design*

Since their inception, the use of personas has expanded from traditional product design (Cooper, 1999) to various other fields.

A review of use cases of personas in design by Salminen et al. (2022) shows personas being applied in areas such as software development, healthcare, and higher education. Their survey also shows personas being used with other methods such as scenarios, interviews, and prototyping. They find that personas are primarily used in the early stages of sensemaking and ideation, as well as in the later stages of communicating important user information to stakeholders.

Personas in professional design practice are created primarily using qualitative, unstructured observations and expert verification. Creating an appropriate persona is also time-consuming, with costs averaging \$49,000 in 2010 (a more recent estimate was not feasible due to the unavailability of relevant data; Drego and Dorsey, 2010; Salminen et al., 2020b). There are a variety of persona creation approaches, both qualitative (Cooper, 1999), such as role-based (e.g., Pruitt and Grudin, 2003), or fiction-based, (e.g., Blythe and Wright, 2006) and quantitative, such as k-means clustering, hierarchical clustering, principal component analysis, and latent semantic analysis (Jansen et al., 2020). These approaches more or less follow the same three distinct phases: data collection, clustering/segmentation of user groups, and persona generation (Mulder and Yaar, 2006; Nielsen, 2019). Historically, data collection involves such methods as ethnographic studies, surveys, and focus groups (Jansen et al., 2021; Pruitt and Grudin, 2003), with modern techniques tapping into large-scale user data from social networks, e-commerce, or other sources (Salminen et al., 2020a).

While there are no standard, established criteria for producing personas, the consensus (e.g. Cooper, 1999; Nielsen and Storgaard Hansen, 2014; Pruitt and Grudin, 2003) is that personas should represent a group of people, but they should be presented as a specific fictional individual, with rich details such as a name, occupation, age, gender, socioeconomic status, interests, stories, and motivations. Despite the emphasis on detail, personas should ideally be short and concise, typically no more than 1–2 pages (Nielsen et al., 2015) of user demographic and behavioral information (Cooper, 1999; Jansen et al., 2020). Achieving this balance between providing as much information as possible to create a realistic impression of the user group while providing as little information as necessary to avoid constraining designer creativity highlights a significant challenge for the application of personas. We contend that much of the hidden biases associated with the use of personas can be attributed to this paradox of realistic and reducible representation.

The psychological foundation of why personas work as a tool for stimulating empathy is the same reason they can also trigger stereotyping: they offer a simplified and generalized representation of user segments, making it easier to relate to but also potentially leading to overgeneralizations (Grudin and Pruitt, 2002). Research has found that personas created from designer intuitions tend to be more stereotypical than those based on actual user data, even when inclusiveness is stated as one of the design objectives (Turner and Turner, 2011). Attributes like gender (Hill et al., 2017; Marsden et al., 2017), race (Salminen et al., 2019b), and age (Marsden et al., 2017) in a persona are particularly prone to inducing stereotyping. Interestingly, novice designers are more susceptible to stereotyping than their experienced counterparts, displaying a higher degree of empathy and trust in persona data when the persona resembles themselves and a reduced inclination when they are dissimilar (Dahiya and Kumar, 2021).

Matthews et al. (2012) found in their study of user-centered design practitioners, that “persona champions” – designers who

used personas as described in the literature and had a positive perception of the method – tended to use personas more effectively.

We refer to this phenomenon as the *Champion bias*, in which the advocate and principal user of the persona tends to get better results from them, likely due to familiarity and comfort with the “personification” process. The persona summary acts as a reductive point of access to the rich source data.

### Persona-based chatbots as Synthetic Users

Chatbots today are increasingly used in such diverse domains as scalable and automated customer service, triage in healthcare systems, or even in gaming as nonplaying characters (Adamopoulou and Moussiades, 2020; Athota et al., 2020). They are also gaining popularity as personal voice assistants on smart mobile devices (Leviathan and Matias, 2018). *Persona-based chatbots* represent a subdomain of chatbot development that emphasizes a cohesive and coherent personality on top of a baseline ability to perform naturalistic conversation (Li et al., 2016). Persona-based chatbots were developed to address three main issues that affect user engagement in chatbot interactions: (1) the lack of a consistent personality, (2) lack of explicit long-term memory, and (3) a tendency to produce non-specific answers (Zhang et al., 2018).

The motivation for imbuing a chatbot with “*persona density*” – in other words, information about a specific profile – is primarily to enhance the user experience by creating realistic and sustainable conversations (Song et al., 2019), stimulate empathy in the user, and improve task outcomes (Zhong et al., 2020). Such persona density is typically achieved either implicitly by training a model to recognize persona-related features directly from personalized datasets (Song et al., 2021), or explicitly by assigning a defined profile (Qian et al., 2018). Hämäläinen et al. (2023) evaluated and observed – via an open-ended HCI questionnaire – text generated by GPT-3, or Generative Pre-trained Transformer 3, an LLM capable of generating text that appears human-generated. They found the text generated by the model in response to the questionnaire to be plausible but less diverse compared to responses by humans to the same questions. The synthetic responses were evaluated as human-like in general and identified as more human than actual human responses in more than 50% of cases. However, GPT-3 also exhibited a tendency to evade questions and occasionally produced false information about the given topics. Despite these limitations, the researchers concluded that synthetic data remains valuable for initial exploration and piloting purposes.

The concept of using AI simulations as “off-the-shelf-users” was first proposed by Vandenberghe (2017) in a thought experiment and listed their potential benefits: they never get tired; they are available at any time; they are scalable in terms of recruitment and interviewing; and they can withstand mistreatment without consequences. In the context of our research, we define “Synthetic Users” as a category of persona-based chatbots that carry unique, humanistic characteristics, backgrounds, values, and preferences. Unlike general persona-based chatbots, Synthetic Users are designed to emulate the behavior and responses of real users that their persona represents, thereby serving as a stand-in for human interaction.

### Design of chatbot as Synthetic User

#### Persona considerations and selection

To inform the design of our Synthetic User, we first collected a variety of off-the-shelf persona summaries from both academia and

industry. A common practice in persona design, particularly when creating a *set* of personas, is to use a spectrum to represent the differences between them, such as using a slider value between introversion and extroversion. Industry personas especially follow this trend (Gaudry, 2022). Modern personas, such as those created by automatic persona generation (Salminen et al., 2019a) or qualitative co-creation techniques (Neate et al., 2019) tend to use very simplified snippets of text and use a greater proportion of graphical content. These quantified and graphical persona descriptions were not suitable as seed data for the Synthetic User. Recent developments in automatic persona generation leverage LLMs, and with the advent of ChatGPT, design researchers are investigating prompt engineering approaches to create effective persona summaries (Schmidt et al., 2024; Paoli, 2023). However, due to the novelty and unproven reliability of these LLM-designed personas, we chose to use personas crafted by human experts. We found the set of personas created by Roussou et al. (2013) to be appropriate for our approach.

Roussou et al.’s persona summaries (2013) possess text descriptions of specific details that acted as good source material for our language model. Due to the significant influence of photorealism on the perception of the persona and unintended stereotyping demonstrated by Salminen et al. (2021b), we decided to exclude portraits from both the Synthetic User and persona summary, aiming to reduce the biases stemming from the images. The finalized persona summary we employed can be viewed in Figure 1. The Synthetic User shares identical text content, with the only change being the transformation of bullet points into complete sentences, all referencing ‘Natalie’ as the subject.

#### Interface design

Prompting, first discovered by Brown et al. (2020) is a technique for efficiently tuning language model outputs through explicit instructions. We adopt this technique for our Synthetic User design, by providing a natural-language description of the task and persona data, as well as selective injection of relevant dialogue as examples when the conversation is ongoing.

We chose GPT-3, the language model most commonly used for prompt-programming at the time of conducting our study (Logan IV et al., 2022; Reynolds and McDonell, 2021; Lester et al., 2021). For this study, we use the “text-davinci-003” model. As with all LLMs, we were constrained by token limitations, which contributed to the lack of a persistent memory to track the progress of the conversation.

To address these limitations, we built a simple prompt-programmer that tracks the conversations and dynamically adjusts prompts with the persona data and relevant snippets of past dialogue. We chose KeyBERT to process the queries and responses. KeyBERT is a keyword extraction method by the creator of BERTopic (Grootendorst, 2022) and relies on BERT-based pre-trained word embeddings (Grootendorst, 2020). Compared to other extraction methods, KeyBERT performs better than the norm for long texts, and decently for short texts (Giarelis et al., 2021). We followed the parameter for KeyBERT tested by Giarelis et al. (2021) that yielded the best results for short text (method = “maxsum”, diversity = 0.7).

Figure 2 shows two sample conversation turns, one query (labeled “1” in the figure) that the participants submitted containing keywords linked to relevant past exchanges and another query (2) that did not. Up to three unique keywords are extracted using KeyBERT (3) and used to retrieve past exchanges containing the same keywords (4) as conversation samples for the prompt (5). If no

## Natalie Schmidt (35) The Young Woman - IT Company Executive

*"The museum is an excellent way to relax between meetings!"*

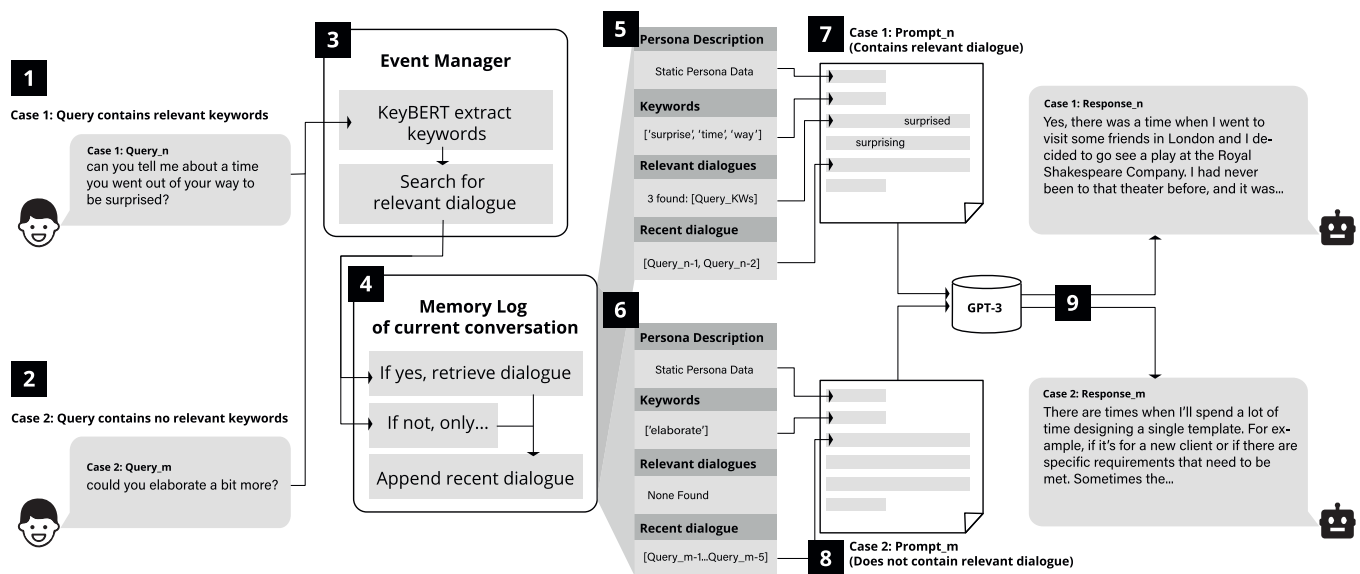
### Summary:

- She works as a web designer at a medium-sized IT company.
- She is interested in various different things, such as art, history, sociology. She tries to read in her spare time.
- Her company participates on EU projects and she travels a lot. Although when traveling, her time is limited, she always finds the time to visit a museum, preferring art museums.
- She tries to have a brief overview of the collections of museums she visits, and if she has time, she focuses on specific exhibits that she finds more appealing.
- She usually does not prefer a guided tour as she finds it stressful if her time is limited.
- She prefers, when possible, to prepare her own visit by searching for information on the web.

