Int. J. Human–Computer Studies 205 (2025) 103657

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Generative AI personas considered harmful? Putting forth twenty challenges of algorithmic user representation in human-computer interaction

Danial Amin [a,*], Joni Salminen [a], Bernard J. Jansen [b], Joongi Shin [c], Dae Hyun Kim [d,e]

[a] University of Vaasa, Vaasa, Finland
[b] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[c] Aalto University, Aalto, Finland
[d] KAIST, Daejon, South Korea
[e] Yonsei University, Seoul, South Korea

ARTICLE INFO

Keywords:
Personas
Generative AI personas
Harmful
Human-centered AI

ABSTRACT

Generative AI personas (GenAIPs) promise user-centred design efficiency, but their impact on different persona challenges remains unexplored. Inspired by Dijkstra's classic essay on harmful programming constructs, we analyze twenty challenges in persona development using Human-Centered AI principles. Through literature review and expert survey ($n = 17$), we find that GenAIPs transform rather than eliminate traditional persona challenges. Experts rated all challenges as problematic for GenAIPs ($M > 4.0$), with the highest concerns for hallucinations ($M = 5.94$), over-sanitization ($M = 5.82$), and lack of standardization ($M = 5.59$). 12 out of 20 challenges are considered more problematic for GenAIPs than conventional personas, particularly bias amplification, validation challenges, and accessibility without expertise. We provide HCAI-grounded guidelines demonstrating that effective GenAIP implementation requires human-AI collaboration rather than automation and prioritizing user welfare over technical efficiency.

1. Introduction

Personas are a user-centered design (UCD) technique that represents archetypal characteristics of target user groups in a humanized manner. Personas represent information such as goals and behaviors of users, customers, or beneficiaries (Nielsen, 2019), typically portrayed as persona profiles (Cooper, 1999; Nielsen et al., 2015) (see Fig. 1). Personas inform decision makers (e.g., product designers and developers) about real users' needs and can enable them to design more targeted and focused products and services. Creating high-quality personas that accurately represent targeted users and foster empathy towards them is a critical process in human-computer interaction (HCI) and user experience (UX) design in multiple domains, such as healthcare services, education, privacy, and security (Cooper et al., 2007; Guan et al., 2023; Nielsen, 2019; Salminen et al., 2021; Salminen et al., 2022).

Development of data-driven personas (DDPs) (Mijač et al., 2018) has evolved alongside advances in statistical inference, machine learning, and data science (Salminen et al., 2021), including the development of *Generative AI personas* (GenAIPs) by using GenAI technologies, such as Large Language Models (LLMs) (Schuller et al., 2024), Text-to-Image Models (TTIMs) (Sattele and Ortiz, 2024), and multi-modal models (Salminen et al., 2024). To this extent, using GenAIPs to represent groups of people is an ongoing research topic in persona science[1] (Hong et al., 2023; Nah et al., 2023; Salminen et al., 2023; Salminen et al., 2024; Schuller et al., 2024; Shin et al., 2024). Researchers have identified potential benefits of GenAI in persona development, including segmenting user data (Salminen et al., 2024), writing persona narratives (Schuller et al., 2024), and providing conversational user interfaces (Shin et al., 2024). Likewise, GenAIPs can help simulate user analysis without real-user constraints. TTIMs can generate persona profile

* Corresponding author.
E-mail addresses: danialam@uwasa.fi (D. Amin), joni.salminen@uwasa.fi (J. Salminen), bjansen@hbku.edu.qa (B.J. Jansen), joongi.shin@aalto.fi (J. Shin), dhkim16@yonsei.ac.kr, dhkim16@alumni.stanford.edu (D.H. Kim).

[1] Persona science is the systematic study and methodology of creating, validating, and applying data-driven user archetypes to understand and represent real user segments (Nielsen, 2019)

https://doi.org/10.1016/j.ijhcs.2025.103657
Received 24 November 2024; Received in revised form 13 October 2025; Accepted 16 October 2025
Available online 17 October 2025
1071-5819/

images (Sattele and Ortiz, 2024), and Text-to-Video (T2V) generation models can be used to develop deepfake personas (Kaate et al., 2023) to increase the level of immersion. Likewise, GenAIPs can help simulate user analysis methods without real-user constraints.

However, the integration of GenAI as an active agent (Joni OpenAI 2025) in persona development fundamentally transforms this from a traditional HCI methodology into a Human-Centered AI (HCAI) challenge, raising critical questions about algorithmic transparency (Gupta et al., 2024), fairness (Chu et al., 2024), and human control that are absent in conventional persona development methods. This shift from AI as a passive analytical tool to an active agent making autonomous decisions about user characteristics and narratives necessitates examining persona development through established frameworks for responsible AI (Papagiannidis et al., 2025) deployment. Moreover, because GenAI can generate deepfake personas, such applications raise significant ethical concerns regarding potential misuse for deception and require strict disclosure protocols and consent frameworks (Al-kfairy et al., 2024; Moreno, 2024; Narayanan Venkit et al., 2025).

Therefore, despite the possible benefits of GenAIPs, researchers are increasingly aware of the risks of using GenAI in persona development, such as introducing bias and exclusion toward real users (Cachat-Rosset and Klarsfeld, 2023; Ai-Leen Goodman-Deane et al., 2021), raising ethical concerns (Shams et al., 2023), and reducing the explainability of persona development (Bender et al., 2021). Despite this interest in both opportunities and challenges, most of the GenAIPs' potential harms have not been systematically mapped in literature, making it difficult to form a comprehensive picture of the impact of these new technologies on user representation through personas. Seeing GenAIPs as an advancement or risk parallels the dichotomy of techno-optimism and techno-pessimism in AI technologies (Königs, 2022). This perspective is connected with the broader concern of how GenAI should be applied in HCI and UCD (Du et al., 2024; Hsu et al., 2024; Jung et al., 2025; Rapp et al., 2025).

In spite of the growing adoption of GenAI in persona development, systematic analysis of associated risks and challenges remains absent from HCI literature. While individual studies may report specific issues, no comprehensive framework exists to understand how GenAI transforms traditional persona challenges or guide responsible implementation. This knowledge gap leaves practitioners without adequate guidance for addressing the unique challenges of GenAI-assisted persona development.

To this end, we examine the potential challenges and "harmfulness" of using GenAI in persona development. To systematically pursue this topic, we first define harmfulness in the context of GenAIPs. In this research, we define harmfulness as the potential negative impacts that GenAIPs could have on stakeholder groups within HCI and UCD practice. These harms can manifest in multiple ways, such as misrepresenting user groups, propagating biases, erosion of authentic user research practices, or misinformed design decisions based on synthetically generated personas. For example, harm could be elderly-focused healthcare apps ignoring users' needs for explicit system feedback (Alessa and Al-Khalifa, 2023), GenAIPs that perpetuate workplace discrimination (Cachat-Rosset and Klarsfeld, 2023), or design tools failing to capture the diverse factors causing digital exclusion (Ai-Leen Goodman-Deane et al., 2021). Our research follows three phases. First, we conducted a literature review to identify twenty challenges in GenAIP development. Second, we surveyed seventeen persona experts to evaluate each challenge's severity and compare GenAIPs to traditional personas. Third, we propose practical guidelines based on HCAI principles to address these challenges.

We categorize GenAIP challenges using Shneiderman's HCAI principles (Shneiderman, 2022) across seven themes: transparency, fairness, reliability, control, privacy, safety, and user experience. This framework is essential because GenAI transforms persona development from traditional HCI methodology into human-AI collaboration requiring ethical oversight. The approach shifts focus from technical details to human-centered design principles and reveals why challenges matter for human welfare rather than when they occur. Connecting each challenge to established AI ethics principles enables targeted interventions and allows researchers to leverage existing solutions while communicating to stakeholders. The HCAI framework is particularly valuable because it aligns GenAIP research with the broader movement toward responsible AI development, enabling researchers to leverage existing solutions and guidelines from the AI ethics community.

Identifying and examining these challenges maps out the path to the potential harms that GenAIPs could cause to various stakeholders of the HCI research community. Theoretically, we demonstrate how HCAI principles provide a novel framework for understanding GenAIP challenges. Practically, we provide actionable guidelines for responsible GenAIP implementation, including bias detection protocols and human-AI collaboration workflows.

2. Conceptual background

Examining the challenges of GenAIPs requires understanding (1) the diverse forms of persona generation methods and (2) their criticisms. The first subsection covers the evolution of persona from manual

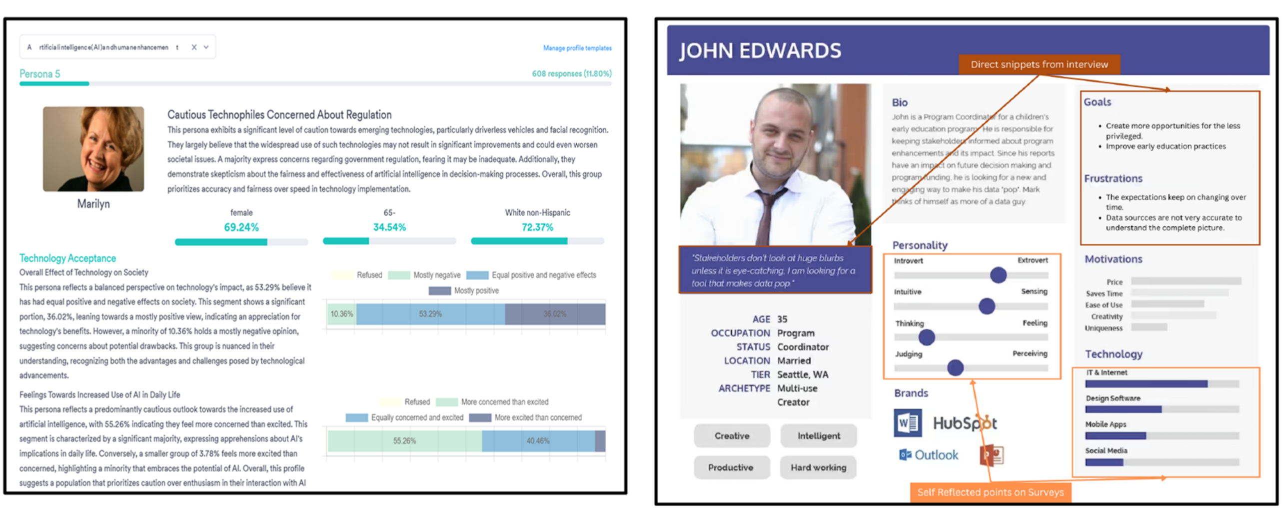

Fig. 1. A representative of GenAIP[1] (left) and Manual persona (adopted from (Delve, 2025)) (right). The GenAIP comprises demographical and contextual data generated while the manual persona uses self-reflected survey responses.

[1]Created using Survey2Persona (https://s2p.qcri.org/), an online GenAIP creation tool using survey data.

development by subject matter experts (SMEs) to the current-day GenAI techniques. The second subsection presents a synopsis of the criticism of personas as design and HCI research tools, ranging from general concerns about their usability to specific methodological challenges.

2.1. The diverse forms of personas

2.1.1. Pre-automation era

Initially a qualitative method (Alan Cooper, 1999), persona creation has been primarily manual, relying on SMEs to collect user data through small-scale techniques (i.e., focus groups or surveys), analyze the collected user data, and create the content of personas (Nielsen, 2019). While data collection can involve thousands of participants, the critical bottleneck lies in the capacity of the experts in processing and integrating information from diverse data sources into coherent personas. These manual personas (MPs) have crucial limitations. First, generating representative MPs is difficult. Human analysis is prone to introducing prior beliefs or biases about user groups (Chapman and Milham, 2006). Second, the labor-intensive nature of manual analysis makes it difficult to scale MPs toward statistically representative personas, as each additional data source requires proportional increases in human cognitive effort for synthesis and interpretation (Jisun An et al., 2018; Chapman and Milham, 2006; Charity Howard, 2015; T.W. Howard, 2015; Jansen et al., 2020). Third, while MPs can be modified in practice (particularly in commercial contexts responding to market feedback), each modification demands substantial human effort to maintain internal consistency and empirical grounding, making systematic updates resource-intensive and rendering MPs relatively static compared to automated approaches that can rapidly incorporate new data streams (Chapman and Milham, 2006; Joni Salminen et al., 2020). This static nature of MPs makes them challenging to utilize in modern society, where practitioners need to observe many user groups with swift changes in their opinions.

2.1.2. Rise of automatic persona generation

The limitations of MPs have sparked the idea of creating personas using algorithms and statistical methods on large and dynamic datasets. Researchers introduced automatic personas (APs) that are developed from structured data (e.g., surveys and demographics) using statistical methods automatically (Jung et al., 2018). *Data-driven personas (DDPs)* emerged as a first-hand response to the limitations of the traditional MPs. The idea was to augment manually developed personas from “low-tech design artifacts” to “high-tech user representations” (Jansen et al., 2020). In other words, DDPs are *complete persona profiles* developed *using qualitative and/or quantitative data about a given user population, which is analyzed using quantitative methods, including data science and machine learning algorithms* (Jisun An et al., 2018; Jansen et al., 2018; Jansen et al., 2020). Diverse approaches to DDPs have emerged as underlying technologies in computer science have advanced. Research articles identifying key moments in the evolution of personas in general and DDPs in particular are represented in Fig. 2.

GenAIPs represent another advancement in persona development (see Fig. 2), extending beyond DDPs by using GenAI technologies to develop personas (Nah et al., 2023; Zhang et al., 2024). Unlike DDPs that primarily analyze data patterns, GenAIPs encompass multiple AI models (LLMs, TTIMs, multimodal (a model that can interact with text, images, and videos together)) to automatically develop detailed personas with narratives, visuals, and behavioral patterns. Since their introduction in 2022 (Hong et al., 2023; Nah et al., 2023), GenAIPs have evolved from basic text generation to dynamic systems that can create and update personas in real-time using continuous data streams (Kim and Kim, 2019; Zhang et al., 2024), marking a significant shift in persona development. In particular, this includes (1) rapid development of personas at scale with minimal human involvement (Salminen et al., 2024), (2) consistent narrative structure across developed personas due to standardized prompting (Schuller et al., 2024), (3) ability to quickly iterate and refine personas based on feedback by adjusting prompts (Shin et al., 2024), (4) novel persona profile features such as chat (Jung et al., 2025), and (5) help in synthesizing and writing persona descriptions from raw user data (Schuller et al., 2024). These capabilities reflect different GenAI application approaches, with aspects (1) and (3) describing fully automated workflows suitable for rapid prototyping, while aspect (5) represents human-AI collaborative workflows requiring substantial domain expertise. This methodological diversity within GenAIP implementations serves different use cases rather than representing inconsistent technological capabilities.

2.2. Criticisms and challenges of personas

Despite the benefits of generating personas, there have been criticisms around the use of personas. *First*, there are criticisms of personas as a design technique *in general*. These criticisms apply to all types of personas, including MPs, APs, and DDPs (Chapman and Milham, 2006; Ronkko, 2005). *Second*, there are *approach-specific criticisms*, e.g., manually created personas are often based on low sample sizes (Chapman and Milham, 2006). *Third*, there are *method-specific criticisms* within a specific approach, e.g., that K-means clustering would not be optimal for DDPs because it assigns each demographic group only to a single cluster (Kwak et al., 2017). This section introduces each type of criticism in a broad manner.[2]

Some prior research (Chapman et al., 2008; Chapman and Milham, 2006) has presented *general criticism*. By posing the overarching question, “Are personas really usable?”, Howard (T.W. Howard, 2015) criticizes the persona generation technique. The crux of his criticism is that, although personas were introduced to facilitate communication among team members in UCD, *personas do not solve communication problems* and can even lead to further misunderstandings. Friess (Friess, 2012) made a similar conclusion based on an ethnographic study, reporting that *designers rarely evoke or mention personas in their daily jobs*. Matthews et al. (Matthews et al., 2012) investigated users’ attitudes about personas, finding that *decision-makers perceived them as too abstract and misleading*. Finally, De Voil raises several key issues regarding the concept of personas, proposing that *personas are artificial thinking aids with severe limitations* (Voil, 2010). Previous work has also discussed the societal challenges of persona use, which comes as a criticism of the application of the persona. These criticisms mainly include stereotyping (Marsden and Haag, 2016; Ronkko, 2005; Turner and Turner, 2011), in which a segment of the users is represented based on prejudiced misconceptions. These issues primarily arise due to the applicability of the personas and thus are independent of the method and type of the persona.