### Goals:

- To relax and enjoy aesthetically beautiful objects
- To learn interesting facts about topics that interests her, related to the museum exhibits
- To have easy access to information that can help her prepare her museum visit.
- To take good advantage of the limited time she has available in the museums.

**Figure 1.** A traditional persona developed by Roussou et al. (2013) adopted for our study.



**Figure 2.** Examples of conversation turns involving (1) user query containing keywords associated with past exchanges and (2) user query that does not. The event manager (3) receives the query and extracts keywords using KeyBERT and (4) retrieves relevant and/or recent exchanges. Different data (5 and 6) for corresponding prompts (7 and 8) are used to inform the GPT-3 text generation (9).

keywords are present as in Case 2, then only recent exchanges are appended as samples (6). The prompts (7) and (8) are passed to the language model and responses served to the chat interface (9). This system ensures the language model will not provide an answer inconsistent with a previous response as the length of the conversation grows. Tests on parameter tweaking and the corresponding responses from the Synthetic User are provided in the supplemental material. We designed a Python-based messaging interface (Figure 3) hosting the interaction.

### Research questions

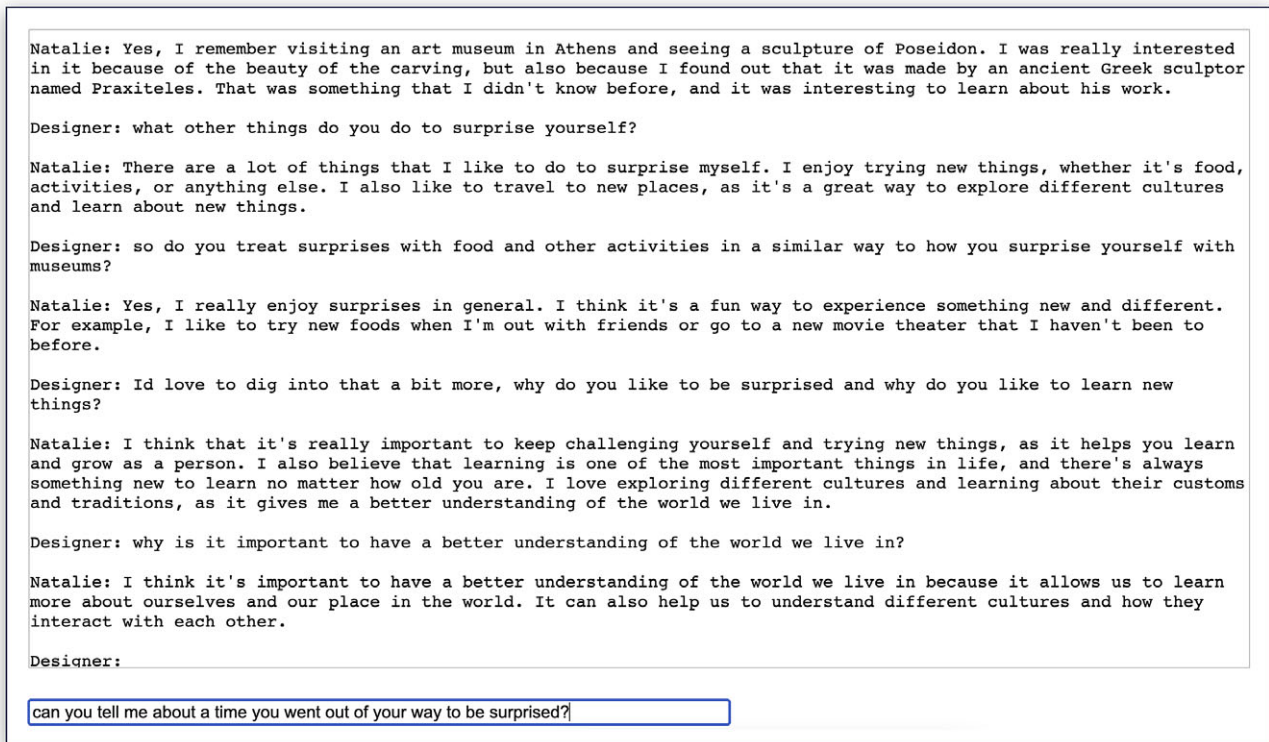
For clarity, we define *Persona* as a fictional representation of a *User Group*, which refers to the final intended consumers or users of a product or service, and a *Synthetic User* is an AI-driven animated

persona designed for interactive engagement with designers. With these definitions, the main research question can be stated broadly as follows: "What is the effect of a Synthetic User on designers' performance in terms of insight about a user group?" This question breaks down into three subquestions:

**RQ1:** What is the effect of a designer's interaction with a Synthetic User related to the empathy that is felt toward the Persona?

**RQ2:** What is the effect of a designer's interaction with a Synthetic User on the designer's understanding of the needs of the user group represented by the Persona?

**RQ3:** What is the effect of a designer's interaction with a Synthetic User on the designer's ideation outcome?




**Figure 3.** Screenshot of the chat interface showing a typical interaction between the participants in the study and the “Synthetic User” or an application of persona-based chatbot. The participants type their questions/remarks into the text input field at the bottom, and see their text input (shown as conversational turns by “Designer”) and the corresponding “responses” from the Synthetic User (shown as conversational turns by ‘Natalie’) in the text window above.

## Study


To address these research questions, we conducted a between-subjects study involving a design task for product or service ideas to benefit a hypothetical user group represented by the persona.

While in professional scenarios we intend for a Synthetic User to augment and not replace traditional personas, we used a traditional persona as a baseline for our study, representing the current standard format in which designers engage with personas. By comparing the participants’ behavior and task outcome between the two groups (Synthetic User vs. Baseline), we could identify whether and in what ways they can enrich the design process and outcome. We believe that in actual usage, a static persona summary could be presented alongside the Synthetic User, but for the study, we excluded providing the static persona summary for the Synthetic User condition to solely assess the impact of the conversational modality.

## Participants

Out of 22 initial participants, we completed the study with 20 participants (6 female, 14 male, 0 nonbinary), with the remaining two (1 female, 1 male) opting out of the study halfway. (Sec. 6.5 discusses the reason for the participants opting to discontinue the study.) The teams were split equally by gender (3 female, 7 male in either condition). The participants comprised 14 design practitioners from industry, 5 Ph.D. students majoring in design, an assistant professor in design, a post-doctoral researcher in design, and a master’s student majoring in design. Participants’ design experience ranged between 1 and 20 years (1  20) with a

median of 6 years’ experience. These data were collected during the study after the participant had already been assigned to their condition; thus, the distribution of design experience was uneven with participants in the Synthetic User condition having a mean experience of 6.3 years ( $SD = 5.54$ ), while those in the baseline condition had an average of 6.4 years ( $SD = 2.07$ ).

All participants indicated they had prior interaction with some form of chatbot, particularly customer service bots (all 20 participants) and voice assistants such as Siri, Alexa, and Cortana (19 participants). Seventeen participants had engaged with personas prior to the study. Based on their responses on a 7-point Likert scale (dislike  like), 11 participants appeared to view personas favorably while four appeared to dislike personas as a design technique.

## Conditions and tasks

As mentioned earlier, the between-subjects study setup meant that 10 of the above participants used the Synthetic User for their tasks, while the remaining 10 used the baseline “traditional” persona representation. Both groups completed a background survey and a warm-up divergent thinking exercise, called “30 Circles” (Kelley, 2018) where they were given a sheet of paper with 30 blank circles and asked to create sketches transforming the circles into different objects. Both groups then followed the procedure outlined below.

**Task 1. Insight Gathering:** Participants had 10 minutes to study and gather as much information from the persona named Natalie as they could in order to understand consumers like her. The Synthetic

User group was asked to interview Natalie, with only some basic information: her name, age (35), profession (IT executive), location (Berlin), and situation (frequently travels for work). The other group was presented with Natalie's persona summary and asked to interpret and extrapolate the customer group she represents. Both groups were asked to think-aloud during this process.

**Task 2. Ideation:** Participants had 20 minutes to imagine as many product or service ideas as possible for customers like Natalie to help them discover new meaningful experiences in their life of frequent work travels. Each group was allowed to reference the persona in the assigned form (Synthetic User or Persona Summary).

Following the exercises, participants answered a NASA TLX usability survey (Hart and Staveland, 1988) and a modified Persona Perception Scale (PPS; Salminen et al., 2020b). The PPS was originally developed to evaluate individual impressions of various persona qualities, such as credibility, completeness, and consistency. The creators highlighted its broad applicability for evaluating persona designs and strategies, making it suitable for assessing alternative persona representations like the Synthetic User. Similar to other research that has applied the PPS to evaluate conversational AI (Ha et al., 2024), we believe it is appropriate for our study as well. We removed two metrics from our version of the PPS because they did not apply to our experimental setup. These metrics are *Consistency*, which measures how consistent the various types of information (i.e., photo and profile) are, and *Willingness to use*, which indicates the participant's desire to use this persona for their task. Since we were not evaluating the quality of the persona content writing, we deemed these two metrics less relevant to our study.