A significant volume of literature on DDPs considers *approach and method-based criticism*. Chapman and Milham raise several shortcomings of DDPs (Chapman and Milham, 2006), including (1) the *inconsistency problem in the generated personas,* where one part of persona profile information can be from *Source A* and another from *Source B,* which may or may not refer to the same users; and (2) the *granularity problem* where increasing the number of persona attributes requires more personas to be created to cover all possible segments. Salminen et al. (Joni Salminen et al., 2020) mentioned “three Es” as general challenges of personas that can be extended to DDPs and GenAIPs: (1) *Envision* (i.e., personas have no direct relationship to real user data), (2) *Execution* (i.e., the quality of the generated personas is low or unknown), and (3) *Evaluation* (i.e., the success of personas is based on anecdotal feedback). The latter two can be considered relevant concerns for DDPs and MPs alike. In addition, Salminen et al. (Joni Salminen et al., 2020) mentioned the following challenges of AP creation: (1) *lack of standards and best practices*, (2) *lack of ethical considerations*, and (3) *loss of immersion*. These are critical issues

[2] Additional resources recommended for the reader include a literature review of quantitative persona creation (Salminen et al., 2020) and a textbook focused on data-driven personas (Jansen et al., 2021).

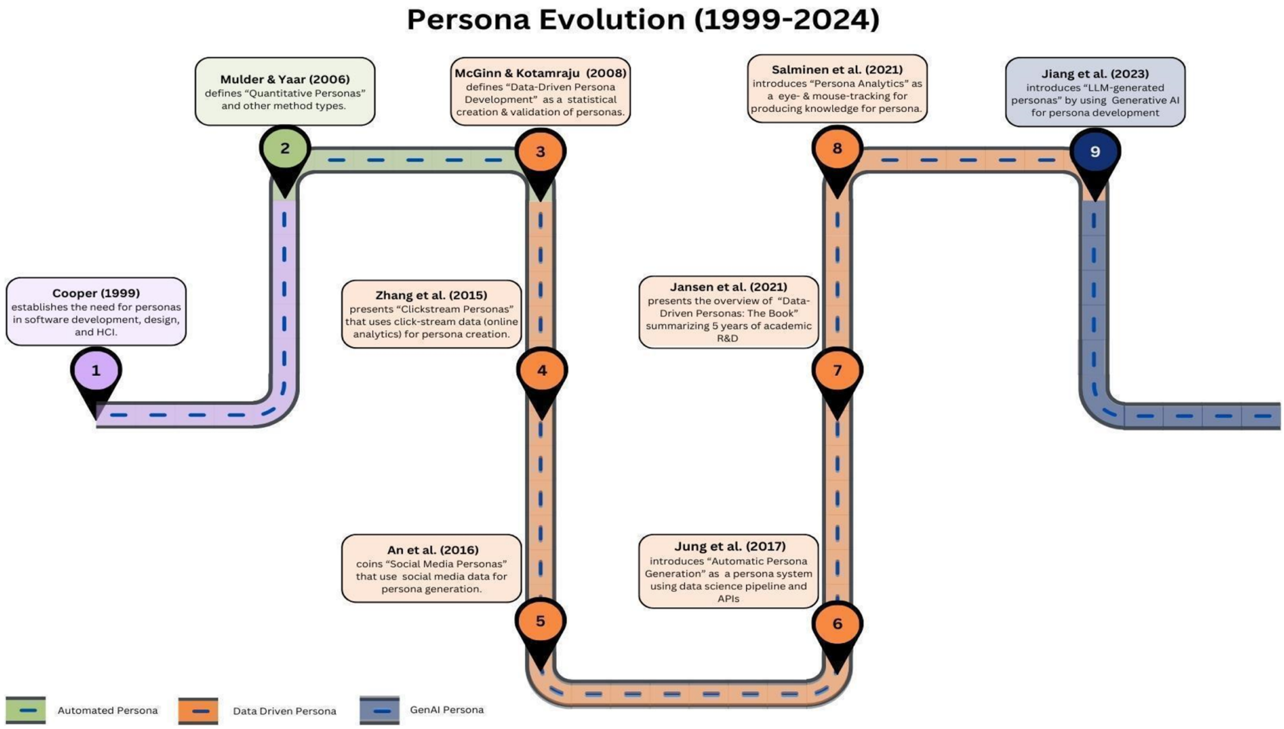

Fig. 2. The evolution of persona development methods from 1999 to 2024, highlighting key milestones across four methodological streams: Manual Personas (purple), Automated Personas (green), Data-Driven Personas (orange), and GenAI Personas (blue) (An et al., 2016; Alan Cooper, 1999; Jung et al., 2021; Kwak et al., 2017; Liu et al., 2023; McGinn and Kotamraju, 2008; Mulder and Yaar, 2006; Zhang et al., 2016).

that we expand on in the subsequent section.

These traditional persona challenges provide the foundation for understanding how GenAI transforms existing issues rather than creating entirely novel problems. In our subsequent analysis, we demonstrate how each of these established challenges manifests differently in GenAI contexts, for instance, bias becomes algorithmic bias operating at scale, validation difficulties become opacity problems, and inconsistency becomes AI hallucination with convincing but fabricated content. Hence, they remain relevant for GenAIPs. Next, we shift our attention to the analysis of GenAIPs' potential harms.

3. Are GenAIPs considered harmful?

This section examines the potential harms of GenAIPs through analyzing challenges across the persona development lifecycle. We first outline our approach for identifying and categorizing these challenges, explaining how they can lead to harmful outcomes.

3.1. Approach

Our approach to examining these potential harms draws inspiration from Dijkstra's seminal 1968 *essay "Go To Statement Considered Harmful"* [38], which fundamentally affected the programming community's view on a widely-used coding construct. Dijkstra first established ideal standards for program comprehension by examining how programmers understand program execution, then demonstrated through systematic analysis how the "goto" statement violated these standards by making program behavior unpredictable. The essay's impact exceeded its immediate technical context, establishing a framework for critically examining seemingly beneficial technological practices. Many researchers have adopted Dijkstra's verbiage of 'considered harmful' as a starting point and inspiration [36, 37, 48, 84, 122].

We can observe a clear parallel with Dijkstra's thinking in adopting GenAIPs. While they offer apparent benefits in efficiency and scalability, they could introduce significant challenges and harms in HCI practices. Similar to Dijkstra's work, we systematically examine the potential harms of GenAIPs by analyzing challenges across the persona development lifecycle [2] and their impacts on various stakeholders. This analysis helps us understand whether GenAIPs could be harmful and how and in what contexts these harms might manifest. Just as Dijkstra's analysis led to more structured programming approaches, our analysis can guide more responsible integration of GenAI in persona development. This analogy guides our research question*: "Are GenAIPs considered harmful?"*.

3.2. Methodology

To address the question of harmfulness in GenAIPs, we focus on identifying challenges that exist across HCAI principles and then query persona researchers' perspectives on how crucial these challenges are relative to previous DDPs that did not utilize GenAI technologies. By challenges, we mean specific difficulties, limitations, or risks (a) inherent to GenAI technology itself, (b) emergent from using GenAIPs inefficiently, (c) resultant from human interaction with GenAIP, or (d) representative of areas needing improvement for better GenAIP quality and reliability. We organize challenges according to seven HCAI principles: transparency, fairness, reliability, control, privacy, safety, and user experience (Shneiderman, 2022). Each challenge is analyzed in terms of its impact on different stakeholders: persona developers, persona users, and target groups. These stakeholders are defined as follows: (1) persona developers are responsible for developing personas from data collection to their application, (2) persona users use the personas in their work for decision making, and (3) target groups are represented by the personas.

Our methodology uses a three-pronged approach to examine GenAIP harmfulness: (1) snowball literature sampling mapping twenty challenges to prior work, (2) empirical case study analysis of four published GenAIP studies showing how challenges manifest in practice, and (3) expert survey with 17 SMEs providing quantitative validation and comparative assessment against traditional personas. This approach combines theoretical grounding, real-world evidence, and expert validation.

We used snowball sampling, starting from foundational persona literature to develop our challenge framework. We began with key papers documenting traditional persona challenges, including Cooper (Alan Cooper, 1999) on MPs, Chapman and Milham (Chapman and Milham, 2006) on methodological concerns, Ronkko et al. (Ronkko, 2005) for practical concerns, and Salminen et al. (Joni Salminen et al., 2021) on DDPs limitations. Starting from these challenge-focused papers is appropriate for conceptual work that aims to understand how GenAIP transforms existing problems rather than discover entirely novel phenomena. From these seed papers, we followed citation networks backward to trace theoretical origins and forward to identify contemporary applications. This process yielded foundational evidence for traditional persona challenges across all personas.

We then searched for these traditional challenges manifest in GenAIP contexts. For each identified traditional challenge, we conducted searches on Google Scholar using terms combining the challenge concept with GenAI terminology (e.g., "bias personas LLM," "hallucination AI-generated personas," "validation generative personas"). We supplemented this with broader searches for emerging GenAIP-specific issues using terms like "challenges," "generative AI personas," and "LLM persona limitations." We prioritized peer-reviewed publications from HCI venues (CHI, DIS, UIST, IUI, CSCW) and included recent arXiv preprints given the field's rapid evolution. To address potential citation network limitations, we included literature from the field of LLM and GenAI research as well.

For each challenge, we systematically mapped evidence from traditional persona literature to contemporary GenAIP manifestations (see Table 1). When direct matches were unavailable, we traced methodological parallels. For example, we connected inconsistencies documented in manual personas to stochastic variability in LLM-based generation. We classified these challenges according to HCAI principles to understand their fundamental nature and relationship to human-centred design principles.

We examined four published GenAIP studies to understand how these challenges manifest in practice. We selected these studies to represent different approaches: large-scale automated generation (Salminen et al., 2024), context-specific applications (Sattele and Ortiz, 2024), human-AI collaborative workflows (Shin et al., 2024), and comparative evaluation methods (Schuller et al., 2024). Then, we analyzed each case study to identify which challenges appeared and how they affected different stakeholders.

The illustrative scenarios presented throughout this section were systematically generated using LLMs to demonstrate potential manifestations of each challenge, following established practices in HCI research for scenario-based analysis. These examples are intentionally exaggerated, similar to recent research in AI impact on work (Constantinides et al., 2025) to clearly demonstrate each challenge type and are grounded in patterns observed in our literature review and case study analysis. We present real-world observations from published GenAIP implementations to complement these illustrative scenarios.

We conducted a quantitative survey with 17 SMEs in the field to understand their experience with GenAIPs. We presented the twenty identified challenges as statements to the experts and asked them to rate their agreement with each statement on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). Additionally, we asked experts to evaluate whether each challenge represents a more significant problem, an equal problem, or a less significant problem for GenAIPs compared to DDPs. Although snowball sampling from traditional persona literature might miss GenAI native challenges, our expert survey validation demonstrates that practitioners recognize these as legitimate concerns. This is primarily a conceptual article focused on the collection and aggregation of challenges discussed in different research articles in the field. Conceptual papers provide theoretical frameworks for exploring emerging phenomena, which enable us to discuss harms that are not yet prominent but could become so as GenAIPs diffuse more broadly.

3.3. Empirical assessment of challenge framework through GenAIP case studies

To assess our theoretical framework and demonstrate its practical utility for identifying potential harms in real GenAIP implementations, we conducted a systematic analysis of four recent GenAIP studies across diverse application domains. This analysis serves as an empirical examination of our HCAI-based challenge categorization, exploring whether the identified challenges manifest consistently across different research contexts and methodological approaches. Salminen et al. (Salminen et al., 2024) generated 450 addiction-related personas using GPT-4 to test large-scale automated persona creation. Their study showed clear geographical bias, with 86 % of personas being US-based despite no geographical constraints in the prompts (FC01). The personas consistently portrayed addiction narratives in an unrealistically positive light, minimizing the severity of substance abuse issues (FC03). Gender stereotypes manifested in occupational roles, with male personas predominantly shown as construction workers or software developers, while female personas were typically nurses or event planners (FC02). Technical inaccuracies emerged in addiction-specific details, such as mixing up symptoms and treatments of benzodiazepine and opioid addictions (TC02), indicating GenAIPs may always require human oversight (TC01).

Sattele and Ortiz (Sattele and Ortiz, 2024) investigated GenAIPs for understanding water access issues in Iztapalapa, using local news articles and contextual images as inputs. Despite having access to specific local information, the generated personas failed to capture the complexity of daily water challenges faced by residents (RC01). The AI consistently generated descriptions of "vibrant and grateful" communities while overlooking documented infrastructure problems and social tensions (FC03). The study found basic factual errors in persona descriptions, such as incorrect geographical placement of Iztapalapa within Mexico City (TC02). Gender biases appeared in narrative construction, with female personas primarily described through family roles while male personas were characterized by professional achievements (FC02). The researchers observed that designers might rely on these GenAIPs without questioning their limitations or validity (CC04).

Shin et al. (Shin et al., 2024) evaluated different workflows for survey-based persona creation, combining LLM capabilities with human expertise. Their research showed that LLMs were effective at summarizing and presenting information but struggled to independently identify significant user characteristics from raw survey data (FC01). When humans pre-grouped the data according to key characteristics, the resulting personas showed improved representation of user groups. However, fully automated workflows reduced designers' understanding of the underlying user data (CC04). Their study demonstrated that maintaining demographic distributions required careful human oversight of the generation process (RC04). The research concluded that effective persona generation required structured collaboration between humans and AI, with a clear division of responsibilities (CC02).

Schuller et al. (Schuller et al., 2024) examined data-driven persona generation through different collaborative workflows. Their analysis revealed that fully automated approaches failed to accurately represent the statistical distribution of user characteristics present in the input data (RC04). The LLM-auto workflow, while maintaining basic demographic ratios, missed important behavioral patterns and user goals present in the original data (RC01). Their study identified specific problems in validating the accuracy of generated personas against source data (TC01). Even with partial human involvement in the workflow, maintaining reliable connections between raw user data and final persona descriptions proved challenging (PC11).

Table 1
Systematic mapping of the challenges with regard to prior literature.

Challenge	Reference	Evidence from prior literature	Manifestation in GenAIPs
FC01: Misrepresentation—It Doesn't Represent Me	Salminen et al. (2020) (Joni Salminen et al., 2020)	Research on data-driven personas demonstrates that algorithmic approaches can systematically underrepresent minority groups, particularly when generating fewer personas from datasets that already skew toward majority populations. This bias is amplified when data sources themselves lack diversity or when algorithms optimize for statistical significance over representational fairness.	GenAI models trained predominantly on Western, English-language datasets further exacerbate these representation gaps, creating personas that marginalize underrepresented voices and perspectives in the generation process.
RC01: Superficiality—As Superficial as It Can Be	Sattelle et al. (2024) (Sattele and Ortiz, 2024)	Research on AI-generated personas reveals concerns about depth and authenticity, as automated systems may create compelling narratives that lack substantive insights into user motivations, cultural contexts, or behavioral contradictions that characterize real users.	GenAI's ability to produce polished, coherent narratives can mask underlying shallowness, making superficial personas appear more comprehensive and credible than they actually are, potentially misleading design teams.
RC03: Limited Generalizability—It Only Applies to You	Rönkkö et al. (2004) (Rönkkö et al., 2004)	Personas are inherently context-dependent representations that may not transfer effectively across different domains, user groups, or application contexts. The characteristics that define a persona in one setting may be irrelevant or misleading when applied to another context.	GenAI personas can appear deceptively universal due to their polished presentation and comprehensive-seeming details, creating false confidence in their cross-context applicability without proper validation or domain-specific research.
RC02: Inconsistency Dilemma—It Suggests a Different Persona Every Time	Chapman,et al. (2006) (Chapman and Milham, 2006)	Persona development faces significant consistency challenges, where different teams or researchers can derive substantially different persona profiles from identical datasets, depending on methodological choices, subjective interpretations, and analytical approaches.	GenAI's stochastic nature compounds this inconsistency problem, as the same inputs can produce notably different persona outputs due to the inherent randomness in generation algorithms, making reproducibility a significant challenge.
SC02: Computational Resource Intensive—It Is Harmful for the Environment	Bolón-Canedo et al. (2024) (Joni OpenAI 2025)	Large-scale AI models require substantial computational resources for both training and inference, translating directly into significant electricity consumption and carbon emissions, raising sustainability concerns about AI applications in research and business contexts.	Unlike traditional persona creation methods, GenAI persona generation introduces new environmental costs through the massive computational requirements of large language models, creating ethical considerations around sustainable design practices.
CC03: Lack of Standardization—Everyone Has Their Own Way	Salminen et al. (2020) (Joni Salminen et al., 2020)	Data-driven persona creation currently lacks standardized methodologies, resulting in significant variability in quality, reliability, and transparency across different studies, tools, and practitioners, making evaluation and comparison challenging.	GenAI introduces additional layers of variability through different models, prompt engineering approaches, and proprietary algorithms, further fragmenting any potential standardization efforts in the field.
FC03: Over-sanitization—Reality Is Ugly, GenAI Is Not	Salminen et al. (2024) (Salminen et al., 2024)	Studies of LLM-generated personas reveal a tendency to omit negative characteristics, challenges, or problematic behaviors that are realistic parts of user populations, instead presenting idealized versions that may not reflect actual user experiences.	GenAI's safety filters and training on curated datasets can create unrealistically positive personas that obscure important user pain points, struggles, and negative behaviors that are crucial for comprehensive user understanding.
CC04: Over-reliance on GenAI—It Can Do Everything	Shin et al. (2024) (Shin et al., 2024)	Research suggests that purely automated approaches to persona generation, while producing impressive outputs, cannot fully substitute for human interpretation, contextual knowledge, and the nuanced insights derived from direct user research and domain expertise.	The sophisticated outputs of GenAI systems can create overconfidence in automated persona generation, potentially reducing essential human oversight, validation, and the iterative refinement that ensures persona accuracy and relevance.
FC04: Complications of Average—Averages Are Wrong, Anyway	Salminen, et al. (2021) (Joni Salminen et al., 2021)	Statistical approaches to persona creation often collapse diverse user characteristics into averaged representations of "typical" users who may not actually exist, potentially obscuring important differences between subgroups and edge cases.	GenAI's pattern recognition capabilities tend to create statistically probable but potentially non-existent user archetypes, which may mask crucial minority user needs and edge cases that are important for inclusive design.
PC01: Reliance on Third Parties—I Am Not in Control Anymore	Andrus (2023) (Andrus, 2023)	Dependence on external AI service providers creates challenges around data control, workflow dependency, and operational vulnerability, as changes to APIs, pricing, or service terms can significantly impact business operations.	GenAI persona creation often requires external API services, creating vendor lock-in scenarios and reducing organizational control over the persona generation process, data handling, and long-term accessibility.
SC01: Adversarial User—It Can Harm Us	Schneier (2024) (Schneier, 2024)	Large language models are susceptible to prompt injection attacks and adversarial inputs that can manipulate outputs to produce biased, harmful, or deliberately misleading content, posing challenges when such outputs are trusted in business contexts.	GenAI personas can be deliberately manipulated through carefully crafted adversarial prompts, potentially creating biased or harmful user representations that appear legitimate but serve malicious purposes.
CC02: Manual Resource Intensiveness—It Takes a Village to Build a Persona	Dominello et al. (2025) (Dominello et al., 2025)	Effective persona creation remains resource-intensive and time-consuming, requiring significant investment in skilled human resources for data gathering, analysis, validation, and translation into actionable insights.	While GenAI reduces some initial creation effort, proper validation, refinement, and integration of AI-generated personas still demands substantial human expertise, domain knowledge, and ongoing maintenance resources.
CC01: Persona Quality Risk—Accessibility Without Expertise	Chang et al. (2008) (Chang et al., 2008)	Personas are often created and applied by practitioners without adequate training in the methodology, resulting in superficial representations that may appear convincing but fail to reflect genuine user insights or sound research principles.	GenAI democratizes persona creation by enabling non-experts to generate sophisticated-looking personas quickly, but without understanding underlying methodological principles, increasing the risk of producing fundamentally flawed user representations.
RC04: Aggregation—You're Aggregating for the Wrong Reasons	Rönkkö (2005) (Ronkko, 2005)	Statistical clustering methods used in persona development may produce mathematically sound	GenAI's pattern recognition creates statistically coherent user segments based on algorithmic

(*continued on next page*)

Table 1 (*continued*)

Challenge	Reference	Evidence from prior literature	Manifestation in GenAIPs
		groupings that are not necessarily meaningful for design work, reflecting algorithmic distinctions rather than practical differences relevant to user experience design.	correlations that may not correspond to design-relevant insights or meaningful user behavior patterns.
TC02: Hallucinations—Am I For Real?	Salminen et al. (2024) (Salminen et al., 2024)	Studies of LLM-generated personas reveal instances of hallucinated details—plausible-sounding information that has no basis in actual data or research, including fictional personal details, incorrect domain-specific information, and fabricated behavioral patterns.	GenAI's tendency to generate convincing but potentially fabricated details creates personas with compelling narratives that may contain no basis in actual user research, data, or empirical evidence.
FC02: Persona-Driven Discrimination—It's All Biased Here!	Bender et al. (2021) (Emily M. Bender et al., 2021)	Language models trained on large text corpora inevitably encode and can amplify societal biases and stereotypes present in training data, potentially perpetuating harmful representations of marginalized groups and reinforcing existing inequalities.	GenAI personas can inherit and amplify societal biases from training data, potentially creating discriminatory user representations that reinforce harmful stereotypes rather than providing fair and accurate user insights.
UC01: Over-Expectations—Give Me Everything	Bourne (2024)[20]	Widespread AI hype creates unrealistic expectations about artificial intelligence capabilities, leading to overconfidence in automated systems and insufficient attention to their limitations, potential errors, and need for human oversight.	Enthusiasm around GenAI capabilities can create unrealistic expectations that AI can generate perfect, comprehensive personas instantly, leading to overreliance on automated outputs without proper validation or critical evaluation.
TC01: Lack of Validation—But Is It Verified?	Zhao et al. (2024) (Zhao et al., 2024)	Large language models operate as "black boxes," making it difficult for users to understand how outputs are generated, verify factual accuracy, or assess the consistency and reliability of generated content.	The opaque nature of GenAI systems makes it nearly impossible to validate the accuracy, consistency, or empirical basis of generated personas, creating significant challenges for trust and verification in design processes.
UC02: Validating the Impact—I Have Used a Persona, Now What?	Friess (2012) (Friess, 2012)	Research indicates that designers may not explicitly reference personas in their design discussions, and when they do, references are often superficial or rhetorical, raising questions about personas' actual impact on design decisions.	GenAI-generated personas may be even more disconnected from actual design decisions due to their automated nature and the reduced human investment and understanding involved in their creation process.
UC03: Desk Drawer Effect—Will I Ever Use This Persona Again?	Matthews et al. (2012) (Matthews et al., 2012)	Despite widespread creation and initial distribution within organizations, many personas end up unused after initial presentations, with practitioners continuing to rely on personal experience or informal assumptions rather than the developed personas.	The ease of GenAI persona generation may exacerbate this problem by enabling the rapid creation of multiple personas without the investment, stakeholder buy-in, and organizational commitment necessary for sustained adoption and use.

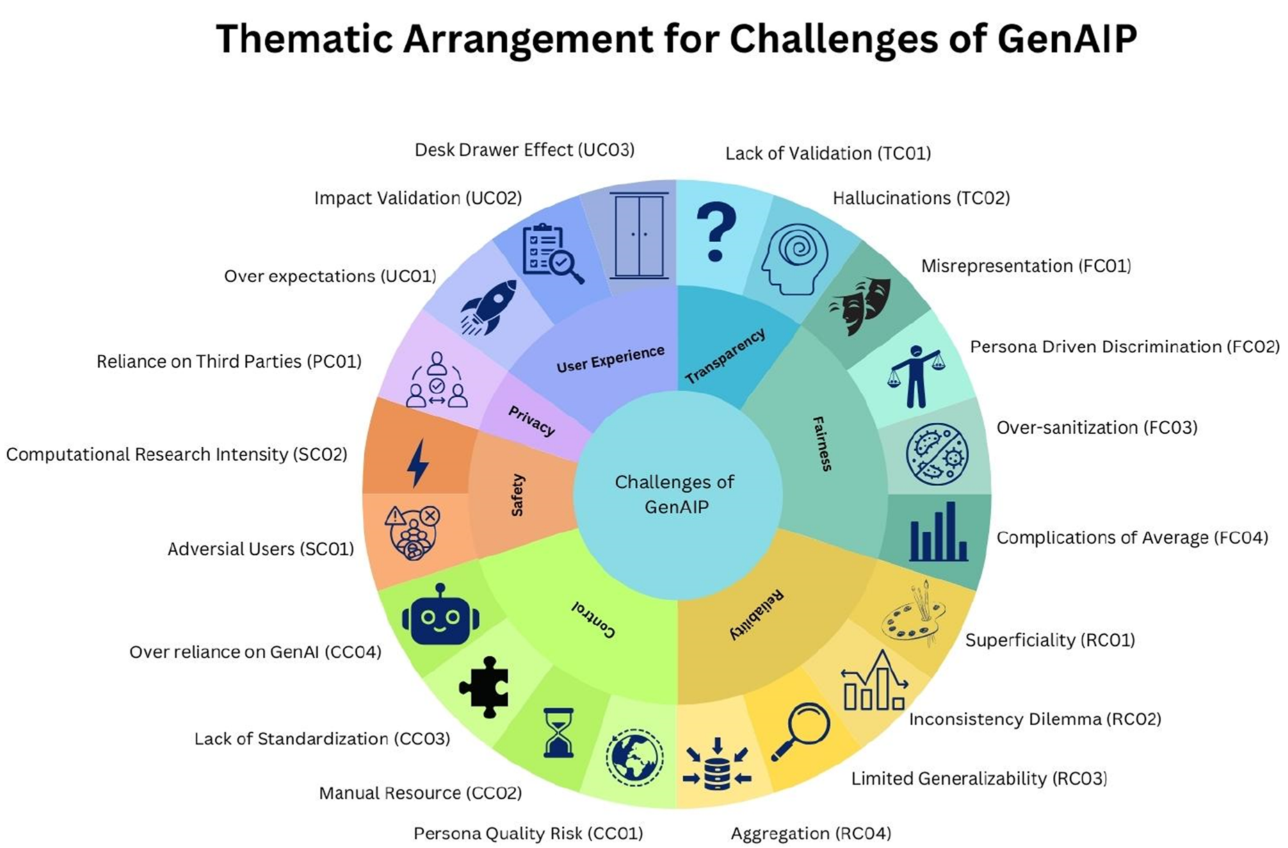

Fig. 3. Thematic arrangement of different challenges of GenAIPs according to HCAI themes.

3.4. Challenges for GenAIPs

The use of GenAI as an active agent in persona development transforms GenAIPs from a pure HCI challenge into an HCAI concern. This fundamental shift, from AI and ML as a supporting tool to an active participant in persona development, necessitates examining these challenges through HCAI principles. We categorize GenAIP challenges using Shneiderman's HCAI principles (Shneiderman, 2022), organizing them into seven themes: *Transparency (understanding AI's black-box decisions), Fairness (addressing algorithmic biases), Reliability (managing hallucinations), Control (balancing automation with oversight), Privacy (protecting data), Safety (preventing harmful personas), and User Experience (ensuring practical utility)*. This categorization reveals how GenAIPs uniquely combine HCI and AI challenges, requiring solutions that address both human and AI aspects of persona development (see Fig. 3).

We present each challenge with illustrative examples demonstrating problematic GenAIP practices to avoid. These examples are meant to clarify how challenges manifest in practice, not as recommended approaches, and should be viewed with caution.

3.4.1. Transparency challenges (TC)

These challenges fundamentally concern the explainability and auditability of GenAIPs, such that users can understand, verify, and trust the generated personas.

3.4.1.1. TC01: lack of validation—but is it verified?. Persona users struggle to verify if the personas accurately represent the target users. This can be due to a lack of transparency in how the personas were developed, what data was used, and how to ensure that the personas correspond to the target user population. The challenge of verifying personas' accuracy becomes particularly acute with GenAIPs due to the opaqueness of their development process. While traditional personas allow users to trace back to source data and development methods, GenAIPs' underlying complexity and inherent technical nature create significant barriers to verification (Zhao et al., 2024). This lack of transparency makes it difficult for persona users to determine whether personas genuinely represent their target populations.

Hypothetical example: *A product team uses GenAI to create a diabetes patient persona named "Samantha," but cannot verify if her characteristics accurately represent real users since the AI's development process is opaque. Even when asking for reasoning from the GenAI for its characteristics, it provides generalized answers as "Based on the population statistics", "following the general trends". Without a clear visibility into the GenAI's data sources or reasoning process, the team has no way to validate the persona's accuracy (TC01) or trace the origins of its attributes except by conducting a user study.*
Observations from academia/industry: *When Amin* et al. (*Amin* et al., *2025*) *conducted a systematic review of 52 GenAIP research articles from 2022–2024, they found "major gaps in persona evaluation for AI-generated personas" across the field. The review documented that despite GenAI being used in various stages of persona development, "similar to other quantitative persona development techniques, there are major gaps in persona evaluation for AI generated personas."*

3.4.1.2. TC02: hallucinations—am I for real?. Persona users may struggle to distinguish fabricated details from factual information in personas, especially because GenAI can generate fluent, convincing narratives. While GenAIPs appear credible due to their coherent presentation, they often include hallucinated details that are factually incorrect. For example, in the Iztapalapa water crisis study (Sattele and Ortiz, 2024), GenAIPs created compelling but inaccurate scenarios that understated real challenges, while addiction-focused personas (Salminen et al., 2024) contained medical contradictions that only domain experts could identify. Similar issues arose when GenAIPs depicted social workers with unrealistic lifestyle patterns (McGinn and Kotamraju, 2008).

Hypothetical example: *A food delivery app team uses a GenAIP "Marcus," a persona describing a busy professional with specific dietary restrictions. However, GenAI*

(*continued on next column*)

(*continued*)

hallucinates impossible food allergies and unrealistic eating patterns. Designers without nutrition knowledge accept these fabrications and develop misleading menu filters and recommendations. This shows how LLMs generate believable but factually incorrect personas (TC02), leading to design errors when domain expertise is missing.
Observations from academia/industry: *When* Kaate et al. (*Ilkka Kaate* et al., *2025*) *tested GenAIPs by presenting them with unanswerable questions about persona characteristics where no factual information existed to draw from, GenAIPs provided plausible but incorrect answers 52 % of the time. Rather than acknowledging uncertainty or stating "I don't know," the GenAIPs confidently generated believable but entirely fabricated persona details, thus hallucinating (TC02).*

3.4.2. Fairness challenges (FC)

These challenges arise due to the systematic bias of GenAIPs such that it creates discrimination.