### Experimental setup

All participants joined the study remotely through Microsoft Teams. Those in the Synthetic User condition interacted with the bot hosted on a 16-inch Macbook Pro with the chatroom interface (Figure 3) running on a web browser (Chrome/Safari). The screen was recorded to capture participants' interactions with the persona along with their think-aloud process. To avoid the participants having to switch between typing and speaking, one researcher transcribed their questions verbatim into the Chatroom interface. Participants in the baseline condition were presented with a digital document of the persona (Figure 1). All participants were recommended the use of pen and paper to generate ideas; two of the 20 participants used an iPad and stylus as a substitute.

### Analysis and results

This section looks at the analysis and results derived from our study, focusing on three research questions: the perceived empathy toward the Persona (RQ1), the understanding of the User Group (RQ2), and the ideation outcomes (RQ3), and on two conditions: participants using a traditional Persona Summary vs. a Synthetic User. Participants' self-reported perceptions were gauged using the PPP, which we further categorized into empathy-related and understanding-related measures. To supplement these self-reports, we used a dictionary-based text analysis tool to understand participants' verbalized thought processes during the study. The subsequent subsections will provide a detailed breakdown of the results, starting with the evaluation of empathy and understanding, followed by insights into ideation outcomes and designer behavior.

### Evaluating empathy and understanding

Participants self-reported their perception of the Synthetic User through the PPS (see Analysis Section). The PPS measured self-reported perceptions of the persona. We group the PPS categories of "empathy," "similarity," and "likeability" together which we categorize under how much empathy the participant thought they felt toward the persona (RQ1), while "completeness," "credibility," and "clarity" are grouped as indicative of how much understanding the participant thought they had about the persona (RQ2). For the full figure of the PPS results, see Figure 5.

To augment and possibly explain these self-reported results, we employed LIWC, or Linguistic Inquiry and Word Count (Tausczik and Pennebaker, 2010) to analyze the transcripts of the participants thinking aloud, that is, a transcript of their thought process during the study. LIWC is a widely used tool that primarily incorporates dictionary-based text analysis methods. The dictionary categories in LIWC use linguistic indicators to identify psychological processes such as sentiment, cognition, and emotion. Based on the LIWC-22 psychometric properties (Boyd et al., 2022), we identified categories of LIWC relevant to RQ1 and RQ2.

#### Empathy toward the User Group (RQ1)

A Mann–Whitney U-test of the data (responses on a 7-point Likert scale) showed no statistically significant difference between the two groups' self-reported scores on the categories related to RQ1. While the above analysis was performed on participants' self-reported data, the participant transcripts were analyzed using LIWC categories related to empathy to determine if the participants exhibited empathy while talking about the persona. Drawing from previous research that used LIWC categories to evaluate the level of Empathy, we chose *affect*, *social*, *cognition*, and *perception* as common metrics associated with empathy (Samrose et al., 2020; Gibson et al., 2015).

*Clout* (relates to authority and confidence) and *authenticity* (reveals the level of self-disclosure) have also been used to study empathy and compassion in professional client conversations (Qiu and Tay, 2023; Mascaro et al., 2023). In these studies, clout was evaluated as a metric of the professional's confidence and authority, indicating the degree to which the professional's language demonstrates expertise and assurance. Higher measurement of clout in this professional relationship corresponds to higher perceived compassion and empathy.

Of all the responses on the PPS metrics, three of the four statements under *Similarity* passed the Shapiro–Wilk test for normality and were evaluated using a two-tailed *t*-test. Responses to the statement *This character feels similar to me* under *Similarity* and all statements under the rest of the categories were analyzed using the Mann-Whitney *U* test. No significant differences were found between the two conditions (see Table 1 for details). Only the *Social* LIWC category from the ones described showed a significant difference between the conditions, with a lower score in the Synthetic User scenario (see Table 2).

#### Understanding of the User Group (RQ2)

The results from the PPS showed no significant difference between the condition, with the exception of the statement, *This character seems like a real person* ( $U = 18.5$ ,  $p < .05$ ) with participants in the Baseline condition reporting a higher score ( $Md = 6.0$ ,  $IQR = 5.0 - 6.0$ ) than those in the Synthetic User condition ( $Md = 3.0$ ,  $IQR = 2.0 - 5.0$ ).

**Table 1.** Medians and interquartile ranges for scores in each PPS category

RQ	PPS Category	PPS Measure	Baseline		Synthetic User	
			Median	IQR	Median	IQR
RQ1	Empathy	I feel I understand this character	5.0	5.00–5.75	5.5	4.25–6.00
		I feel strong ties to this character	4.0	2.25–5.00	4.0	2.00–4.00
		I can imagine a day in their life	6.0	5.00–6.00	6.0	5.00–6.00
	Likeability	I find this character likable	5.5	4.00–6.00	4.0	4.00–6.00
		I could be friends with this character	5.5	5.00–6.00	4.0	3.25–5.00
		I could spend time with this character	5.5	4.25–6.00	4.0	3.25–5.00
		This character is interesting	5.5	3.50–6.00	3.5	2.25–4.75
	Similarity	This character feels similar to me	5.0	4.25–5.00	3.5	1.50–5.00
		This character and I think alike	3.5	3.00–5.00	4.0	3.25–5.00
		This character and I share similar interests	5.0	4.25–5.75	5.0	3.25–6.00
I feel I would agree with this character...		4.5	4.00–5.00	4.0	2.25–4.75	
RQ2	Completeness	The character is detailed enough...	4.5	3.00–6.00	4.0	2.50–5.75
		The character provides enough info...	5.5	3.50–6.00	4.0	3.25–4.75
		The character seems complete	3.0	2.00–5.00	3.5	2.25–5.00
		The character is not missing vital info	4.5	3.25–6.00	4.5	3.25–5.00
	*Credibility	This character seems like a real person	6.0	5.25–6.00	3.0	2.00–5.00
		I've met people like this character	6.0	6.00–6.00	6.0	5.25–6.00
		They seem to have a personality	6.0	3.25–6.00	4.0	3.00–6.00
	Clarity	Information is well presented	5.0	4.25–5.75	5.5	5.00–6.00
		Info about them is easy to understand	6.0	5.00–6.00	6.0	5.25–6.00
		The character is memorable	4.0	3.25–5.75	4.5	2.25–5.00

Note: A \* against a measure indicates a statistically significant ( $p < .05$ ) difference between the conditions.

Similar to the process in analysis of Empathy, we analyzed transcripts of the participants' think-aloud data using LIWC categories that would relate to expressions of understanding. We determined the subcategories of *allure* and *curiosity*, along with *need* and *want* as those related to participants' understanding of the persona's motivations and aspects that influence their behavior. *Curiosity* words reflect the speakers' search or interest in new knowledge, while *allure* refers to words commonly used in persuasive advertising communication to indicate desirable qualities (Boyd et al., 2022). Both wants and needs are associated with language that necessitates action or behavior. These categories were chosen as they align with evaluating the design requirements for user groups.

Additionally, based on work by Boyd et al. (2022), we considered categories like Cognitive Processes (especially *insight* and *causation*), *all-or-none*, and *certitude* which corresponded to participants' level of certainty regarding their understanding. The use of *causal* words (e.g., because, effect, hence) and *insight* words (e.g., think, know, consider), two subcategories of cognitive mechanisms, in describing a past event can suggest the active process of reappraisal. Absolutist language (e.g., all, none, never) refers to over-generalized, weakly-correlated with certitude, as indicators of confidence and certainty.

Among these categories, we observed a significant difference between the groups only in *Insight*, showing a lower score in the Synthetic User condition (see Table 2).

### Ideation outcome and designer behavior

A total of 274 concepts were generated from the 20 participants, with exactly 137 in each condition. The average number of ideas for both groups was the same ( $M = 13.7$ ), though the spread of idea counts across participants was narrower for the Baseline group ( $SD = 4.24$ ) than for the Synthetic User group ( $SD = 8.66$ ), indicating a greater consistency in idea fluency in the Baseline group. We wanted to know whether or not the Synthetic User can impact the diversity of ideas generated by the participants.

We turned to topic modeling using BERTopic (Grootendorst, 2022) to determine the themes within the ideas generated. These themes were compared across the two conditions to get a sense of the diversity of ideas: more themes within a condition could mean more diverse ideas. Figure 6 offers a comparison of the two conditions, showing a comparable diversity of ideas in both conditions, each with two unique topics not present in the other as well as the same number of ideas in the "Outlier" group.

The large "Outlier" group mainly resulted from model limitations. For example, most of the ideas related to "dating" and "museums" were classified as outliers due to tokens like "Tinder" and "matchmaking" being absent from the model. Niche technology terms like "Chatbot," "telepresence," "AR/VR," and "Geolocation" also contributed to outliers due to their absence in the model's tokens list. Some ideas that could fit existing categories were missed by automated clustering, illustrating model

**Table 2.** Comparison of LIWC categories in Baseline and Synthetic User conditions

RQ	Category	LIWC Category	Condition	Value		Examples
				Mean	SD	
RQ1	Empathy	Clout	Baseline	51.83	18.38	--
			Syn. User	37.07	20.17	--
		Authentic	Baseline	32.08	15.35	--
			Syn. User	37.94	22.76	--
		Cognition	Baseline	19.51	1.44	"There's <b>no</b> image so <b>obviously wondering</b> what the person looks like."
			Syn. User	19.46	2.83	"I'm <b>thinking</b> of doing <b>other ideas</b> ..."
		Affect	Baseline	3.65	0.53	"I <b>would like</b> to have group of ideas which <b>help</b> to give her..."
			Syn. User	4.41	1.93	"It sounds like you <b>love to</b> travel..."
		* Social	Baseline	11.58	1.35	"Could <b>you</b> have a, <b>you</b> know, like a <b>shared personal assistant</b> makes an itinerary for <b>you</b> ?"
			Syn. User	9.55	2.17	"So it sounds like <b>you love learning</b> about a range of different things"
		Perception	Baseline	7.79	1.09	"I'm starting to <b>visualize</b> what the <b>warehouse</b> and the rooms would <b>look</b> like."
			Syn. User	8.63	1.33	" <b>Sounds</b> like her smartphone..."
		Conversation	Baseline	5.17	1.71	" <b>um</b> so I'm <b>gonna</b> just start by..."
			Syn. User	6.18	2.37	" <b>um</b> so I'm just <b>gonna</b> like <b>um</b> ..."
RQ2	Certainty	allnone	Baseline	0.71	0.30	"If she <b>always</b> travel a lot to run the project..."
			Syn. User	1.00	0.44	"She already has <b>everything</b> ..."
		certitude	Baseline	0.93	0.36	"The solution to this seems <b>self-evident</b> ..."
			Syn. User	1.37	0.76	" <b>Obviously</b> she loves her friends..."
	Cognitive Processing	* insight	Baseline	4.98	0.59	"I <b>find brainstorming</b> out loud a lot <b>trickier</b> ."
			Syn. User	3.94	1.05	"I'm <b>thinking</b> what else can I can ask her to get to <b>know</b> more."
		cause	Baseline	1.57	0.29	"I can kind of <b>infer</b> that <b>because</b> she wants"
			Syn. User	1.45	0.57	" <b>Because</b> there's obviously..."
	Motives	curiosity	Baseline	1.11	0.45	"Just <b>curious</b> about her as a person"
			Syn. User	1.04	0.72	" <b>Trying to find</b> other like-minded people..."
		allure	Baseline	9.34	1.07	"to relax and <b>enjoy</b> aesthetically <b>beautiful</b> objects to <b>learn</b> interesting facts."
			Syn. User	11.20	2.51	"It sounds <b>like</b> you love to <b>travel</b> ..."
	States	need	Baseline	0.34	0.19	"She's a power woman that <b>needs</b> to go fast."
			Syn. User	0.53	0.39	"So maybe Natalie <b>needs</b> some help with..."
want		Baseline	0.53	0.22	"She doesn't <b>want</b> guided tours."	
		Syn. User	0.53	0.50	"A <b>desire</b> to do..."	