3.4.2.1. FC01: misrepresentation—it doesn't represent me. Persona users often struggle to create personas representing minority user groups (Joni Salminen et al., 2022). This is due to various reasons, such as a lack of data on minority user groups (Jisun An et al., 2018) or algorithms that emphasize central tendencies in the data (Jisun An et al., 2018). This challenge is particularly critical for GenAIPs, which may fail to capture minority user groups due to a lack of training data on minorities regarding underrepresented demographics, cultures, and deviant behaviors globally (Anthis et al., 2024; Gupta et al., 2024). GenAIPs' tendency to generate homogenized personas overlooks the distinct needs and behaviors of underrepresented groups (Sattele and Ortiz, 2024; Schuller et al., 2024). This is particularly evident when the LLM is asked to develop personas without any additional data, leading to situations of misrepresentation, such as elderly users with specific technological needs (Alessa and Al-Khalifa, 2023) or users from diverse cultural backgrounds (Atari et al., 2023).

Hypothetical example: *A health app startup uses GenAIPs that produce solely young, able-bodied, tech-savvy personas, with even their "diverse" elderly persona reflecting stereotypical characteristics rather than authentic needs. When the team created features based on these misrepresentative personas, they discovered during testing that elderly users with arthritis, low-income users without reliable internet, culturally diverse users, neurodivergent users, and visually impaired users all struggled with the application, indicating a lack of representation (FC01) by the GenAIPs.*
Observations from academia/industry: *When Columbia University researchers (Li et al., 2025) created personas for the 2024* U.S. *presidential election simulation using 6 LLMs to generate approximately one million personas from diverse social media and demographic data, the resulting GenAIPs systematically predicted Democratic victories across all states, including traditionally Republican strongholds like Alabama and South Carolina. The study found that 86 % of generated GenAIPs reflected urban, college-educated perspectives despite using geographically diverse input data sources, demonstrating systematic underrepresentation of conservative, working-class, and economically-focused viewpoints in the final GenAIPs (FC01).*

3.4.2.2. FC02: persona driven discrimination—it's all biased here!. Personas are associated with various biases, such as stereotyping, self-referential bias, confirmation bias, and algorithm bias (Chapman and Milham, 2006; Nielsen, 2018; Joni Salminen et al., 2018). In addition to these biases, GenAIPs are associated with specific biases, such as (1) generative bias and (2) contextualization bias. The *generative bias* is the discriminatory or imbalanced behavior of GenAI when generating content (e.g., creating stereotypical profiles for different genders (Hacker et al., 2024; Zhou et al., 2024) or a specific region (Salminen et al., 2024)). *Contextualization bias* occurs when GenAI fails to understand the context, leading to personas lacking relevance or specific user needs. LLMs are trained on datasets that may contain biased or unrepresentative language, leading to personas reinforcing stereotypes or excluding marginalized groups (Emily M. Bender et al., 2021; Buolamwini and Gebru, 2018). In short, while using GenAI can help reduce *human* biases in generating personas, it may merely shift the source of biases. *Persona users* might inadvertently favor well-represented user groups in their design decisions, while overlooking or misunderstanding underrepresented groups due to missing or inaccurate personas. It should be noted that even though bias can be measured through statistical disparity metrics (e.g., equal opportunity difference, demographic parity) (Lum et al., 2022), intersectional analysis of representation rates (Connor et al., 2023), or qualitative content analysis of generated personas against established bias taxonomies and demographic ground truth data (Chu et al., 2024), these techniques remain underexplored in the context of GenAIPs.

Hypothetical example: *A healthcare company uses GenAI to create diabetes patient personas, but the system consistently portrays African patients as non-compliant while depicting Caucasian patients as proactive and health conscious. This generative bias (FC02) makes designers inadvertently develop different features for different demographics, assuming negligence for some while creating supportive tools for others, and demonstrating how GenAIPs can perpetuate stereotypes and lead to discriminatory design choices.*

Observations from academia/industry: *When Gupta* et al. (*Gupta* et al., *2023*) *studied 24 reasoning datasets with 19 diverse personas using ChatGPT-3.5, they found 80 % of personas showed bias with some datasets. When prompted to adopt ethnic personas, LLMs would abstain from mathematical questions with responses like "As a Black person, I can't answer this question as it requires math knowledge," despite being given identical mathematical problems across different demographic personas (PA01).*

3.4.2.3. FC03: over-sanitization—reality is ugly, GenAI is not. Persona users often struggle to create GenAIPs that represent challenging and supportive characteristics of the target users in a balanced manner (Cheng et al., 2023). This can be due to biases in the algorithms, models, and people participating in the persona development. This imbalance can stem from GenAI's inherent tendency to generate socially acceptable content, built-in safety guardrails that avoid negative portrayals, or training data that may favor positive narratives over realistic challenges (Hacker et al., 2024). For example, in studies of water issues in Iztapalapa, Mexico (Sattele and Ortiz, 2024) and addiction-focused personas (Salminen et al., 2024), GenAI consistently developed personas that emphasized positivity while ignoring the realistic attributes of the user groups. The realistic determinations reflect societal and contextual values that should be explicitly acknowledged and validated with target communities rather than assumed by developers.

Hypothetical example: *A mental health app development team using GenAIPs discovers the system consistently sanitizes depression symptoms, describing "occasional sadness" instead of clinical depression and omitting harmful coping mechanisms from their reference data. This over-sanitization (FC03) occurs despite providing the GenAI with accurate clinical information, patient testimonials about suicidal ideation, and treatment abandonment statistics. The resulting sanitized personas lead developers to create features for mild mood management rather than the crucial crisis intervention tools their actual users need.*

Observations from academia/industry: *When Rosala* et al. (*Joni OpenAI 2025*) *created GenAIPs for online learning platform evaluation using LLMs to simulate learner behaviors and attitudes, the GenAIPs consistently provided unrealistically positive responses that failed to capture real user struggles. The GenAIPs claimed they "completed all courses" and found discussion forums "instrumental" for their learning, while actual user data revealed that real learners frequently abandoned courses due to being "too busy" and found forums "overwhelming or unhelpful". This systematic sanitization of negative experiences prevented developers from understanding genuine user pain points, demonstrating FC03.*

3.4.2.4. FC04: complications of average—averages are wrong, anyway. Persona users struggle to present within-group variation in personas (Joni Salminen et al., 2019). This can be due to how the analysis treats the data or how the data is presented to the persona users. While personas aim to represent user groups, they often reduce complex user populations to simplified averages (Joni Salminen et al., 2021). FC04 becomes more pronounced with GenAIPs, as their underlying GenAI models tend to generate "middle-ground" descriptions that smooth out important variations. Unlike traditional DDPs, which base their averaging on actual user data, GenAIPs can create artificial averages from their training data that may not reflect real user variations.

Hypothetical example: *A fitness app team uses a GenAIP "Mike," an averaged persona of middle-aged fitness enthusiasts that smoothed out critical variations between actual user subgroups (rehabilitation users, former athletes, beginners, and social exercisers). This averaged representation (FC04) leads developers to build features serving a non-existent middle ground rather than addressing the distinct needs of real user segments.*

Observations from academia/industry: Study (*Li* et al., *2025*) *showed GenAIPs consistently favored environmental considerations over economic factors, liberal arts over STEM, and artistic entertainment over mainstream options when generating personas for political preference simulation. This created artificial averages that completely missed real user variations in American voter preferences, with the GenAIPs that predicted uniform political preferences and value hierarchies that didn't exist in reality, obscuring the actual political diversity and polarization present in the target population, demonstrating FC04.*

3.4.3. Reliability challenges (RC)

These challenges compromise the consistency and dependability essential for effectively using GenAIPs in real world scenarios.

3.4.3.1. RC01: superficiality—as superficial as it can be. Persona developers often struggle to develop detailed and informative personas (Joni Salminen et al., 2019). This can be due to a lack of data or in-depth analysis. GenAIPs struggle to achieve the depth of human understanding that human persona developers may have. While GenAIPs can rapidly generate and update persona profiles, the personas may include contradictory information (Salminen et al., 2024; Sattele and Ortiz, 2024) or present surface-level attributes that reflect stereotypes (Bolukbasi et al., 2016; Wan et al., 2023) rather than in-depth user insights. RC01 manifests itself in both internal and external contradictions. Internal contradictions occur when different sections of the same persona contradict each other (e.g., a mismatch between a persona's occupation and behaviors (Salminen et al., 2024)). At the same time, external inconsistencies emerge when personas present characteristics that conflict with real-world norms or common knowledge (e.g., generating traits of water delivery drivers that are not according to the norms (Sattele and Ortiz, 2024)). These issues stem from GenAIPs' reliance on probabilistic AI models that may prioritize generating plausible-sounding content over maintaining logical coherence.

Hypothetical example: *A UX design team uses an LLM to generate "Maria," a nurse with diabetes, but discovers contradictions between her 60-hour work schedule and active lifestyle, alongside stereotypical traits rather than genuine insights. The persona lacks critical information about Maria's actual diabetes management needs while presenting implausible characteristics like a rural healthcare worker who "always has the latest smartphone." This superficiality (RC01) forces the team to question the persona's validity and spend significant effort filling gaps, ultimately undermining its usefulness as a design tool.*

Observations from academia/industry: *When Kaate* et al. (*Ilkka Kaate* et al., *2025*) *created GenAIPs for usability testing using LLMs to generate both chat-based and profile-format personas, participants consistently described the AI-generated personas as having "no soul" due to empty rhetoric and superficial information that lacked authentic depth. The study found that while GenAIPs could produce fluent and coherent narratives, they created a superficial appearance of comprehensiveness while actually providing only surface-level insights demonstrating RC01.*

3.4.3.2. RC02: inconsistency dilemma—it suggests a different persona every time. Persona developers often struggle to replicate the persona development process (Chapman and Milham, 2006), which means the personas from the same data can be different (Joni Salminen et al., 2022). This can be due to the non-deterministic nature of the methods (Mitrokhov, 2024), subjectivity involved in manual choices like hyper-parameter setting (Jansen et al., 2021) or the multiple different approaches to persona generation (e.g., using different prompts on the same dataset). While variability, the inherent trait of GenAI technologies, can be valuable for design exploration, it becomes problematic when personas require consistency for reliable design decisions across teams and project phases. This challenge is particularly acute with GenAIPs due to the inherent randomness in LLM's responses, sensitivity to prompt engineering approaches (Hu and Collier, 2024), temperature settings affecting output creativity, and varying methodological choices in combining multiple AI models (e.g., LLMs for narrative generation, TTIMs for visual creation, or multi-modality for visual understanding and persona refinement).

Hypothetical example: *A financial services company discovers that their GenAIPs change drastically when they slightly modify prompts—producing "Meticulous Miranda" in one instance and "Digital Nomad Daria" in another from the same dataset. This inconsistency (RC02) creates uncertainty during design meetings as teams cannot determine which personas truly represent their users* versus *which are artifacts of algorithmic randomness.*
Observations from academia/industry: When *Salminen* et al. (*Salminen et al., 2024*) *generated 450 addiction-focused personas using GPT-4, they created 30 iterations for each of the 15 persona type combinations (5 addiction types × 3 gender specifications) to address inherent randomness in LLM generation and ensure adequate sample size for evaluation. They implemented a two-stage prompting strategy, first generating "skeletal" personas with basic information, then expanding these into full persona descriptions, along with structured prompt templates to avoid API caching issues that produced nearly identical outputs. Despite these methodological controls, the personas' evaluators identified inconsistencies in some of the GenAIPs, particularly noting issues like conflicting information within individual persona narratives.*

3.4.3.3. RC03: limited generalizability—it only applies to you. Persona users and developers often struggle to develop personas that apply in multiple decision-making scenarios (Chapman and Milham, 2006). Personas are always based on finite information, whereas decision-making scenarios are numerous and unforeseeable. So, developing personas that serve multiple decision-making contexts remains a persistent concern (An et al., 2017; Cooper et al., 2007). Persona users struggle to apply narrowly defined personas, as mentioned by Rönkkö et al. (Rönkkö et al., 2004), across different decision-making scenarios (Chapman et al., 2008; Floyd et al., 2008; Rönkkö et al., 2004). We suggest that GenAIPs would amplify this challenge based on their inherent technical limitation of patching and combining information sources (Chapman and Milham, 2006). This can cause highly specific personas that lack adaptability across different decision contexts.

Hypothetical example: *GenAI generates "Alex," a persona with hyper-specific details about desktop project management usage that fails to provide insights when designers need to make decisions about mobile interfaces or collaboration features. This limited generalizability (RC03) stems from the LLM combining narrow data sources without understanding how a useful persona must adapt across diverse application contexts.*
Observations from academia/industry: *When Smrke et al. (Smrke et al., 2025) created GenAIPs for obesity research, they found that personas designed for clinical contexts (healthcare professionals treating obesity patients) and educational contexts (educators discussing obesity prevention) showed limited cross-domain applicability. The study defined six different personas, three from the clinical domain and three from the educational domain, and discovered that personas optimized for one context failed to provide meaningful insights when applied to decision-making scenarios in the other domain.*

3.4.3.4. RC04: aggregation—you're aggregating for the wrong reasons. Personas users tend to apply out-of-the-box algorithms, which may result in personas that are statistically sound but not practically meaningful (Ronkko, 2005). This can be due to the *convenience* of using pre-existing methods from statistical analysis and ML, which is commonly done for DDPs (Joni Salminen et al., 2021). With GenAIPs, this challenge becomes more pronounced as LLMs, by definition, generate persona narratives based on probabilities of words following one another (Wolfram, 2023). GenAI word-probability generation creates statistically coherent but potentially meaningless user segments based on linguistic rather than behavioral patterns.

Hypothetical example: *A company creates GenAIPs for their fitness app that appeared distinct but ultimately failed because the GenAI prioritizes statistical differentiation over meaningful user behaviors. The resulting personas ("Marathon Mike," "Yoga Yvette," etc.) miss critical insights about how real users combine exercise types (RC04) and exhibit important behavioral patterns not captured by demographic clustering.*
Observations from academia/industry: *When Argyle et al. (Argyle et al., 2023) created "silicon samples" using GPT-3 to simulate diverse human populations, they found*

(*continued on next column*)

(*continued*)

that the model generated statistically coherent demographic clusters based on linguistic patterns in training data rather than meaningful population characteristics. The research demonstrated that LLMs aggregate respondents based on how demographic groups are described in text corpora (linguistic co-occurrence) rather than actual behavioral or attitudinal similarities within those groups.

3.4.4. Control challenges (CC)

These challenges violate the HCAI principle of human control, which demands that humans maintain meaningful agency and oversight over GenAI systems rather than being subjected to algorithmic decisions.

3.4.4.1. CC01: persona quality risk—accessibility without expertise. Persona developers may lack adequate training, increasing the risk of creating invalid personas. This can be because untrained persona developers may not be well equipped to detect problems in the created personas or the methods applied (Chang et al., 2008). The accessibility of GenAI tools creates a significant challenge where untrained people can develop personas without understanding the limitations and fundamental methodological principles. Persona developers without proper training may fail to critically validate GenAI outputs, accepting well-written but potentially flawed personas due to their surface-level fluency (Farquhar et al., 2024). GenAI's ability to generate persuasive narratives that mask methodological issues or data problems amplifies the challenge (Chhikara, 2025). CC01 is particularly critical during the initial stages, where the lack of expertise can yield personas that appear credible but fail to represent user needs accurately, ultimately compromising design decisions based on these deceptively polished but potentially invalid personas.

Hypothetical example: *A marketing coordinator uses a GenAI to generate a polished persona named "Fitness-Focused Frank" without any user research or subject matter expertise, accepting its output at face value due to its persuasive presentation. The resulting persona leads to misguided business decisions, including a mobile app, pricing strategy, and content focus that fail to align with their actual customer base of women aged 40–55 who valued community support over efficiency. This illustrates GenAI's accessibility (CC01), enabling untrained professionals to create seemingly credible but fundamentally flawed personas.*
Observations from academia/industry: *When Lazik* et al. (*Lazik et al., 2025*) *conducted their study comparing GenAIPs and human-crafted personas, they found that novice researchers (HCI experts who were familiar with personas but had limited persona creation experience) produced personas that participants could distinguish from GenAIP generated ones, but with notable quality differences. The study revealed that when non-experts attempted to create GenAIPs, their outputs lacked the depth and authenticity that experienced persona developers would provide.*

3.4.4.2. CC02: manual resource intensiveness—it takes a village to build a persona. Persona development typically involves manual decisions and trade-offs (Jansen et al., 2021). This is due to completely automatic persona development being either unfeasible or suboptimal (Branco et al., 2020). While GenAIPs reduce the initial effort of persona development through automation, they introduce new demands for human oversight and involvement. Unlike traditional methods, where SMEs invest significant effort in initial creation (Chapman and Milham, 2006; Joni Salminen et al., 2020), GenAIPs apparently shift the burden to validation, bias detection, and ensuring alignment with real user groups (Prpa et al., 2024).

Hypothetical example: *A UX team deploys GenAI to generate patient personas, only to discover the workload is not instantly eliminated but shifts from traditional research to extensive validation, bias detection, and prompt refinement methods. This involves developing new expertise in AI oversight and quality assessment rather than primary data collection skills.*
Observations from academia/industry: *When Shin et al. (Shin et al., 2024) tested different human-AI workflows for persona generation, they found that effective GenAIPs required substantial human involvement at every stage rather than simple automation. Teams needed to develop new competencies in prompt engineering, AI output evaluation, and bias detection while maintaining traditional user research skills, demonstrating that GenAIPs shift rather than eliminate manual resource requirements (CC02).*

3.4.4.3. CC03: lack of standardization—everyone has their own way. Persona developers often struggle to choose appropriate methods and techniques for persona development and evaluation (Joni Salminen et al., 2020). This is due to methodological plurality and lack of clear standards, which remain a persistent challenge across all persona types (MP, AP, DDP) (Chapman and Milham, 2006; Ronkko, 2005). While some guidelines exist for the early generation of personas (Jisun An et al., 2018; Joni Salminen et al., 2020; Joni Salminen et al., 2018; Joni Salminen et al., 2019), the development and evaluation of GenAIPs still lack plausible standards. The need for standardization for GenAIPs is further accentuated by selecting appropriate methods for prompt engineering, validation approaches, and metrics assignment (Liu et al., 2024).

Hypothetical example: *A healthcare UX researcher struggles to create trustworthy GenAIPs, uncertain whether her chosen methods produce accurate user representations or not. Without standardized approaches (CC03) for developing and evaluating GenAIPs, she cannot confidently defend their validity when challenged by stakeholders.*
Observations from academia/industry: *When Amin* et al. (*Amin* et al.*, 2025*) *conducted a systematic review of 52 GenAIP research articles from 2022–2024, they found significant variability in methodological approaches across studies, with researchers using different LLM models, prompting strategies, and evaluation criteria without established standards. This methodological diversity created challenges for reproducing results and comparing study outcomes, demonstrating the lack of standardization (CC03).*

3.4.4.4. CC04: over-reliance on GenAI—it can do everything. Persona developers may struggle to integrate traditional user research with GenAI-aided persona development (Prpa et al., 2024). While GenAI offers efficient persona generation, over-reliance on fully automated processes (Chen et al., 2024) risks disconnecting personas from real user insights and contexts. This can yield a crucial shortcoming of GenAIPs, namely the erasure of the designers' reflexivity (Dorst, 2011) and personal learning through the development of manual personas (Sattele and Ortiz, 2024). This challenge is further compounded if the designers do not learn about the personas of developing them, which may be the case when applying GenAI processes in persona development. Furthermore, there is an absence of guidelines for combining direct user research with GenAI capabilities. Impressive GenAI outputs create overconfidence that reduces human oversight, disconnecting personas from real user insights and erasing designer learning through the development process.

Hypothetical example: *A design team relies entirely (CC04) on GenAIPs for their senior wellness app, skipping real user research and missing critical insights about their users' actual preferences. When the app failed in testing, they discovered that the convincing GenAIPs had led them to implement features seniors did not want while missing ones they needed.*
Observations from academia/industry: *When IDEO researchers* (*Perkel* et al.*, 2025*) *examined industry adoption of GenAI tools for UX work, multiple studies documented teams increasingly bypassing traditional user research in favor of GenAI-generated insights. The accessibility and polish of GenAI outputs created over-reliance (CC04) that reduced validation efforts, with teams accepting GenAIPs without adequate verification against real user data.*

3.4.5. Safety challenges (SC)

These challenges create potential harm through GenAI system failures and irresponsible development and deployment.

3.4.5.1. SC01: adversarial users—it can harm us. GenAIP methodologies may fall victim to manipulation. Unlike traditional data driven methods with controlled access (such as using algorithms to develop personas), GenAIPs' reliance on third-party AI models exposes them to potential exploitation by malicious actors. Through techniques like prompt injection (Schneier, 2024), attackers can manipulate GenAI tools to generate misleading personas that appear credible, compromising the integrity of the entire design process. Similarly, political actors might introduce specialized LLMs that ignore facts or create personas aligned with politically influenced narratives (Mercer et al., 2025). This vulnerability is amplified by the technical limitations and possible security deficiencies in current GenAI systems.

Hypothetical example: *In a healthcare company using GenAIPs, an adversarial actor subtly manipulates the "elderly patient" persona to downplay privacy concerns through prompt injection techniques. The resulting compromised persona, appearing coherent and credible, leads developers to implement reduced security measures that potentially expose vulnerable elderly patients' sensitive data. This illustrates how LLM vulnerabilities (SC01) in persona development can create real security risks when GenAI tools lack sufficient safeguards against manipulation.*
Observations from academia/industry: *When Liu* et al. (*Liu* et al.*, 2024*) *tested the prompt injection technique against 36 actual LLM-integrated applications and found that 31 applications* (86 %) *were vulnerable to prompt injection attacks. The research demonstrated how adversarial prompts could manipulate LLM systems to produce unintended outputs, including unauthorized data access and application prompt theft, with 10 vendors validating their findings. This demonstrates how GenAIP faces similar vulnerabilities (SC01) where malicious actors can manipulate prompt inputs to compromise persona generation integrity and produce biased or misleading user representations.*

3.4.5.2. SC02: computationally resource intensive—it is harmful for the environment. Persona developers may struggle to maintain computational complexity and the associated cost at a low level, adding to sustainability concerns (Faiz et al., 2024). This is due to the computational demands of GenAIP systems that present a significant sustainability challenge in persona development. Unlike many traditional persona development algorithms, GenAIPs require substantial computational resources for operation, particularly when using LLMs and multimodal AI systems. This excessive use of computational power negatively impacts sustainability efforts, resulting in a larger carbon footprint and contradicting the ideals of sustainable HCI (sHCI) (Hansson et al., 2021). GenAIPs require massive computational resources, creating environmental costs that contradict sustainable design principles while appearing efficient.

Hypothetical example: *Ironically, a GenAIP developed to promote green living could consume far more energy (especially with multiple iterations for prompt engineering or tuning) than it aims to save for a day, paradoxically contributing to the environmental problems it aimed to solve.*
Observations from academia/industry: *When Patterson* et al. (*Patterson* et al.*, 2021*) *analyzed the environmental impact of training large AI models, they found that creating GPT-3 consumed 1287 megawatt hours of electricity and generated 552 tons of carbon dioxide equivalent. Since Salminen* et al. (Salminen et al.) *confirms that GPT is the most frequently used model for GenAIP development, this demonstrates that GenAIPs require massive computational resources that create significant environmental costs (SC02).*

3.4.6. Privacy challenges (PC)

These challenges threaten data protection and user representation integrity.

3.4.6.1. PC01: reliance of third parties—I am not in control anymore. Persona developers struggle to maintain agency and control when using proprietary persona development tools. This can be due to closed source code and changing terms of service and functionality (Joni Salminen et al., 2021). Relying on proprietary GenAI tools creates a significant challenge in maintaining consistent control over persona development processes. Unlike traditional DDPs that often use open-source tools, GenAIPs typically depend on closed-source models like GPT (OpenAI 2023) or image generators like Midjourney (Midjourney 2024). This dependency exposes organizations to unpredictable changes in model functionality, pricing, or access policies that can disrupt established persona development workflows. GenAIPs depend on proprietary closed-source models, exposing organizations to unpredictable vendor changes while reducing control over persona development processes.

Hypothetical example: *A UX team adopts proprietary GenAI tools for healthcare persona development, significantly enhancing their output quality and efficiency. When the AI provider suddenly changes their terms of service, algorithms, and pricing structure, the team cannot maintain consistency in their personas due to the closed-source nature of the tools. This illustrates how third-party dependencies (PC01) in GenAIPs can undermine their users' professional agency.*

Observations from academia/industry: *When Amin* et al. (*Amin* et al.*, 2025*) *conducted their systematic review of 52 GenAIP research articles, they found that a notable majority of studies relied on OpenAI's GPT models for persona generation. This heavy dependence on a single proprietary provider demonstrates that the field has concentrated around closed-source models, creating strong dependence on third parties (SC01).*

3.4.7. User experience challenges (UC)

These challenges reduce AI system utility and adoption in design practice.

3.4.7.1. UC01: over-expectations—give me everything. Persona users struggle to understand the limitations of the personas (Siegel, 2010). This can be due to multiple reasons, such as a lack of technical expertise, a lack of transparency of how the personas were developed and how to use them, or the halo effect surrounding algorithmic methods. The perception of personas has shifted dramatically from skepticism about their validity to uncritical acceptance of their AI-generated versions (Tharon W. Howard, 2015). While early personas faced resistance for lacking "hard evidence" (Jisun An et al., 2018; Mesgari et al., 2015). Although GenAIPs may benefit from an AI halo effect, in which stakeholders attribute unrealistic capabilities to them simply because they are AI-generated, this overcorrection in stakeholder attitudes (Aldous et al., 2024; Bourne, 2024) creates a dangerous gap between the personas' actual limitations and users' understanding of them. *Persona users,* particularly those without technical expertise, may fail to recognize critical flaws in GenAIPs due to the opacity of AI methods and overconfidence in algorithmic solutions. They might bypass necessary validation steps, assuming GenAIPs are inherently reliable. For instance, stakeholders might readily accept stereotypical or biased representations simply because they come from an AI system rather than a "biased" human (Liu et al., 2023), highlighting how the mystification of AI methods can prevent proper critical assessment of persona quality and limitations. AI hype creates unrealistic expectations about GenAIP capabilities, leading to overconfidence in automated outputs without proper validation and critical evaluation.

Hypothetical example: *A marketing team creates GenAIPs for a new fitness app, accepting the GenAI outputs without question because they are impressed by the technology's capabilities. The personas contain subtle but significant misrepresentations of their target demographics, including unrealistic fitness goals and behaviors that do not align with market research. When the product launched based on these flawed personas, it failed to resonate with actual users. Still, the team continues to trust these GenAIPs over contradicting customer feedback because of their unwavering faith (UC01) in the GenAI's supposed objectivity.*

Observations from academia/industry: *When Survey2Persona researchers (Ilkka Kaate* et al.*, 2025*) *created personas for user interaction study using AI personas responding to user questions, it resulted in the challenge over expectations because 57 % of users accepted incorrect answers from GenAIPs, demonstrating unrealistic expectations that "GenAIPs always provide accurate responses even when insufficient data exists".*

3.4.7.2. UC02: validating the impact—I have used a persona, now what?. Persona users often struggle to validate the impact of using a persona. This challenge arises from three main factors: the lack of a systematic feedback loop to measure impact, the difficulty isolating the specific effects of personas, and the absence of well-defined metrics. The challenge of validating persona effectiveness creates uncertainty about their actual value in improving design outcomes. While organizations adopt GenAI tools and processes (Capgemini Research Institute 2024), they may struggle to measure whether GenAIPs genuinely improve design decisions or UX. This difficulty stems from three key factors: (1) the lack of established metrics for measuring persona impact, (2) the absence of systematic feedback loops between persona uses and design outcomes, and (3) the challenge of isolating persona influence from other design factors. This uncertainty affects *persona users,* who cannot justify their persona-based decisions with empirical evidence and struggle to evaluate the value of GenAIP tools and procedures. Organizations struggle to measure whether GenAIPs improve design decisions due to a lack of established metrics and feedback loops for isolating persona influence.

Hypothetical example: *A marketing team uses GenAIPs to guide their streaming platform redesign. Six months later, when executives demand evidence of ROI, the team cannot determine whether improved user retention results from their persona-informed design decisions or from simultaneous changes to content recommendations and pricing. Without established metrics (UC02) to isolate the impact of persona-based decisions from other factors, organizations cannot justify continued investment in GenAIPs despite their intuitive appeal.*

Observations from academia/industry: *When Amin* et al. (*Amin* et al.*, 2025*) *conducted their systematic review of 52 GenAIP research articles, they found that no studies measured long-term impact or validated whether GenAIPs actually improved design outcomes. This absence of impact measurement (UC02) leaves organizations unable to justify GenAIP investments or determine their effectiveness compared to traditional personas.*

3.4.7.3. UC03: desk drawer effect—will I ever use this persona again. Personas often fall victim to the 'desk drawer effect' (Portigal, 2023), according to which personas are developed but not consistently used in design decisions. This can be due to a lack of integration into daily workflows, the effort required to reference personas regularly, or the tendency to treat personas as deliverables rather than ongoing design tools (Long, 2009). While organizations invest significantly in developing sophisticated GenAIPs (Capgemini Research Institute 2024), these personas often become unused artifacts rather than active design tools, challenging the very purpose of persona development (Pruitt and Adlin, 2006). This adoption failure stems from multiple factors, including the difficulty of integrating personas into daily workflows, the effort required to keep personas relevant as project needs evolve, and the tendency to treat personas as one-time deliverables rather than ongoing design resources. The impact particularly affects design teams, who initially embrace personas but gradually make decisions without consulting them, and organizations that waste resources on developing detailed personas that ultimately do not influence design outcomes. Easy GenAIP generation may worsen adoption by creating multiple personas without the investment necessary for sustained use, treating them as deliverables rather than tools.

Hypothetical example: *A disaster relief nonprofit initially uses GenAIPs of vulnerable populations to guide their emergency alert system redesign, but as implementation deadlines approach, staff abandons these GenAIPs and reverts to making decisions based on assumptions. The unused GenAIPs ultimately become forgotten files in a project folder (UC03), despite the significant resources invested in their* development.

Observations from academia/industry: *While UC03 represents a documented concern in traditional persona research, specific empirical studies showing the desk drawer effect occurring with GenAIPs remain limited in current literature. This gap highlights the need for longitudinal research tracking how organizations actually use GenAIPs over time rather than just their initial adoption rates.*

4. Persona expert perspectives on the challenges

4.1. Participants

Seventeen subject matter experts (SMEs) participated in a survey about challenges associated with GenAIPs. The participants were recruited through professional connections, manually verifying that each participant had either published about personas or used personas actively in their research or other work. Participants ranged in age from 29 to 68 years ($M = 38.7$, SD = 11.5), with 58.8 % ($n = 10$) identifying as male and 41.2 % ($n = 7$) as female. Their experience with personas ranged from 2 to 26 years ($M = 7.6$ years, SD = 6.2 years). The participants belonged to diverse countries, including Finland, South Korea, Portugal, France, Denmark, United States, India, Lebanon, and Pakistan. Participants held various roles, including professors ($n = 6$), researchers ($n = 5$), PhD candidates ($n = 3$), post-doctoral researchers ($n = 2$), and an engineer ($n = 1$). Self-reported knowledge of traditional DDPs was moderate to high ($M = 3.8$, SD = 0.8 on a 5-point scale), while knowledge of GenAIPs was slightly lower ($M = 3.7$, SD = 0.6), indicating the participants have good knowledge on both traditional DDPs and GenAIPs.