Note: – A \* indicates a statistically significant ( $p < .05$ ) difference between the conditions.

– LIWC Dictionary category words are shown in **bold italics** in the *Examples* column.

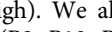
– "Clout" and "Authentic" are aggregated measures and do not have dictionary entries.

– RQ3 is not mentioned above as it has to do with ideation outcomes.

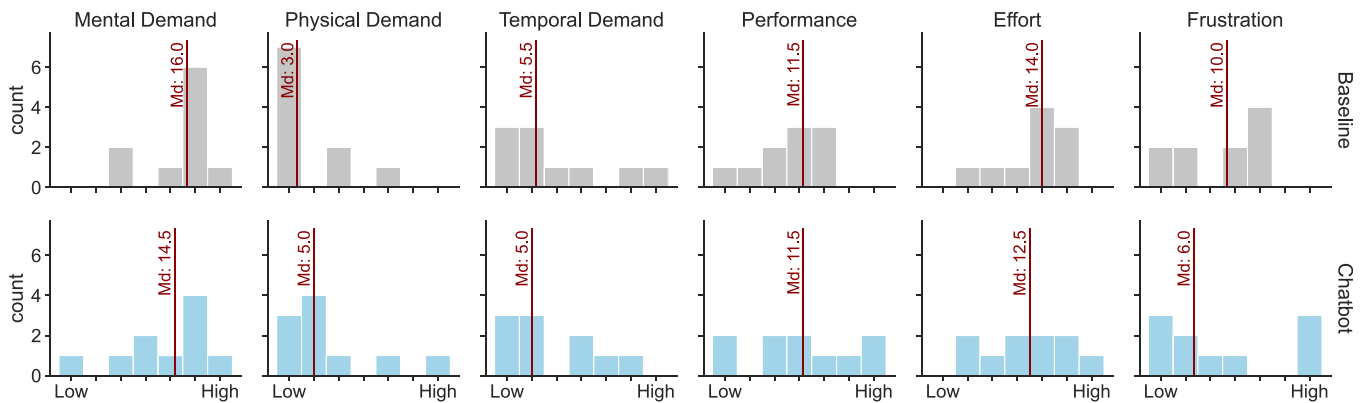
limitations. For instance, the idea of "Book or material of all permanent collections in museums" could have been categorized under "books, podcasts, posters" but was likely excluded due to the specificity of "permanent collections," which may have carried a higher weight within the sentence. A more robust model could reduce the "Outlier" cluster. However, no significant factors affecting comparability of the two sets of ideas generated by the two conditions were identified.

Response to the NASA TLX scales showed minimal differences between the ideation with the persona summary and the Synthetic User. A Shapiro–Wilk test shows normal distributions for the mental, physical, and temporal demands, but not for the rest, but


both the t-test and U-test found no significance in the TLX results (see Figure 4).

The spread of the participant responses on the "frustration" component of the TLX is worth noting as the participants in the Synthetic Users group showed a greater spread than the Baseline group. Examining the distribution for the Synthetic Users group, we see that most participants reported neutral-to-low frustration levels for the task (low  high). We also see a bimodal distribution with three participants (P9, P10, P27) reporting the highest frustration levels, and an equal number (P17, P23, P31) reporting the lowest, centered around a median of 6.0 on a 21-point scale, representing low frustration. Compare this to the baseline





**Figure 4.** Participant responses for the task described in the Study section on the NASA TLX scale (Hart and Staveland, 1988), separated by condition (Baseline vs. Chatbot). Median values are shown for each scale item and condition.

group where the spread is narrower (low  high) with a moderate median of 10.0.

Of the 10 minutes allocated to “Insight Gathering” (Task 1), participants in the Baseline condition used an average of 6 minutes and 51 seconds, with only 20% of participants using all 10 minutes.

All Synthetic User participants used the full 10 minutes. Nine out of 10 Synthetic User participants also used the full 20 minutes allocated to “Ideate” (Task 2), with the one dropout being P9, who reported the highest frustration for the Synthetic User.

In the Baseline group, 8 out of 10 participants used the full 20 minutes. Two Synthetic User participants dropped out from the study during “Ideation” (Task 2).

## Discussion

In our investigation of the impact of Synthetic User interaction on designers, all three research questions yielded results that were inconclusive. We found that while Synthetic User interaction may influence perceptions of the persona in terms of credibility, clarity, and likeability, it neither significantly enhances a designer’s empathy toward the persona nor the quantity or diversity of ideation outcomes. Building on these findings, we explore further to find potential explanations and implications of our results.

### “Blurring” of designers’ perceptions of personas

As we defined in the Background, the notion of “Synthetic Users” aims to generate relevant information about the user group they represent and stimulate insights in the designer. In this section, we attempt to understand the designers’ perception of the Synthetic User, from the level of basic Persona summaries up to the level of bona-fide clients.

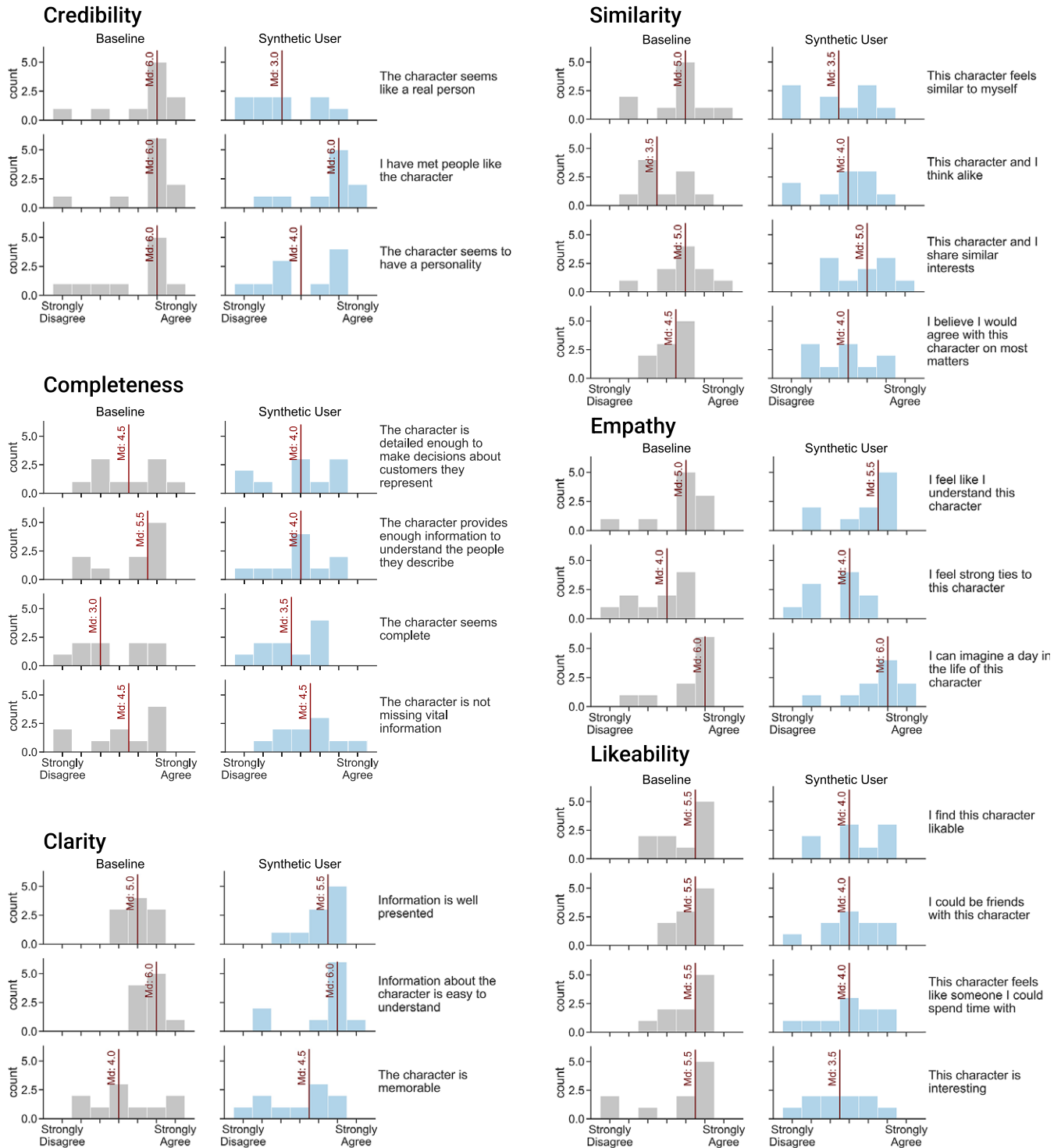
The results from the PPP, as detailed in the Analysis section, indicate a significant difference between the two conditions, with the Synthetic User condition scoring lower in perceived credibility (see Table 1), as evidenced by responses to *This character seems like a real person*. However, it is unclear if this response is due to the quality of the participants’ interactions with the chatbot or their knowledge that they were interacting with one since we made no attempt to conceal the fact that they were interacting with a chatbot.

Examining differences in participants’ think-aloud verbalizations as measured via the LIWC category scores showed lower scores in the *social* and *insight* categories for the Synthetic User condition.