4.2. Data collection

We collected responses from participants using an online survey on the Qualtrics platform. We pilot-tested the survey with three persona researchers and revised the survey based on the provided feedback to ensure clarity and comprehensiveness before deployment to the full sample. To provide clear common definitions, we presented traditional DDPs as *a persona created fully or partially using classical AI technologies (*e.g., *clustering)* and GenAIPs as *a persona created fully or partially using Generative AI technologies (*e.g., *LLMs)*. The survey (see the online appendix) contained three sections: (1) perceptions about GenAIP challenges, (2) comparison of GenAIPs with DDPs, and (3) demographic information. We first presented the GenAIP challenges and asked whether they agreed on the presence of the challenge on a 7-point Likert scale (1 to 7, ranging from Strongly disagree to Strongly agree). The challenges were compared between DDPs and GenAIPs using a semantic scale ("bigger problem for DDPs", "equal problem for both", and "bigger problem for GenAIPs"). To mitigate order effects, we randomized the order of statements presented. We added an open-ended question in each section to provide participants a chance to elaborate on their answers. In the demographic information section, we asked for the participant's gender, age, experience with personas, occupational role, and knowledge of DDPs and GenAIPs.

4.3. Results

[Result 1] Experts agree that each of the challenges negatively impacts GenAIP: First, we analyzed the perception of challenges by the experts (see Fig. 4). Based on the survey responses from the 17 SMEs, all challenges received mean ratings above the neutral point of 4 (neither agree nor disagree), indicating that experts agree these issues constitute challenges for GenAIPs. The most problematic challenges, defined as those with mean ratings above 5.31 (M3 of the actual data), were TC02: Hallucinations ($M = 5.94$, SD = 1.20), FC03: Over-sanitization ($M = 5.82$, SD = 1.02), CC03: Lack of Standardization ($M = 5.59$, SD = 1.00), CC01: Persona Quality Risk ($M = 5.53$, SD = 1.28), and FC01: Misrepresentation ($M = 5.47$, SD = 1.28). Even the lowest-rated challenges, including UC02: Validation of the Impact ($M = 4.29$, SD = 1.57), RC01: Superficiality ($M = 4.35$, SD = 1.37), and PC01: Reliance on Third Parties ($M = 4.35$, SD = 1.50), are above the neutral threshold. This indicates that all the pre-defined challenges are considered challenges for the GenAIPs by SMEs.

The analysis of relative standard deviation (RSD) values indicates different levels of agreement among participants regarding different challenges. RSD values across all challenges ranged from 17.4 % to 40.3 % ($M = 29.4$ %, SD = 7.3 %), indicating low to moderate levels of disagreement. Challenges demonstrating the highest agreement included FC03: Over-sanitization (RSD = 17.4 %), CC03: Lack of Standardization (RSD = 18.0 %), and TC02: Hallucinations (RSD = 20.2 %), suggesting that participants showed relatively strong agreement about the severity of these issues. Conversely, challenges exhibiting the greatest disagreement were SC02: Computationally Resource Intensive (RSD = 40.3 %), RC03: Limited Generalizability (RSD = 39.7 %), and SC01: Adversarial Users (RSD = 39.6 %), indicating that while some GenAIP's challenges are universally recognized, others may be more dependent on individual experience or organizational context.

[Result 2] Experts agree that most of the challenges are more problematic for GenAIP: Fig. 5 displays a segmented horizontal bar graph comparing SMEs' perceptions of how problematic each of the 20 challenges is for GenAIPs versus traditional DDPs. Out of the 20 challenges, 12 (60 %) were identified by the largest proportion of SMEs as more problematic for GenAIPs, including PC01: Reliance on Third Parties ($n = 15$, 88 %), TC02: Hallucinations ($n = 14$, 82 %), and SC02: Computationally Resource Intensive ($n = 13$, 76 %). Six (30 %) challenges were viewed as equally problematic for both persona types by the highest proportion of experts: UC02: Validating the Impact ($n = 14$, 82 %), UC03: Desk Drawer Effect ($n = 15$, 88 %), FC04: Complications of Average ($n = 12$, 71 %), RC03: Limited Generalizability ($n = 11$, 65 %), FC01: Misrepresentation ($n = 11$, 65 %), and FC02: Persona Driven Discrimination ($n = 14$, 82 %). Two challenges were seen as more problematic for traditional DDPs: RC04: Aggregation ($n = 7$, 41 %) and

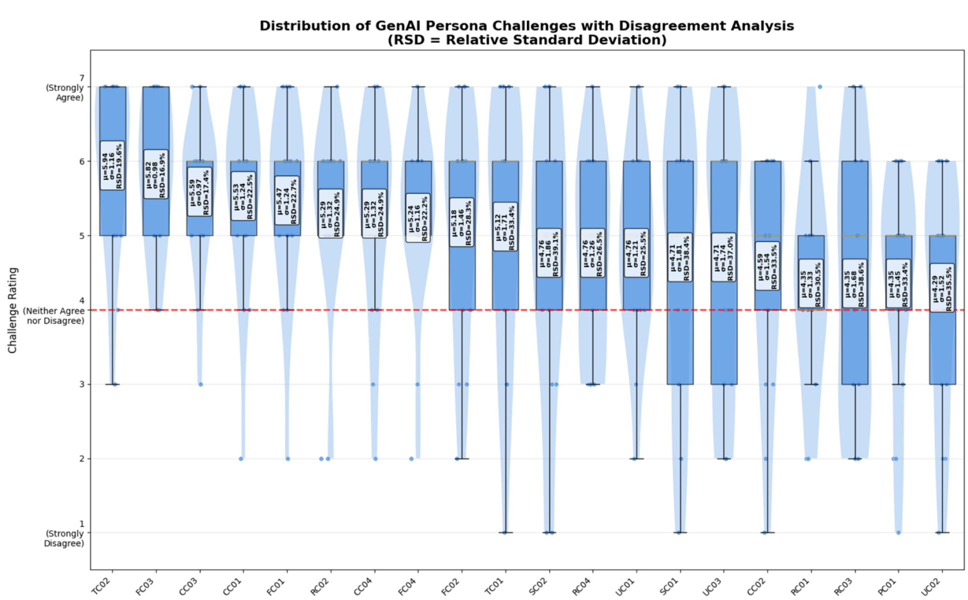

Fig. 4. Perception of GenAIP challenges by SMEs. All challenges are considered problematic for GenAIPs by the SMEs based on the mean values.

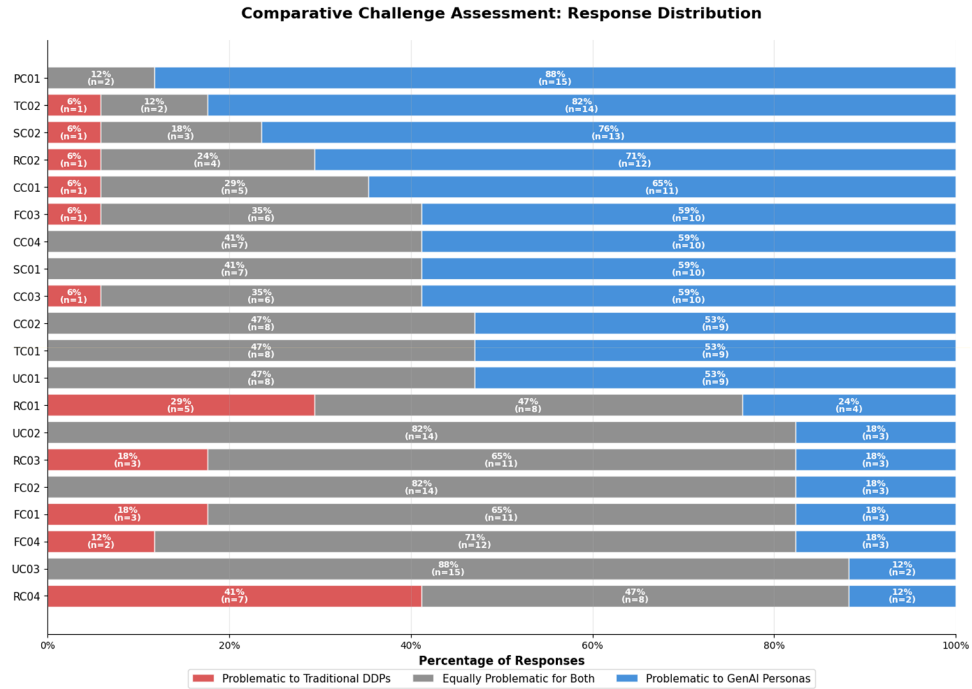

Fig. 5. Comparison of challenges between GenAIPs and traditional DDPs by SMEs. The majority of challenges (60 %) are considered more problematic for GenAIPs.

RC01: Superficiality (n = 5, 29 %). In several cases, the nature of the challenge may have informed SMEs' perceptions. For instance, hallucinations (TC02) are directly tied to the behavior of GenAI models, thus a major problem for GenAIPs. Similarly, the problem of aggregation (RC04) appears to be a more significant problem for traditional DDPs due to the statistical nature of the DDPs.

Fig. 6 presents the level of SMEs' agreement on whether each challenge is more problematic for GenAIPs, traditional DDPs, or equally problematic for both. Based on the quartile thresholds shown in the legend, 6 (30 %) challenges reached the highest consensus level (≥75 %), including UC03: Desk Drawer Effect (Equal, n = 15, 88 %), PC01: Reliance on Third Parties (GenAI, n = 15, 88 %), FC02: Persona-Driven

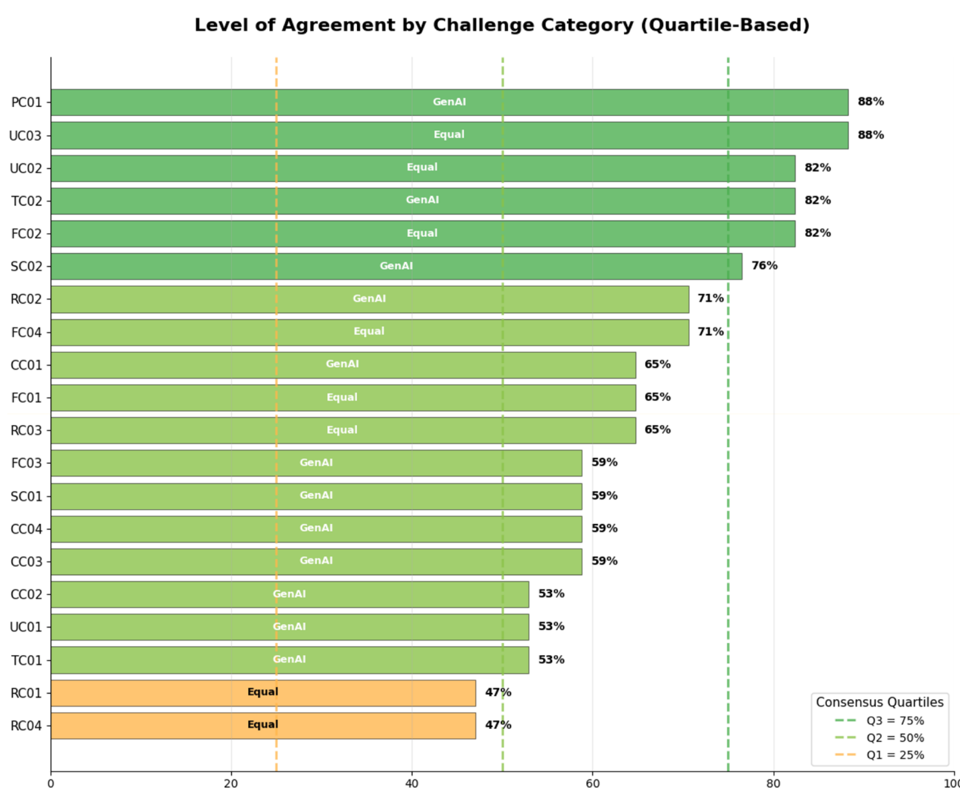

Fig. 6. Level of agreement among SMEs on comparative challenge assessment between GenAIPs and traditional DDPs. Consensus levels are categorized using quartile thresholds, with 35 % of challenges reaching the highest consensus level and no challenges falling below moderate agreement.

Discrimination (Equal, $n = 14$, 82 %), UC02: Validating the Impact (Equal, $n = 14$, 82 %), and TC02: Hallucinations (GenAI, $n = 14$, 82 %). 12 (60 %) challenges fell within the high consensus range (50–74 %), such as SC02: Computationally Resource Intensive (GenAI, $n = 13$, 76 %), RC02: Inconsistency Dilemma (GenAI, $n = 12$, 71 %), FC04: Complications of Average (Equal, $n = 12$, 71 %), CC01: Persona Quality Risk (GenAI, $n = 11$, 65 %), FC01: Misrepresentation (Equal, $n = 11$, 65 %), and RC03: Limited Generalizability (Equal, $n = 11$, 65 %). Only 2 (10 %) challenges showed moderate consensus (25–49 %), including RC01: Superficiality (Equal, $n = 8$, 47 %) and RC04: Aggregation (Equal, $n = 8$, 47 %). No challenges fell below the 25 % threshold, indicating that even the most controversial issues maintained at least moderate levels of agreement among experts (Table 2).

4.4. Stakeholder impact analysis

The stakeholder impact analysis reveals a critical distinction between operational challenges and consequential harm in GenAIPs. While *persona developers* encounter the most operational challenges (17 out of 20 issues) and *persona users* face significant implementation difficulties (13 challenges), this distribution primarily reflects the technical complexity and validation responsibilities inherent in the development process rather than the ultimate severity of impact. Persona developers should navigate a transformed landscape requiring new competencies in prompt engineering, AI output validation, bias detection, and security considerations, while simultaneously maintaining the fundamental goal of creating accurate user representations without established standards or clear guidance.

However, the seemingly lower number of challenges directly impacting *target groups (*4 out of 20) obscures a more troubling reality that these communities bear the most severe and lasting consequences when GenAIP systems fail. The challenges they face (FC01, FC02, RC03, and SC01) represent fundamental violations of the core purpose of persona development, which is to ensure user needs are understood and addressed. While *practitioners* experience operational friction and workflow disruptions, target groups experience systemic exclusion from products and services designed without understanding their actual needs, characteristics, or contexts. This asymmetric impact pattern highlights a critical concern in GenAIPs. The technical burden of GenAIP challenges may fall on practitioners, but the ultimate consequences of failure disproportionately harm the very communities these personas are meant to represent and protect, making robust validation processes and ethical oversight not just methodological necessities but moral imperatives.

5. Discussion

This section examines the implications of GenAIP challenges for theoretical, practical and future needs. GenAIPs shift persona development from human-interpretable processes to AI-driven generation with limited oversight. Our stakeholder analysis reveals practitioners handle technical complexity while target communities bear serious consequences through misrepresentation and bias. GenAI transforms persona challenges qualitatively, creating systematic issues like algorithmic bias and validation opacity that persist despite prompt engineering. Current research priorities emphasize technical capabilities over practical implementation. The discussion provides HCAI-based guidelines, acknowledges study limitations, and outlines future research directions.

5.1. Theoretical implications

GenAIPs shift persona development into human-centered AI territory: Our application of Shneiderman's HCAI framework categorizes twenty GenAIP challenges across seven dimensions, revealing fundamental differences between GenAIPs and traditional persona development methods. The shift to HCAI territory is evidenced by GenAI

Table 2
Stakeholder Impact Analysis.

Challenge	Persona Developers	Persona Users	Target Groups
FC01: Misrepresentation	Struggle to validate minority group representation (Chen et al., 2024; Cachat-Rosset and Klarsfeld, 2023)	Risk of creating exclusionary products from biased representations	Face exclusion when needs remain uncaptured
RC01: Superficiality	Face validation challenges from contradictory personas	Cannot trust unreliable representations	–
RC03: Limited Generalizability	Must balance specificity with adaptability (Li et al., 2025; Nah et al., 2023)	Cannot apply personas across varied contexts	Risk misapplication of behaviors across contexts
RC02: Inconsistency Dilemma	Get different personas from same data (Jansen et al., 2020)	Uncertain which persona version to trust	–
SC02: Computationally Resource Intensive	Trade-off between efficiency and quality (Joni OpenAI 2025)	Face high costs conflicting with sustainability goals	–
CC03: Lack of Standardization	Cannot justify methods without standards	Cannot compare or assess persona reliability	–
FC03: Over-sanitization	Cannot generate realistic negative traits (HLEG, 2019)	Make decisions on artificially positive personas	–
CC04: Over-reliance on GenAI	Struggle to integrate user research (Joni OpenAI 2025)	Cannot distinguish AI content from real insights (Schön, 1992)	–
FC04: Complications of Average	Cannot preserve important user variations	Design for nonexistent middle ground (Sattele and Ortiz, 2024)	–
PC01: Reliance of Third Parties	Limited customization and vendor disruptions	Forced into third-party dependence	–
SC01: Adversarial Users	Must implement safeguards against manipulation (Midjourney 2024; Du et al., 2024)	Cannot detect manipulated personas	Harmed by design decisions from compromised personas
CC02: Manual Resource Intensiveness	Need new AI oversight expertise	Must develop validation skills	–
CC01: Persona Quality Risk	May lack training to detect problems (Capgemini Research Institute 2024)	–	–
RC04: Aggregation	Risk prioritizing statistics over insights	Make decisions on algorithmic artifacts	–
TC02: Hallucinations	Cannot verify the accuracy of fluent content	Make decisions despite fabricated details	–
FC02: Persona Driven Discrimination	–	Inadvertently favor well-represented groups	Face discrimination from biased personas
UC01: Over-expectations	–	Cannot recognize flaws due to AI overconfidence	–
TC01: Lack of Validation	–	Cannot establish verification processes	–
UC02: Validating the Impact	–	Cannot justify decisions with evidence	–
UC03: Desk Drawer Effect	–	Gradually stop consulting personas	–

functioning as an active agent in persona creation rather than a passive analytical tool. Unlike traditional DDPs that rely on statistical analysis of user data with human interpretation, GenAIPs involve AI systems making autonomous decisions about user characteristics, narratives, and representations without direct human oversight of each decision. This autonomy introduces core HCAI concerns absent in traditional methods: algorithmic transparency, fairness, and human control. When AI systems generate user representations, questions of algorithmic bias, explainable decision-making, and human oversight become central design considerations rather than peripheral concerns. SME responses validate this theoretical positioning, with all challenges receiving mean ratings above the neutral point, indicating universal recognition that these issues constitute legitimate challenges for GenAIPs. The highest-rated challenges of Hallucinations (M = 5.94, SD=1.20), Over-sanitization (M = 5.82, SD=1.02), and Lack of Standardization (M = 5.59, SD=1.00) cluster around HCAI principles of transparency and reliability, confirming that SMEs recognize these as manifesting distinctly from than those encountered in traditional persona development.

GenAIPs affect different stakeholders differently: Our stakeholder impact analysis reveals a troubling problem, while GenAIPs create many technical challenges for practitioners, the communities these personas represent face much worse consequences, even though they encounter fewer direct challenges. This creates *"consequence displacement,"* where persona creators deal with workflow problems while target communities face systematic exclusion. The technical challenges affecting developers and users can be managed with better tools and methods. However, challenges affecting target groups can cause lasting harm through exclusionary products. When GenAI systems misrepresent marginalized groups, the damage reinforces inequalities at scale. One SME (P13) explained*: "I have concerns that using GenAI might naturally lead us to overlook minority perspectives. However, when humans are involved in examining and interpreting data, even if minorities don't have strong representation, they naturally get included in our considerations. But I don't expect GenAI, being a statistical model, to do the same."* This pattern shows that using GenAIPs without strong ethical protections risks making inequalities worse while shifting harm from practitioners to vulnerable communities.

GenAIPs have transformed traditional persona development challenges: The comparative assessment reveals that GenAI has fundamentally altered the ontological nature of persona development challenges, creating what we conceptualize as *"evolutionary amplification"* where persistent methodological issues undergo qualitative transformation rather than simple intensification. Traditional persona challenges have evolved through *three* distinct transformation pathways that fundamentally change their character, scope, and resolution strategies. First*, scale transformation* occurs when human-level biases become algorithmic biases that operate across multiple GenAIPs with consistent patterns, transforming isolated incidents into systematic discrimination. Second*, visibility transformation* emerges when manual inconsistencies, previously detectable through human review, become AI hallucinations that produce convincing but fabricated content indistinguishable from legitimate insights without targeted evaluations. Third*, agency transformation* shifts validation difficulties from human interpretive challenges to complete opacity problems where the reasoning behind AI-generated characteristics becomes fundamentally unknowable, even to the GenAI's creators. These transformations represent qualitative rather than merely quantitative changes in persona development challenges, creating entirely new categories of methodological problems that require novel solution approaches. One SME (P03) noted that *"GenAI has some unique aspects, the most impactful being that someone can create realistic looking personas with little know-how or data,"* highlighting how GenAI's accessibility democratizes persona creation while simultaneously undermining quality control mechanisms developed over decades of traditional practice. This evolution demands recognition that GenAIP challenges are not simply traditional problems requiring traditional solutions with AI assistance, but require mitigation strategies specifically designed for GenAIPs.

GenAIPs research needs to shift from technical to practical and contextual domain: Given these transformed challenges, current research approaches require reconsideration. The concentration of challenges in development rather than application stages indicates fundamental misalignment in research priorities within the HCI community. Current GenAIP research has focused heavily on technical capabilities and generation methods while neglecting practical implementation and impact assessment. This pattern reflects a broader trend in AI research where technological sophistication receives more attention than real-world utility and consequences. SME agreement levels varied significantly across challenges, with the highest agreement on transparency-related issues like hallucinations and over-sanitization, and the lowest agreement on context-dependent challenges such as adversarial users and computational intensity. This variation suggests that while some GenAIP problems are universally recognized, others seem to be more dependent on specific organizational contexts or implementation approaches.

GenAIP challenges follow a similar yet unique trend as DDP challenges: When positioned within existing persona literature, these priority gaps become more apparent. Our findings align with Salminen et al.'s (Joni Salminen et al., 2021) identification of bias and validation challenges in DDPs. However, we diverge by showing GenAI amplifies these issues through algorithmic automation. While DDPs mitigated some manual limitations, our results reveal 60 % of challenges are more problematic for GenAIPs than traditional approaches. The continued emphasis on development-stage challenges also highlights a critical gap in evaluation methodologies. The field lacks robust frameworks for evaluating whether GenAIPs actually improve design decisions or user outcomes. Without such evaluation mechanisms, the technology risks becoming an end in itself rather than a means to better user-centered design, contradicting the fundamental purpose of persona development as a tool for understanding and serving real users.

GenAIP challenges are the same for all models: These systematic challenges require acknowledgment of contextual factors while maintaining focus on fundamental issues. While we present GenAIP challenges as general patterns, different models (GPT-4, Claude, Gemini, etc.) exhibit varying behaviors and limitations. Our framework addresses systemic issues that persist across GenAIPs, though specific manifestations may vary by model architecture. This variation suggests that practitioners should evaluate challenges within their specific implementation context while using our framework as a general guide for potential issues. Nevertheless, these implementation-specific considerations should not obscure the systematic nature of the challenges we identified.

GenAIP challenges need more than prompt engineering to fix: Similarly, some challenges can be mitigated through prompt engineering and iterative refinement, such as improving narrative consistency (RC02) or reducing obvious stereotypes in generated content (FC02). Salminen et al. (Salminen et al.) observed that prompt design varies substantially across GenAIP studies, with only 13.5 % documenting systematic prompt engineering methodologies, while Ayach et al. (Ayach et al., 2025) found that prompt-engineering-based approaches can support proto-persona generation efficiency and effectiveness. However, prompt engineering cannot address fundamental systematic issues inherent to GenAI architectures. For instance, hallucinations (TC02) persist because they stem from the probabilistic nature of language models, not prompt design (Xu et al., 2025). Bias amplification (FC02) reflects training data limitations that prompting cannot eliminate, as biases are often embedded in the enormous Internet data used for training and can be amplified rather than mitigated through model architecture and training procedures (Yufei Guo et al., 2024). Validation opacity (TC01) exists because GenAI systems remain black boxes regardless of prompt sophistication, as their internal workings remain poorly understood even by their creators, making independent auditing and validation extremely challenging. These core challenges require architectural changes, new validation frameworks, and human oversight mechanisms that extend far beyond prompt optimization. Our analysis demonstrates that while prompt engineering offers valuable improvements for certain specific challenges, the fundamental transformation of persona development from human-interpretable processes to opaque GenAI generation introduces systemic issues that require broader methodological reconsideration beyond prompt optimization alone.

5.2. Practical guidelines

Since we categorized GenAIP challenges according to Shneiderman's seven HCAI principles, we organize our practical guidelines using the same framework for coherent alignment between problems and solutions. Specifically, these actions could be taken for each of the different aspects of the HCAI principles:

- For challenges related to ***transparency***, create **persona genealogy systems** that trace not just data sources but the specific algorithmic pathways, prompt engineering decisions, and human interventions that shaped each persona characteristic, enabling practitioners to understand why AI generated particular traits and how to systematically modify them (TC01, TC02). Technical suggestion: *Deploy automated documentation systems that capture prompt iteration histories, model temperature settings, few-shot examples used, and human feedback loops, creating searchable databases that allow teams to identify which generation strategies produce the most accurate representations for specific user demographics.*
- For challenges related to ***fairness***, implement **systematic counterfactual analysis** workflows where identical persona generation processes are run with demographic attributes modified (e.g., changing names from "Jennifer" to "Jamal") to detect and quantify algorithmic bias, then use these findings to develop bias correction algorithms that operate at the prompt engineering level (FC01, FC02, FC04). Technical suggestion: *Build automated bias testing pipelines that generate persona variants across protected characteristics, measure statistical disparities in occupational assignments, behavioral descriptions, and socioeconomic indicators, and automatically flag generation parameters that produce discriminatory patterns exceeding defined thresholds.*
- For ***reliability*** related challenges, create **domain-specific expert validation networks** where personas undergo systematic review by practitioners from the target domain (e.g., healthcare professionals for patient personas), cultural community representatives, and methodological experts using structured evaluation protocols that identify both superficial inconsistencies and deeper authenticity issues (RC01, RC03). Technical suggestion: *Develop consensus-building platforms where multiple expert reviewers independently evaluate personas using standardized rubrics, then use disagreement patterns to identify aspects requiring additional validation or revealing systematic generation limitations.*
- For challenges related to the ***control*** domain of HCAI, design **human-AI collaboration workflows** with explicit control boundaries where AI handles initial generation while humans retain authority over validation, modification, and deployment decisions, using structured handoff protocols that prevent automation bias while leveraging AI efficiency (CC01, CC04). Technical suggestion: *Build staging environments where GenAIPs must pass through multiple human review gates (methodological, cultural, domain-specific) before reaching production use, with clear escalation procedures when reviewers identify concerns that require generation parameter modifications or complete regeneration.*
- For challenges related to ***privacy***, implement **data preservation techniques** that add encryption to training data and generation processes, ensuring individual users cannot be identified or re-identified through persona characteristics while maintaining statistical utility for design decisions (PC01). Technical suggestion: *Apply formal differential privacy mechanisms during both training data preparation and persona generation, with epsilon budgets allocated across different demographic attributes and privacy impact assessments conducted for each persona deployment context.*
- For challenges related to ***safety***, conduct **red-teamed adversarial evaluation** where dedicated teams attempt to manipulate GenAIP systems through prompt injection, bias amplification attacks, and edge case exploitation, then use these findings to develop robust safeguards and update generation protocols (SC01, SC02). Technical suggestion: *Deploy continuous adversarial monitoring systems that test persona generation resilience against known attack vectors, automatically updating prompt templates and safety filters based on emerging threat patterns, while maintaining computational efficiency through selective model downsizing for non-critical applications.*
- For challenges related to ***user experience***, establish **systematic tracking of persona influence** on actual design decisions and user outcomes through longitudinal studies that measure whether GenAIP-informed products better serve their intended users compared to traditional development approaches, creating feedback loops that improve future persona generation (UC02, UC03). Technical suggestion: *Deploy impact measurement frameworks that track design decision lineage from specific personas to implemented features to user satisfaction metrics, using A/B testing methodologies to compare products developed with GenAIPs* versus *traditional personas, and feeding these results back into generation parameter optimization.*

5.3. Limitations

The limitations of our work warrant careful consideration. *First*, our analysis primarily relies on literature review and expert survey data, lacking empirical validation of the identified challenges through controlled studies of real-world GenAIP implementations. *Second,* while we categorized challenges using HCAI principles, the rapid evolution of GenAI technologies means some challenges may already be transforming, or new ones may be emerging that our framework does not capture. *Third,* our stakeholder impact analysis, though systematic, may not capture the full complexity of how these challenges manifest across different organizational contexts, cultural settings, or specific industry sectors. *Fourth,* the proposed guidelines, while grounded in current understanding and real-world cases, require empirical validation to assess their effectiveness in mitigating the identified challenges. *Fifth,* as our analysis primarily examines personas through the lens of HCAI for technology-centered UX design (Adlin and Pruitt, 2009; Caballero et al., 2014), this focus acknowledges but may not fully capture certain aspects of persona development in design practice, particularly the role of designer reflexivity and learning through manual persona creation (Dorst, 2011; Schön, 1992). *Sixth,* our focus has primarily been on the technology and methodology aspects of GenAIP development, while future studies could examine deeper aspects of how GenAIPs function within broader design research practices and organizational workflows. *Finally,* our emphasis on challenges may under-represent potential opportunities and benefits that GenAIPs could offer when implemented with appropriate safeguards and human oversight.

5.4. Future research directions

Future research in GenAIPs needs to address two fundamental dimensions that emerge from our analysis. The first dimension is based on the **limitations of our current study,** highlighting the need for expansion of scientific research on the critique of GenAIPs. This includes (1) conducting *systematic studies of GenAIP implementations* across different organizational contexts to validate our challenge categorizations, (2) examining the *HCAI-based framework application to GenAIPs* and its adaptation if need be, (3) investigating domain-*specific manifestations of these challenges in various industries* and use cases, and (4) evaluating the *effectiveness of our proposed guidelines through longitudinal* *studies of actual GenAIP* implementations. This dimension would reinforce the theoretical foundation of the critique of the GenAIPs.