The LIWC categories, which were intended to support or explain self-reported results, did not yield clear insights. Only the *Social* and *Insight* categories showed significant results, both recording lower scores in the Synthetic User scenario. This suggests that these linguistic traits, which are associated with empathy and understanding, were less prominent. This leads us to conclude that this representation of the Synthetic User may result in a lower empathy and understanding compared to a static persona representation.

Post-study interviews support these results of lowered empathy and understanding, evidenced by complaints and higher average self-reported frustration in the TLX index of the Synthetic User condition. It is possible that the prejudice toward Personas as a technique compounded the resistance to interacting with the Synthetic User. We mention *Champion Bias* in the Background section, where advocates of personas tend to get better results from them. In our study, we observed the converse to also be the case: those who disliked personas tended to get less out of them. For instance, participants P9 and P17 from the Synthetic User condition had negative views about personas in general. P9 admitted, *I don’t like personas, which I mentioned earlier. That made it harder to get into the flow of the task*. P17 had stronger views, stating, *Personally, I find personas annoying. I understand the idea, but I feel like it’s an oversimplification. Users often use products in ways or by groups that designers never anticipated. Personas can play on stories and assumptions without questioning where they come from. I try not to stereotype. I understand the need for a focal point, but I’ve always felt it’s limiting*. P23 had a more moderate, though similarly negative view, noting *I’ve never really used personas that much in my design process often because I have an ideal person already in mind*.

However, we only found a weak correlation (one that was not statistically significant) between participants’ views about personas in general and their level of agreement on whether the persona had sufficient information, both reported on a 7-point Likert Scale (Kendall’s  $\tau = -0.03, p > 0.05$ ). This weak correlation could partly be due to their perception that the Synthetic User was often not forthcoming. For instance, P31 – one of the participants who considered personas as a valuable method – was unhappy with the lack of strong responses to their questions, stating (of the chatbot persona) *Some of her answers were wishy-washy. Like her interest in cooking; she responded with “not really” to many things*. P10 had a similar reaction to the perceived caginess of the persona: *You have a person that doesn’t want to give anything away and ... it doesn’t feel nice if you want to get to the bottom of something*.

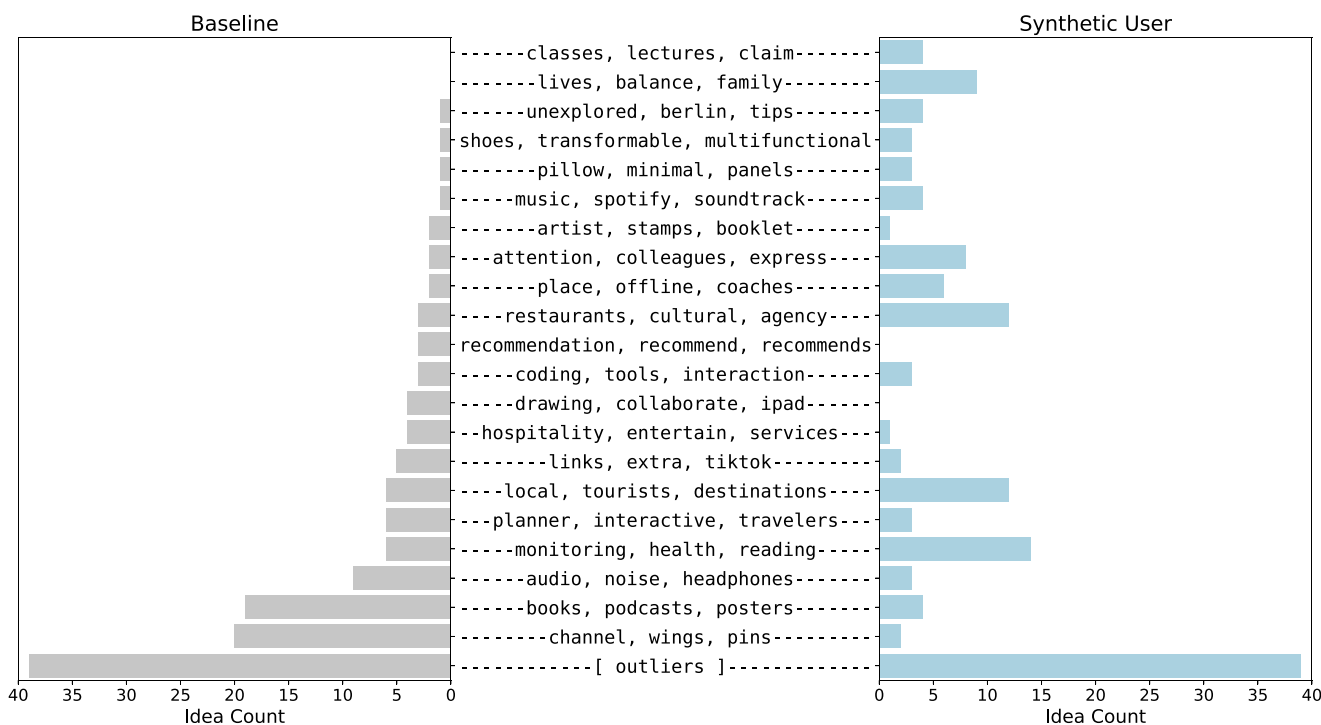


**Figure 5.** Participant responses to each item on the Persona Perception Scale (PPS) (Salminen et al., 2020b) grouped by category and separated by condition (Baseline vs. Chatbot). Median values are shown for each category. See also Table 1: for details.

Such impressions can be puzzling, as an effective persona needs to be *deliberately shallow*, being a simplified representation of user segments (Grudin and Pruitt, 2002). However, the designers do not make the same allowance for the Synthetic User as they do for the persona summary. This seeming inconsistency leads us to believe that the conversational interface of a Synthetic User created an expectation of humanistic behavior, which was not fulfilled for many of the participants. This expectation was evident through

the way participants formulated questions with the Synthetic User, suggesting they viewed it as a potential client, aligning with design facilitation principles (Martin et al., 2012), which emphasize open-ended questions to gain further insights into what motivates consumers.

Despite instructions by the researcher to treat the Synthetic User as a representative of a target user group, it seems the interactive nature of the Synthetic User blurred this distinction for some



**Figure 6.** The distribution of ideas, as determined by topic modeling (using BERTopic). The top three keywords are shown for each topic. The Baseline condition seems to span a slightly broader range of unique topics compared to the Synthetic User Condition, but the distribution of ideas across the topics appears more spread out in the Synthetic User condition than the Baseline.

designers. For example, P17, who expressed significant enjoyment during the study, consistently personalized the Synthetic User in their ideation, referring to it as “Natalie” and tailoring solutions around her responses:

- A product to assist *Natalie* in finding a mentor/mentee support group.
- A service to help *Natalie* coordinate outings with friends based on mutual interests.
- Recommendations for *Natalie* to explore activities during work travels.
- Solutions tailored for activities *Natalie* and her children would both enjoy.

Although anthropomorphism in chatbot and robotics design is typically employed to foster empathy (Darling, 2017; Janson, 2023) – manifested as increased personification – the personification observed here diverged from our intended use of Personas as mere representative archetypes.

On the flip side, some participants were openly reproofing and blunt with the Synthetic User, seemingly taking advantage of the fact that the chatbot had no feelings to hurt. For instance, P41’s critique, *Natalie, that feels weak*, or an instance from a pilot session where the participant probed into the chatbot’s personal relationships (e.g., asking the chatbot *Would you cheat on your boyfriend (on a trip)?*) were observed. Participants acknowledged that they posed questions to the Synthetic User they wouldn’t typically ask a human. This aligns with Vandenberg’s assertion that chatbot representations of potential users can be helpful due to their ability to withstand mistreatment without consequences (Vandenberghe, 2017).

In summary, this study’s findings highlight a significant challenge in the designers’ perception and interaction with Synthetic Users. Results imply that the same information as a persona

summary, when presented as a Synthetic User, may not be seen as sufficient, and the interactive component seems to increase participants’ expectation of information from the persona. We also observed a range of perception responses, from skepticism to reproof, toward the Synthetic User. This indicates a need for re-thinking the design and implementation of Synthetic Users as qualitatively different from personas in order to effectively portray end users.

### Impact on ideation

As highlighted in the [Analysis section](#), participants interacting with the Synthetic User were more motivated than their Baseline counterparts to use up all the available time for ideation. This is likely attributed to the ability to continuously query for potentially new information from the Synthetic User, whereas the Persona Summary remains static.

However, the Baseline group displayed greater consistency in idea fluency, which can possibly be explained by fewer instances of cognitive switching between querying and ideation compared to the continuous conversation with the Synthetic User.

Both groups produced a comparable number and diversity of ideas, as analyzed using BERTopic. However, the Synthetic User condition introduced two unique interaction mechanics: the capability to pose follow-up questions and to “pitch” ideas for feedback. Every participant in the Synthetic User group engaged in pitching their ideas to the Synthetic User for feedback. For instance, one participant refined the basic concept of “Tinder for museums” – a common idea among the participants across conditions – into a more nuanced idea through dialogue with the Synthetic User.

Participant: Would you like to go on museum dates with strangers?

Synthetic User: I'm not sure, it would depend on the situation. I think it could be fun to go with new people that I don't know, but if it's a group of strangers I don't know how well it would work out.

Participant: What if they're all museum nerds? (*sic*)

Synthetic User: Then I would be in good company! I'm definitely interested in art and history, but don't consider myself a museum nerd.

Participant proposes the idea 'Museum Social App'

Interactions with the Synthetic User often led participants down new ideation pathways. For instance, in one interaction, the Synthetic User's mention of having a cat inspired a participant to conceive a "Cat food dispenser with water, synchronized with an app."

Participant: Do you have any pets?

Synthetic User: I have a cat, but I also grew up with dogs and horses, so I love animals in general.

Participant: What does your cat do when you're on travel?

Synthetic User: My cat is usually taken care of by my parents, but I do try to call them often to check-in. They usually just keep her inside and feed her, but she's used to it since I'm often traveling for work.

Participant: is your cat an indoor or outdoor cat?

Synthetic User: My cat is an indoor cat.

(Participant idea: 'Cat food dispenser including water synced with an app')

In summary, though interactions with the Synthetic User departed from what was prescribed to the participants, we observed instances where the deviation likely increased their motivation for sustained ideation, facilitating unique idea development through continuous information queries and idea pitching. This differed from the Baseline group, which exhibited more fluent ideation but a tendency to cease when ideas were exhausted. While the Synthetic User's interactivity effectively aided in refining and inspiring new concepts, it may necessitate participants to oscillate between analytical thinking (gathering relevant information) and creative thinking (generating ideas), potentially resulting in an increased cognitive load. Research on working memory, cognitive load, and creativity (Redifer et al., 2019) has shown that increased cognitive load is a predictor of a decrease in creative outcome and may explain the reduced ideation fluency in the Synthetic User condition.

Again, this appears to indicate that the approach to designing with a Synthetic User cannot simply be transplanted from a corresponding approach with a traditional persona summary. The notion of the Synthetic User as an "interactive persona" may need to be modified to one that is closer to a stakeholder, borrowing methods from participatory design.

### Synthetic User inconsistencies

Synthetic Users are expected to maintain a consistent character and provide accurate information. Unlike traditional Persona Summaries, they sometimes exhibit inconsistencies that can be particularly jarring for designers using them.

### LLM inconsistencies

Predominantly, these challenges stem from LLM hallucinations, referring to the generation of information that isn't rooted in its training data or deviates from the expected persona data, as documented in

studies by Scao et al. (2022); Radford et al. (2019); Bang et al. (2023). LLMs are shown to have limitations in analytical and numerical reasoning (Liu et al., 2022) and a tendency to produce hallucinated statements as facts (Rawte et al., 2023).

For Synthetic Users, this can disrupt the illusion of an effective persona in stimulating empathy and creativity. In our study, this manifested as the chatbot presenting conflicting details about herself, potentially introducing fixation and disrupting the ideation process. Informing users about these potential pitfalls can help manage expectations. For instance, participant P41 observed: *There were a couple of inconsistencies, but you told me they were going to be there, so I wasn't so thrown by it.*

On the other hand, we also observed certain details that remained broadly consistent in the output from the Synthetic User. While certain details varied (e.g., its birthplace), its current residence in Germany remained constant. This ability of the Synthetic User to introduce new persona details, while adhering to its foundational data, is an important feature. This consistent core is crucial for Synthetic Users to be effective as a dynamic yet consistent representation of user segments aids designers' understanding of the user group without misleading them.

### Synthetic User limitations

LLMs don't self-initiate deeper exploration unless directly prompted. For Synthetic Users, this manifests as "shallow conversation", marked by a noticeable absence of conversational depth or user insights. The following example showcases such a shallow conversation, likely influenced by the open-ended nature of questioning that designers are trained to employ. The Synthetic User, however, doesn't meet the expectation of proactively diving deeper and responds to probing inquiries with surface-level answers.

Participant: How do you truly feel about creating novel items?

Synthetic User: It's invigorating! Witnessing reactions to fresh creations and endeavoring to craft unseen designs are aspects I cherish.

Participant: Why do you relish observing people's reactions?

Synthetic User: Observing reactions is intriguing. It's gratifying to believe I can pleasantly surprise them and that they derive as much joy from my creations as I do.

Participant: Why is evoking positive surprise in others pivotal for you?

Synthetic User: Offering pleasant surprises contributes to people's happiness. It's fulfilling to be a source of their joy.

...

(... several exchanges between the participant and the Synthetic User follow where the latter responds with similar words like "gratifying", "stimulating", etc. to questions about motivation without offering more detail.)

...

Participant: Why emphasize unprecedented designs when your inclination is towards people's happiness? Aren't they sometimes divergent aims?

Synthetic User: Introducing people to innovative designs can be a source of their happiness. The novelty often excites them, and that joy is something I cherish.

The presence of conversational dead-ends highlights the need for a more refined design and clearer instructions for Synthetic Users. Simply animating a general LLM with personality data is not enough to establish realistic nuances expected by the designer.

Conversational shallowness can also inadvertently lead designers to stereotype. When the synthetic stakeholder merely affirms designers' assumptions without delving deeper, it can falsely boost their confidence, as evidenced by the increased confidence metrics we observed through LIWC, which showed higher values in the Synthetic User condition for the categories of *allnone* ( $M = 1.00, SD = 0.44$ ) and *certitude* ( $M = 1.37, SD = 0.76$ ) compared to the Baseline condition ( $M = 0.71, SD = 0.30$  and  $M = 0.93, SD = 0.36$  respectively). These assumptions are evident in participants' statements such as this one by P9: *I think the problem with her is that she's superficial. She doesn't have depth. There are lots of people like this, especially here in (city name)*. Similarly, P31's conclusion on the persona was, *She's a millennial. She likes to do things like I do, such as going to a museum ... I didn't ask her what her job was, but I believe I could guess*. Combined with the lower metric of *insight* ( $M = 3.94, SD = 1.05$ ) and *causation* ( $M = 1.45, SD = 0.57$ ) in Synthetic Users compared to the Baseline condition ( $M = 4.98, SD = 0.59$  and  $M = 1.57, SD = 0.29$  respectively), indicating a diminished process of reappraisal, this reveals participants felt more confident in their understanding. However, this confidence might have been based on shallow or incorrect assumptions, as illustrated by participant statements that reinforced stereotypes.

### Summary

Our study suggests that the utility of integrating Synthetic Users in ideation is as a complement to static personas, rather than a replacement. By adding an interactivity to the personas, the Synthetic Users do influence designers' perceptions, but not in the intended outcome of enhancing empathy, with lower credibility and empathy scores. Synthetic Users encourage sustained ideation but possibly increase cognitive load, contrasting with the shorter but more consistent ideation flow provided by static personas. Designers' interactions with Synthetic Users varied, ranging from anthropomorphizing to critical detachment. These insights suggest a need for the design and implementation of Synthetic Users to optimize their effectiveness in the design process.

### Reflection on the "Designerly Use" of Synthetic Users

Our study critically assesses the use of Synthetic Users in design ideation. We found that while a minimally designed Synthetic User did not significantly enhance empathy or understanding, we identified design requirements to better integration into the design process and fit with designers' perception and utilization.

We pose a conjecture that the traditional abductive nature of design thinking, as described by Cross (1982) may conflict with the structured interactions required by LLM-powered Synthetic Users. Designers are trained to conduct interviews with users in a flexible structure that allows for open-ended exploration, and to be careful to not influence the interviewee's mental processes; they are encouraged to employ "neutral prompts" in order to be non-leading and non-directive, such as "Tell me about ..." and "How do you feel about ..." (Wilson, 2013).

This creates a dichotomy in thinking modes – the intuitive, emergent thinking of designers versus the deliberate, structured prompting needed to effectively engage with LLMs (Cross, 2023). Our study results suggest that this clash may limit the effectiveness of Synthetic Users in iterative, constructive dialogues, essential in the design process.

While designers are accustomed to rapidly switching between different modes of thinking – from highly creative ideation to practical

problem-solving (Cross, 2023), interaction with Synthetic Users potentially demands an additional mode of thinking, that is, prompt design. At the time of the study, participants lacked familiarity with this technology, resulting in a learning curve similar to observations made by Zamfirescu-Pereira et al. (2023), where nonexperts struggle with understanding and effectively designing prompts for instructing AI. The study indicates that exposure and experience allowed users to develop the skillset. Consequently, we anticipate improved participant proficiency in future iterations of our Synthetic Users study.

Additionally, designers expect a level of contribution that the Synthetic User in the study struggled to meet. LLM-powered Synthetic Users are not yet capable of engaging in the kind of iterative, constructive dialogue that human interactions typically offer. We noted that Synthetic Users were limited in building upon ideas or providing the kind of constructive feedback and iteration crucial in the design process. Their responses often merely expanded on the designers' queries, using the present tokens as a foundation for their responses.

Our study reveals a dichotomy in the use of LLMs by designers, necessitating a reevaluation of how Synthetic users in particular, and generative AI in general, are utilized within the framework of design thinking. This dichotomy, characterized by different cognitive modes, highlights the need for further exploration to fully integrate LLMs into the design process.

Moreover, we observed designer biases toward Synthetic Users, similar to the *Champion Bias* we noted for personas. An added bias that did not exist for the analog, some designers exhibited a *Personhood Bias*. Addressing this by setting clear expectations about the non-human nature of Synthetic Users could help streamline interactions and mitigate unproductive personification.

Expanding on the original observations on "off-the-shelf users" by Vandenberghe (2017), we propose additional benefits of Synthetic Users. For instance, Synthetic Users could potentially preserve the privacy of real human counterparts by leveraging the intrinsic data abstraction of LLMs, which learn patterns without retaining specifics, potentially lessening the need for exhaustive real-user data. Challenges such as LLM inconsistencies as outlined in the Discussion section could be addressed in a few ways. Arzberger et al. (2022) – in their study of pairs of designers interacting with a text generator powered by an LLM – observed that insights were obtained when the designers (a) used the output from the text generator to expand one's understanding of the problem, and (b) used seeming ambiguities in the output as a stimulus to explore the situation further. This approach underscores the potential of Synthetic Users to broaden designers' perspectives. By leveraging their inherent creative thinking, designers can interpret and use AI outputs in innovative ways, potentially leading to more diverse and enriched design outcomes.

### Limitations and future work

Our study, exploring Synthetic Users powered by LLMs, did not demonstrate significant improvements in design outcomes or enhanced insight into user groups. However, it is important to note that our research was limited by a small sample size and the use of only one persona, which may limit the generalizability of our findings. Additionally, the within-subjects design, while advantageous in controlling for individual differences, introduced the potential for the Pygmalion effect – whereby high expectations lead to better performance and vice versa – and limited our ability to compare individual user experiences across different conditions

effectively. Future research should address these limitations by involving larger participant groups and incorporating multiple personas to obtain more comprehensive insights into how Synthetic Users influence designers' understanding, processes, and outcomes. Moreover, conducting open-ended, longitudinal studies that span the full lifecycle of a design project could more accurately capture the complexity of real-world design scenarios and interactions with Synthetic Users.

To further explore the potential of Synthetic Users, extending their application beyond initial brainstorming sessions to other stages of the design process, such as concept validation and usability studies, might reveal additional applications and benefits. Additionally, further examination of Synthetic Users from a psychological standpoint is necessary to gauge their capabilities in grasping the intentions and thought processes of both designers and the users they aim to represent. Additionally, examining the consistency of AI-generated personas across various contexts could make Synthetic User applications more relevant, as well as developing evaluation methods tailored to Synthetic Users, beyond subjective human benchmarks, to accurately assess their effectiveness and utility. Achieving this level of understanding is vital for Synthetic Users to effectively contribute to design contexts.

It is important to approach the evaluation of Synthetic Users with caution. As highlighted by the work of Raji et al. (2021), applying tests designed for humans to benchmark or evaluate AI models may not be appropriate. Such benchmarks often mistakenly equate linguistic skills with broader commonsense reasoning and genuine comprehension. This insight validates our decision to apply LIWC only to data generated by humans rather than responses from Synthetic Users and underscores the need for developing specific methodologies for assessing the effectiveness and impact of Synthetic Users.

Our vision for future Synthetic User dialogues encompasses a range of domains, modalities, and levels of expressiveness, approaching the complexity of holistic human interactions or user studies. This study, therefore, lays the groundwork for meaningful future research aimed at refining Synthetic User capabilities and exploring their broader impact in various design contexts.

## Conclusion

This study has embarked on an exploratory journey to understand the potential and limitations of Synthetic Users powered by LLMs in the design process. Our findings have revealed the nuanced ways in which designers interact with and perceive these particular AI-driven tools. While the study did not conclusively demonstrate that Synthetic Users significantly enhance empathy or ideation diversity, it has provided valuable insights into the cognitive and psychological dynamics at play in designer-AI interactions.

We observed that while Synthetic Users can stimulate extended engagement in ideation, they do not necessarily translate into greater ideation fluency or diversity compared to traditional persona summaries. This underscores the complexity of the design process and the challenges of integrating AI into creative workflows. The study also revealed biases and expectations among designers that affect their interaction with Synthetic Users, highlighting the need for clearer guidelines and better design of these tools to align with cognitive processes present in design thinking.

Looking ahead, our research points to several key areas for future development. These include expanding the participant base and personas for more comprehensive insights, enhancing the

technical capabilities of Synthetic Users for greater realism/credibility, and exploring their application across various stages of the design process. Perhaps most critically, the study suggests the need for Synthetic Users that are more human-like in their interactions, aware of the expectations placed upon them, fostering a more natural and productive collaboration with designers.

This research represents just the beginning of a broader inquiry into the role of AI in the design process. As technology evolves, so too will the ways in which designers and AI tools co-create. The journey ahead is filled with possibilities and challenges, and we look forward to continued exploration and discovery in this dynamic and exciting field.

**Data availability.** The data from the study is not made public due to the conditions of the University Ethics Approval.

**Funding statement.** This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Competing interest.** The author(s) declare none.

## References

- Adamopoulou E and Moussiades L (2020) Chatbots: History, technology, and applications. *Machine Learning with Applications* 2, 100006. doi: 10.1016/j.mlwa.2020.100006.
- An J, Kwak H, Jung S, Salminen J and Jansen BJ (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), 1–19. doi: 10.1007/s13278-018-0531-0.
- Arzberger A, van der Burg V, Chandrasegaran S, and Lloyd P (2022) Triggered: Using human-AI dialogue for problem understanding in collaborative design. In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Volume 6: 34th International Conference on Design Theory and Methodology (DTM)*, American Society of Mechanical Engineers, V006T06A038. doi: 10.1115/DETC2022-89273.
- Athota L, Shukla VK, Pandey N and Rana A (2020) Chatbot for healthcare system using artificial intelligence. In *IEEE International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*. Piscataway, NJ: IEEE, 619–622. doi: 10.1109/ICRITO48877.2020.9197833.
- Bang Y, Cahyawijaya S, Lee N, Dai W, Su D, Wilie B, Lovenia H, Ji Z, Yu T, Chung W, et al. (2023) A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, 675–718. doi: 10.18653/v1/2023.ijcnlp-main.45.
- Blythe MA and Wright PC (2006) Pastiche scenarios: Fiction as a resource for user-centred design. *Interacting with Computers* 18(5), 1139–1164.
- Boyd RL, Ashokkumar A, Seraj S and Pennebaker JW (2022) *The Development and Psychometric Properties of LIWC-22*. Austin, TX: University of Texas at Austin. <https://liwc.app>.
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. (2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33, 1877–1901.
- Calabretta G and Kleinsmann M (2017) Technology-driven evolution of design practices: Envisioning the role of design in the digital era. *Journal of Marketing Management* 33(3–4), 292–304. doi: 10.1080/0267257X.2017.1284436.
- Cooper A (1999) The inmates are running the asylum. In Udo A, Edmund E and Knut P (eds.), *Software-Ergonomie '99: Design von Informationswelten*. Wiesbaden, Germany: Vieweg+Teubner Verlag, 17. doi: 10.1007/978-3-322-99786-9\_1.
- Cross N (1982) Designerly ways of knowing. *Design Studies* 3(4), 221–227. doi: 10.1016/0142-694X(82)90040-0.

- Cross N.** (2023) *Design Thinking: Understanding How Designers Think and Work*. London, UK: Bloomsbury Publishing.
- Daalhuizen J, Timmer R, van der Welie M and Gardien P** (2019) An architecture of design doing: A framework for capturing the ever-evolving practice of design to drive organizational learning. *International Journal of Design* 13(1), 37–52.
- Abhishhek D and Jyoti K** (2021) Effect of persona stereotyping on design solutions: Observations from an empirical study with novice designers. In Chakrabarti A, Poovaiah R, Bokil P and Kant V (eds.), *Design for Tomorrow—Volume 3*. Singapore: Springer, 175–185. doi: 10.1007/978-981-16-0084-5\_14.
- Darling K** (2017) “Who’s Johnny?” Anthropomorphic framing in human–robot interaction, integration, and policy. In Lin P, Abney K and Jenkins R (eds.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. New York, NY: Oxford University Press, 173–188. doi: 10.1093/oso/9780190652951.003.0012.
- Dortheimer J, Martelaro N, Sprecher A and Schubert G.** (2024) Evaluating large-language-model chatbots to engage communities in large-scale design projects. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 38, e4. doi: 10.1017/S0890060424000027.
- Drego V and Dorsey M** (2010) The ROI of personas. <https://www.forrester.com/report/The-ROI-Of-Personas/RES55359>.
- Gaudry S** (2022) Tool - personas: Older adults. <https://policycommons.net/artifacts/2244118/tool-personas/3002263/>.
- Giarelis N, Kanakaris N and Karacapilidis N** (2021) A comparative assessment of state-of-the-art methods for multilingual unsupervised keyphrase extraction. In Maglogiannis I, Macintyre J and Iliadis L (eds.), *Artificial Intelligence Applications and Innovations*, volume 627 of *IFIP Advances in Information and Communication Technology*. Cham: Springer, 635–645. doi: 10.1007/978-3-030-79150-6\_50.
- Gibson J, Malandrakis N, Romero F, Atkins DC and Narayanan SS** (2015) Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In *Sixteenth Annual Conference of the International Speech Communication Association*, 1947–1951. doi: 10.21437/interspeech.2015-429.
- Grootendorst M** (2020) KeyBERT: Minimal keyword extraction with BERT. <https://doi.org/10.5281/zenodo.4461265>.
- Grootendorst M** (2022) BERTopic: Neural topic modeling with a class-based TF-IDF procedure. <https://arxiv.org/abs/2203.05794>.
- Grudin J and Pruitt J** (2002) Personas, participatory design and product development: An infrastructure for engagement. In Binder T, Gregory J and Wagner I (eds.), *Proceedings of the Participatory Design Conference*, vol. 2, 144–152.
- Ha J, Jeon H, Han D, Seo J and Oh C** (2024) CloChat: Understanding how people customize, interact, and experience personas in large language models. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 1–24. doi: 10.1145/3613904.3642472.
- Haller E and Rebedea T** (2013) Designing a chat-bot that simulates an historical figure. In *IEEE International Conference on Control Systems and Computer Science*. Piscataway, NJ: IEEE, 582–589. doi: 10.1109/CSCS.2013.85.
- Hämäläinen P, Tavast M and Kunnari A** (2023) Evaluating large language models in generating synthetic HCI research data: a case study. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 1–19. doi: 10.1145/3544548.3580688.
- Hart SG and Staveland LE** (1988) Development of NASA-TLX (task load index): Results of empirical and theoretical research. In Hancock PA and Meshkati N (eds.), *Human Mental Workload*, vol. 52 of *Advances in Psychology*. Amsterdam: Elsevier, 139–183. doi: 10.1016/S0166-4115(08)62386-9.
- Heck J, Rittiner F, Meboldt M and Steinert M** (2018) Promoting user-centricity in short-term ideation workshops. *International Journal of Design Creativity and Innovation* 6(3-4), 130–145. doi: 10.1080/21650349.2018.1448722.
- Hill CG, Haag M, Oleson A, Mendez C, Marsden N, Sarma A and Burnett M** (2017) Gender-inclusiveness personas vs. stereotyping: Can we have it both ways? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6658–6671.
- Howard A and Borenstein J** (2018) The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics* 24(5), 1521–1536.
- Jansen BJ, Jung S and Salminen J** (2019) Capturing the change in topical interests of personas over time. *Proceedings of the Association for Information Science and Technology* 56(1), 127–136.
- Jansen BJ, Salminen JO and Jung S** (2020) Data-driven personas for enhanced user understanding: Combining empathy with rationality for better insights to analytics. *Data and Information Management* 4(1), 1–17.
- Jansen BJ, Salminen J, Jung S and Guan K** (2021) The data-driven persona revolution. In *Data-Driven Personas*. Cham: Springer, 3–27. doi: 10.1007/978-3-031-02231-9\_1.
- Janson A** (2023) How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification. *Computers in Human Behavior* 149, 107954.
- Jung S, Salminen J and Jansen BJ** (2019) Personas changing over time: Analyzing variations of data-driven personas during a two-year period. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6.
- Kelley T** (2018) Build your creative confidence: 30 circles exercise. <https://www.ideo.com/blog/build-your-creative-confidence-thirty-circles-exercise>. Retrieved 14 Sep 2022.
- Lee D-K and Na S-Y** (2018) Present the Celeb-Bot model using artificial intelligence. *The Journal of the Korea institute of electronic communication sciences* 13(4), 765–776.
- Lester B, Al-Rfou R and Constant N** (2021) The power of scale for parameter-efficient prompt tuning. In Moens M-F, Huang X, Specia L and tau Yih SW (eds.), *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 3045–3059. doi: 10.18653/v1/2021.emnlp-main.243.
- Leviathan Y and Matias Y** (2018) Google Duplex: an AI system for accomplishing real-world tasks over the phone. <https://research.google/pubs/google-duplex-an-ai-system-for-accomplishing-real-world-tasks-over-the-phone/>.
- Li J, Galley M, Brockett M, Spithourakis G, Gao J and Dolan B** (2016) A persona-based neural conversation model. In Erk K and Smith NA (eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Berlin, Germany, 994–1003. doi: 10.18653/v1/P16-1094.
- Liu Y, Wei W, Liu J, Mao X, Fang R and Chen D** (2022) Improving personality consistency in conversation by persona extending. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. New York, NY: Association for Computing Machinery, 1350–1359. doi: 10.1145/3511808.3557359.
- Logan R IV, Balažević I, Wallace E, Petroni F, Singh S and Riedel S** (2022) Cutting down on prompts and parameters: Simple Few-Shot learning with language models. In *Findings of the Association for Computational Linguistics: ACL 2022*. ACL, 2824–2835.
- Marsden N, Pröbster M, Haque ME and Hermann J** (2017) Cognitive styles and personas: designing for users who are different from me. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, 452–456.
- Martin B, Hanington B and Hanington BM** (2012) *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Beverly, MA: Rockport Pub.
- Mascaro JS, Palmer PK, Willson M, Ash MJ, Florian MP, Srivastava M, Sharma A, Jarrell B, Walker ER, Kaplan DM, et al.** (2023) The language of compassion: Hospital chaplains’ compassion capacity reduces patient depression via other-oriented, inclusive language. *Mindfulness* 14, 2485–2498.
- Matthews T, Judge T and Whittaker S** (2012) How do designers and user experience professionals actually perceive and use personas? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1219–1228. doi: 10.1145/2207676.2208573.
- McGinn J and Kotamraju N** (2008) Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1521–1524. doi: 10.1145/1357054.1357292.
- Motoki F, Neto VP, and Rodrigues V** (2024) More human than human: Measuring ChatGPT political bias. *Public Choice* 198(1), 3–23.
- Mulder S and Yaar Z** (2006) *The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web*. Berkeley, CA: New Riders.
- Neate T, Bourazeri A, Roper A, Stumpf S and Wilson S** (2019) Co-created personas: Engaging and empowering users with diverse needs within the

- design process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.
- Nielsen L** (2019) Personas in use. In *Personas-User Focused Design*. Cham: Springer, 83–115.
- Nielsen L and Hansen KS** (2014) Personas is applicable: A study on the use of personas in Denmark. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 1665–1674. doi: [10.1145/2556288.2557080](https://doi.org/10.1145/2556288.2557080).
- Nielsen L, Hansen KS, Stage J and Billestrup J** (2015) A template for design personas: analysis of 47 persona descriptions from Danish industries and organizations. *International Journal of Sociotechnology and Knowledge Development* 7(1), 45–61.
- Paoli SD** (2023) Improved prompting and process for writing user personas with LLMs, using qualitative interviews: Capturing behaviour and personality traits of users. <https://arxiv.org/abs/2310.06391>.
- Pruitt J and Grudin J** (2003) Personas: Practice and theory. In *Proceedings of the ACM Conference on Designing for User Experiences*. New York, NY: Association for Computing Machinery, 1–15. doi: [10.1145/997078.997089](https://doi.org/10.1145/997078.997089).
- Qian Q, Huang M, Zhao H, Xu J and Zhu X** (2018) Assigning personality/profile to a chatting machine for coherent conversation generation. In *IJCAI*, 4279–4285.
- Qiu H and Tay D** (2023) A mixed-method comparison of therapist and client language across four therapeutic approaches. *Journal of Constructivist Psychology* 36(3), 337–360.
- Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al.** (2019) Language models are unsupervised multitask learners. *OpenAI blog* 1(8), 9.
- Raji ID, Denton E, Bender EM, Hanna A and Paullada A** (2021) AI and the everything in the whole wide world benchmark. In *Conference on Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks*. pp.17
- Rawte V, Sheth A and Das A** (2023) A survey of hallucination in large foundation models. <https://arxiv.org/abs/2309.05922>.
- Redifer JL, Bae CL and DeBusk-Lane M** (2019) Implicit theories, working memory, and cognitive load: Impacts on creative thinking. *Sage Open* 9(1), 1–16. doi: [10.1177/2158244019835919](https://doi.org/10.1177/2158244019835919).
- Reynolds L and McDonell K** (2021) Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7.
- Roussou M, Katifori A, Pujol L, Vayanou M and Rennick-Egglestone SJ** (2013) A life of their own: Museum visitor personas penetrating the design lifecycle of a mobile experience. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 547–552.
- Rozado D** (2023) The political biases of ChatGPT. *Social Sciences* 12(3), 148.
- Salminen J, Jansen BJ, An J, Kwak H and Jung S** (2019a) Automatic persona generation for online content creators: Conceptual rationale and a research agenda. *Personas-User Focused Design*, 135–160.
- Salminen J, Jung S, An J, Kwak H, Nielsen L and Jansen BJ** (2019b) Confusion and information triggered by photos in persona profiles. *International Journal of Human-Computer Studies* 129, 1–14. doi: [10.1016/j.ijhcs.2019.03.005](https://doi.org/10.1016/j.ijhcs.2019.03.005).
- Salminen J, Guan K, Jung S, Chowdhury SA and Jansen BJ** (2020a) A literature review of quantitative persona creation. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 1–14. doi: [10.1145/3313831.3376502](https://doi.org/10.1145/3313831.3376502).
- Salminen J, Santos JM, Kwak H, An J, Jung S and Jansen BJ** (2020b) Persona Perception Scale: Development and exploratory validation of an instrument for evaluating individuals' perceptions of personas. *International Journal of Human-Computer Studies* 141, 102437. doi: [10.1016/j.ijhcs.2020.102437](https://doi.org/10.1016/j.ijhcs.2020.102437).
- Salminen J, Guan K, Jung S and Jansen BJ** (2021a) A survey of 15 years of data-driven persona development. *International Journal of Human-Computer Interaction* 37(18), 1685–1708. doi: [10.1080/10447318.2021.1908670](https://doi.org/10.1080/10447318.2021.1908670).
- Salminen J, Jung S, Santos JM, Kamel AMS and Jansen BJ** (2021b) Picturing it!: The effect of image styles on user perceptions of personas. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 1–16.
- Salminen J, Guan KW, Jung S and Jansen B** (2022) Use cases for design personas: A systematic review and new frontiers. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 1–21. doi: [10.1145/3491102.3517589](https://doi.org/10.1145/3491102.3517589).
- Samrose S, Anbarasu K, Joshi A and Mishra T** (2020) Mitigating boredom using an empathetic conversational agent. In *Proceedings of the ACM International Conference on Intelligent Virtual Agents*. New York, NY: Association for Computing Machinery, 1–8.
- Scao TL, Fan A, Akiki C, Pavlick E, Ilić S, Hesslow D, Castagné R, Luccioni AS, Yvon F, Gallé M, Tow J, Rush AM, Biderman S, Webson A, Ammanamanchi PS, Wang T, Sagot B, Muennighoff N, del Moral AV, Ruwase O, Bawden R, Bekman S, Mcmillan-Major A, Beltagy I, Nguyen H, Saulnier L, Tan S, Suarez PO, Sanh V, Laurençon H, Jernite Y, Launay J, Mitchell M, Raffel C, et al.** (2022) Bloom: A 176b-parameter open-access multilingual language model. <https://inria.hal.science/hal-03850124>.
- Schmidt A, Elagrouty P, Draxler F, Kreuter F and Welsch R** (2024) Simulating the human in HCD with ChatGPT: Redesigning interaction design with AI. *Interactions* 31(1), 24–31.
- Schmitt O and Buschek D** (2021) Characterchat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot. In *Creativity and Cognition*, 1–10.
- Schneidewind L, Hörold S, Mayas C, Krömker H, Falke S and Pucklitsch T** (2012) How personas support requirements engineering. In *IEEE International Workshop on Usability and Accessibility Focused Requirements Engineering (UsARE)*, 1–5. doi: [10.1109/UsARE.2012.6226786](https://doi.org/10.1109/UsARE.2012.6226786).
- Song H, Zhang W, Cui Y, Wang D and Liu T** (2019) Exploiting persona information for diverse generation of conversational responses. In *International Joint Conference on Artificial Intelligence*, 5190–5196.
- Song H, Wang Y, Zhang K, Zhang W and Liu T** (2021) Bob: BERT over BERT for training persona-based dialogue models from limited personalized data. In Zong C, Xia F, Li W and Navigli R (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, 167–177. doi: [10.18653/v1/2021.acl-long.14](https://doi.org/10.18653/v1/2021.acl-long.14).
- Stoimenova N and Kleinsmann M** (2020) Identifying and addressing unintended values when designing (with) artificial intelligence. In Boess S, Cheung M and Cain R (eds.), *Synergy - DRS International Conference 2020*. Design Research Society. doi: [10.21606/drs.2020.222](https://doi.org/10.21606/drs.2020.222).
- Stolterman E** (2021) The challenge of improving designing. *International Journal of Design* 15(1), 65.
- Tausczik YR and Pennebaker JW** (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29(1), 24–54.
- Turner P and Turner S** (2011) Is stereotyping inevitable when designing with personas? *Design Studies* 32(1), 30–44.
- Vandenberghe B** (2017) Bot personas as off-the-shelf users. In *Proceedings of the ACM CHI Conference Extended Abstracts on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 782–789.
- Wilson C** (2013) *Interview Techniques for UX Practitioners: A User-Centered Design Method*. Oxford, UK: Newnes.
- Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, Jurafsky D, Szolovits P, Bates DW, Abdulnour REE, et al.** (2024) Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *The Lancet Digital Health* 6(1), e12–e22.
- Zamfirescu-Pereira JD, Wong RY, Hartmann B and Yang Q** (2023) Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, 1–21.
- Zhang S, Dinan E, Urbanek J, Szlam A, Kiela D and Weston J** (2018) Personalizing dialogue agents: I have a dog, do you have pets too? In Gurevych I and Miyao Y (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, 2204–2213. doi: [10.18653/v1/P18-1205](https://doi.org/10.18653/v1/P18-1205).
- Zhong P, Zhang C, Wang H, Liu Y, and Miao C** (2020) Towards persona-based empathetic conversational models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 6556–6566. doi: [10.18653/v1/2020.emnlp-main.530](https://doi.org/10.18653/v1/2020.emnlp-main.530).



**Author biographies.** (Eric) **Heng Gu** is a PhD candidate in the Faculty of Industrial Design Engineering at TU Delft. His research focuses on exploring the use of generative AI, specifically language models, to aid design processes in complex design contexts.

**Senthil Chandrasegaran** is Assistant Professor in the Faculty of Industrial Design Engineering at TU Delft. His research focuses on using computational and data visualization approaches to make sense of how designers work,

with a focus on how aspects of designing can be revealed via the language designers use.

**Peter Lloyd** is Professor of Integrated Design Methodology in the Faculty of Industrial Design Engineering at TU Delft, Chair of the Design Research Society, President of IASDR, and former Editor-in-Chief for the journal *Design Studies*. His research focuses on how designers talk in practice, and on storytelling in the design process.