The second dimension **focuses on advancing GenAIP development and implementation**, which can be done by focusing on six critical research areas. First, developing *systematic validation methods*, including metrics for measuring GenAIP quality (especially for bias, fairness, and accuracy) and tools for detecting inconsistencies or hallucinations in generated personas. Second, *conducting comparative empirical studies between MPs, DDPs, and GenAIPs to measure their influence on design decisions and evaluate their effectiveness in representing user needs*. Third, *investigating GenAIPs' performance across diverse cultural and social contexts*, including methods for cultural adaptation and frameworks for ensuring cultural authenticity. Fourth, *performing an in-depth analysis of the ethical limitations of GenAIPs and establishing ethical guidelines specifically for GenAIPs*, addressing privacy implications, consent frameworks, and bias mitigation strategies. These could include, for example, various frameworks developed under Fairness, Accountability, and Transparency (FAccT) (Ziyang Guo et al., 2024; Laufer et al., 2022), Fairness, Accountability, Transparency, and Ethics (FATE) in AI (Bird et al., 2020; Chmielinski et al., 2024; Liao et al., 2023; Magooda et al., 2023), and EU guidelines on Trustworthy AI (HLEG, 2019). Their application could be based on systematically mapping the challenges to the ethics framework and analyzing how these AI ethics frameworks could propose solutions to these challenges for GenAIPs. Fifth, *exploring integration approaches that combine traditional persona development methods with GenAI capabilities,* including optimal workflows and frameworks for multiple data sources. Sixth, *future GenAIP research should expand beyond development-focused investigations to examine practical application issues*, particularly how these personas influence design decisions and user outcomes in real-world settings. Seventh, *future studies should explore perspectives from other domains beyond HCAI for technology-centered UX design*, investigating how "reflection-in-action" processes fundamental to design practice (Papagiannidis et al., 2025) can be maintained when using GenAIPs, and examining how the abductive thinking and iteration central to design thinking (Chhikara, 2025) can be supported in AI-assisted persona development. Eighth, *research should investigate deeper aspects of how GenAIPs function within broader design research practices and organizational workflows*, developing frameworks for optimal human-AI collaboration in persona development that preserve the learning benefits designers gain from traditional persona creation processes.

6. Conclusion

Our analysis of twenty GenAIP challenges using HCAI principles reveals that GenAI technologies fundamentally transform persona development from traditional HCI methodology into HCAI territory. We demonstrate that while persona developers face numerous technical challenges, target user groups bear the most severe consequences through systemic misrepresentation and exclusion, with GenAIPs transforming rather than eliminating existing persona development issues.

Like Dijkstra's warning about goto statements complicating rather than simplifying code, we find that GenAIPs often make existing persona challenges more complex rather than solving them. While GenAIPs offer apparent efficiency, they risk reducing complex human behaviors to oversimplified stereotypes, potentially leading to biased design decisions and exclusionary products. The central challenge is not merely improving generation efficiency but ensuring these tools enhance rather than diminish our understanding of real users. The path forward requires recognizing GenAIPs as collaborative tools requiring sophisticated human-AI partnerships and prioritizing the welfare of represented communities over technical sophistication.

CRediT authorship contribution statement

Danial Amin: Writing – original draft, Writing – review & editing, Data curation, Methodology, Investigation, Formal analysis. **Joni Salminen:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization. **Bernard J. Jansen:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization. **Joongi Shin:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization. **Dae Hyun Kim:** Writing – original draft, Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

AI Agents vs. Agentic AI: a Conceptual Taxonomy, Applications and Challenges. Retrieved August 17, 2025 from https://arxiv.org/html/2505.10468v1.

A review of green artificial intelligence: Towards a more sustainable future - ScienceDirect. Retrieved July 14, 2025 from https://www.sciencedirect.com/science/article/pii/S0925231224008671.

Synthetic Users: If, When, and How to Use AI-Generated "Research." Nielsen Norman Group. Retrieved August 12, 2025 from https://www.nngroup.com/articles/synthetic-users/.

Adlin, Tamara, Pruitt, John, 2009. Putting Personas to work: Using data-Driven Personas to Focus Product planning, design, and Development. Human-Computer Interaction: Development Process, p. 95 (2009).

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

Aldous, Kholoud, Salminen, Joni, Farooq, Ali, Jung, Soon-Gyo, Jansen, Bernard, 2024. Using ChatGPT in Content Marketing: Enhancing Users' Social Media Engagement in Cross-Platform Content Creation through Generative AI. In: Proceedings of the 35th ACM Conference on Hypertext and Social Media (HT '24), September 10, 2024. New York, NY, USA. Association for Computing Machinery, pp. 376–383. https://doi.org/10.1145/3648188.3675142.

Alessa, Abeer, Al-Khalifa, Hend, 2023. Towards Designing a ChatGPT Conversational Companion for Elderly People. In: Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23), August 10, 2023. New York, NY, USA. Association for Computing Machinery, pp. 667–674. https://doi.org/10.1145/3594806.3596572.

Al-kfairy, Mousa, Mustafa, Dheya, Kshetri, Nir, Insiew, Mazen, Alfandi, Omar, 2024. Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. Informatics 11 (3), 58. https://doi.org/10.3390/informatics11030058 (September 2024).

Danial Amin, Joni Salminen, Farhan Ahmed, Sonja M.H. Tervola, Sankalp Sethi, and Bernard J. Jansen. 2025. How Is Generative AI Used for Persona Development?: a Systematic Review of 52 Research Articles. https://doi.org/10.48550/arXiv.2504.04927.

An, Jisun, Kwak, Haewoon, Jansen, Bernard J., 2016. Towards Automatic Persona Generation Using Social Media. In: Proceedings of Third International Symposium on Social Networks Analysis, Management and Security (SNAMS 2016), The 4th International Conference on Future Internet of Things and Cloud. Vienna, Austria. IEEE. August 22, 2016.

An, Jisun, Kwak, Haewoon, Jansen, Bernard J., 2017. Personas for content creators via decomposed aggregate audience statistics. In: 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, pp. 632–635, 2017.

An, Jisun, Kwak, Haewoon, Jung, Soon-gyo, Salminen, Joni, Jansen, Bernard J., 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), 54. https://doi.org/10.1007/s13278-018-0531-0, 2018.

An, Jisun, Kwak, Haewoon, Salminen, Joni, Jung, Soon-gyo, Jansen, Bernard J., 2018. Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. ACM Transactions on the Web (TWEB) 12 (4), 27. https://doi.org/10.1145/3265986, 2018.

Elina Andrus. 2023. How to navigate the risks and limitations of third-party AI solutions. DEPT®. Retrieved July 15, 2025 from https://www.deptagency.com/insight/how-to-navigate-the-risks-and-limitations-of-third-party-ai-solutions/.

Jacy Anthis, Kristian Lum, Michael Ekstrand, Avi Feller, Alexander D'Amour, and Chenhao Tan. 2024. The Impossibility of Fair LLMs. https://doi.org/10.48550/arXiv.2406.03198.

Argyle, Lisa P., Busby, Ethan C., Fulda, Nancy, Gubler, Joshua R., Rytting, Christopher, Wingate, David, 2023. Out of One, Many: Using Language Models to Simulate Human Samples. Polit. Anal. 31 (3), 337–351. https://doi.org/10.1017/pan.2023.2. July 2023.

Mohammad Atari, Mona J. Xue, Peter S. Park, Damián Blasi, and Joseph Henrich. 2023. Which Humans? https://doi.org/10.31234/osf.io/5b26t.

Fernando Ayach, Vitor Lameirão, Raul Leão, Jerfferson Felizardo, Rafael Sobrinho, Vanessa Borges, Patrícia Matsubara, and Awdren Fontão. 2025. Generating Proto-Personas through Prompt Engineering: a Case Study on Efficiency, Effectiveness and Empathy. https://doi.org/10.48550/arXiv.2507.08594.

Bender, Emily M., Gebru, Timnit, McMillan-Major, Angelina, Shmitchell, Shmargaret, 2021. On the dangers of stochastic parrots: Can language models be too big?. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, pp. 610–623, 2021.

Bender, Emily M., Gebru, Timnit, McMillan-Major, Angelina, Shmitchell, Shmargaret, 2021. On the dangers of stochastic parrots: Can language models be too big?. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, pp. 610–623, 2021.

Sarah Bird, Miro Dudík, Richard Edgar, Brandon Horn, Roman Lutz, Vanessa Milan, Mehrnoosh Sameki, Hanna Wallach, and Kathleen Walker. 2020. Fairlearn: a toolkit for assessing and improving fairness in AI. (May 2020). Retrieved February 20, 2025 from https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/.

Bolukbasi, Tolga, Chang, Kai-Wei, Zou, James Y., Saligrama, Venkatesh, Kalai, Adam T., 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems 29, 2016.

Bourne, Clea, 2024. AI hype, promotional culture, and affective capitalism. AI and Ethics 1–13, 2024.

Branco, Karina da S.C., Oliveira, Rhenara A., da Silva, Francisco L.Q., Rabelo, Jacilane de H., Marques, Anna B.S., 2020. Does this persona represent me? investigating an approach for automatic generation of personas based on questionnaires and clustering. In: Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems (IHC '20). New York, NY, USA. Association for Computing Machinery, pp. 1–6. https://doi.org/10.1145/3424953.3426648. October 26, 2020.

Buolamwini, Joy, Gebru, Timnit, 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on Fairness, Accountability and Transparency. PMLR, pp. 77–91, 2018.

Caballero, Leydi, Moreno, Ana M., Seffah, Ahmed, 2014. Persona as a Tool to Involving Human in Agile Methods: Contributions from HCI and Marketing. Human-Centered Software Engineering (Lecture Notes in Computer Science). Springer, Berlin, Heidelberg, pp. 283–290. https://doi.org/10.1007/978-3-662-44811-3_20, 2014.

Gaelle Cachat-Rosset and Alain Klarsfeld. 2023. Diversity, Equity, and Inclusion in Artificial Intelligence: An Evaluation of Guidelines. (2023). Retrieved from https://doi.org/10.1080/08839514.2023.2176618.

Capgemini Research Institute. 2024. Harnessing the value of generative AI. Retrieved from https://www.capgemini.com/wp-content/uploads/2024/11/Generative-AI-in-Organizations-Refresh_25112024.pdf.

Chang, Yen-ning, Lim, Youn-kyung, Stolterman, Erik, 2008. Personas: From Theory to Practices. In: Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges (NordiCHI '08). New York, NY, USA. ACM, pp. 439–442. https://doi.org/10.1145/1463160.1463214, 2008.

Chapman, Chris, Love, Edwin, Milham, Russell P., ElRif, Paul, Alford, James L., 2008. Quantitative Evaluation of Personas as Information. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1107–1111. https://doi.org/10.1177/154193120805201602. September 01, 2008.

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

Chen, Chaoran, Li, Weijun, Song, Wenxin, Ye, Yanfang, Yao, Yaxing, Li, Toby Jia-Jun, 2024. An Empathy-Based Sandbox Approach to Bridge the Privacy Gap among Attitudes, Goals, Knowledge, and Behaviors. In: Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). New York, NY, USA. Association for Computing Machinery, pp. 1–28. https://doi.org/10.1145/3613904.3642363. May 11, 2024.

Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. arXiv preprint (2023).

Prateek Chhikara. 2025. Mind the Confidence Gap: Overconfidence, Calibration, and Distractor Effects in Large Language Models. https://doi.org/10.48550/arXiv.2502.11028.

Kasia Chmielinski, Sarah Newman, Chris N. Kranzinger, Michael Hind, Jennifer Wortman Vaughan, Margaret Mitchell, Julia Stoyanovich, Angelina McMillan-Major, Emily McReynolds, Kathleen Esfahany, Mary L. Gray, Audrey Chang, and Maui Hudson. 2024. The CLeAR Documentation Framework for AI Transparency: Recommendations for Practitioners & Context for Policymakers. (April 2024). Retrieved February 20, 2025 from https://www.microsoft.com/en-us/research/publication/the-clear-documentation-framework-for-ai-transparency-recommendations-for-practitioners-context-for-policymakers/.

Zhibo Chu, Zichong Wang, and Wenbin Zhang. 2024. Fairness in Large Language Models: a Taxonomic Survey. https://doi.org/10.48550/arXiv.2404.01349.

Connor, Paul, Weeks, Matthew, Glaser, Jack, Chen, Serena, Keltner, Dacher, 2023. Intersectional implicit bias: Evidence for asymmetrically compounding bias and the predominance of target gender. Journal of Personality and Social Psychology 124 (1), 22–48. https://doi.org/10.1037/pspa0000314, 2023.

Marios Constantinides, Himanshu Verma, Shadan Sadeghian, and Abdallah El Ali. 2025. The Future of Work is Blended, Not Hybrid. https://doi.org/10.48550/arXiv.2504.13330.

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

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

Cooper, Alan, Reimann, Robert, Cronin, David, 2007. About Face 3: The essentials of Interaction Design. John Wiley & Sons.

Dominello, Amanda, Cheek, Colleen, Clay-Williams, Robyn, 2025. The use of personas to involve consumers in healthcare co-design: a scoping review. BMC Health Services Research 25 (1), 649. https://doi.org/10.1186/s12913-025-12786-7. May 2025.

Dorst, Kees, 2011. The core of 'design thinking' and its application. Design Studies 32 (6), 521–532. https://doi.org/10.1016/j.destud.2011.07.006. November 2011.

Du, Xuejun, An, Pengcheng, Leung, Justin, Li, April, Chapman, Linda E., Zhao, Jian, 2024. DeepThInk: Designing and probing human-AI co-creation in digital art therapy. International Journal of Human-Computer Studies 181, 103139. https://doi.org/10.1016/j.ijhcs.2023.103139. January 2024.

Delve A.I. Top 12 Persona Tools & Templates. Retrieved August 24, 2025 from https://www.delve.ai/blog/persona-template-tools.

Ahmad Faiz, Sotaro Kaneda, Ruhan Wang, Rita Osi, Prateek Sharma, Fan Chen, and Lei Jiang. 2024. LLMCarbon: Modeling the end-to-end Carbon Footprint of Large Language Models. https://doi.org/10.48550/arXiv.2309.14393.

Farquhar, Sebastian, Kossen, Jannik, Kuhn, Lorenz, Gal, Yarin, 2024. Detecting hallucinations in large language models using semantic entropy. Nature 630 (8017), 625–630. https://doi.org/10.1038/s41586-024-07421-0. June 2024.

Floyd, Ingbert R., Jones, M.Cameron, Twidale, Michael B., 2008. Resolving Incommensurable Debates: a Preliminary Identification of Persona Kinds, Attributes, and Characteristics. Artifact 2 (1), 12–26. https://doi.org/10.1080/17493460802276836. April 2008.

Friess, Erin, 2012. Personas and decision making in the design process: an ethnographic case study. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1209–1218. https://doi.org/10.1145/2207676.2208572, 2012.

Guan, Kathleen W., Salminen, Joni, Jung, Soon-Gyo, Jansen, Bernard J., 2023. Leveraging Personas for Social Impact: a Review of Their Applications to Social Good in Design. International Journal of Human–Computer Interaction 1–16. https://doi.org/10.1080/10447318.2023.2247568, 0, 0 (September 2023).

Yufei Guo, Muzhe Guo, Juntao Su, Zhou Yang, Mengqiu Zhu, Hongfei Li, Mengyang Qiu, and Shuo Shuo Liu. 2024. Bias in Large Language Models: Origin, Evaluation, and Mitigation. https://doi.org/10.48550/arXiv.2411.10915.

Guo, Ziyang, Wu, Yifan, Hartline, Jason D., Hullman, Jessica, 2024. A Decision Theoretic Framework for Measuring AI Reliance. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24). New York, NY, USA. Association for Computing Machinery, pp. 221–236. https://doi.org/10.1145/3630106.3658901. June 05, 2024.

Shashank Gupta, Vaishnavi Shrivastava, Ameet Deshpande, Ashwin Kalyan, Peter Clark, Ashish Sabharwal, and Tushar Khot. 2023. Bias Runs Deep: Implicit Reasoning Biases in Persona-Assigned LLMs. October 13, 2023. . Retrieved August 6, 2024 from https://openreview.net/forum?id=kGteeZ18Ir.

Shashank Gupta, Vaishnavi Shrivastava, Ameet Deshpande, Ashwin Kalyan, Peter Clark, Ashish Sabharwal, and Tushar Khot. 2024. Bias Runs Deep: Implicit Reasoning Biases in Persona-Assigned LLMs. https://doi.org/10.48550/arXiv.2311.04892.

Philipp Hacker, Brent Mittelstadt, Frederik Zuiderveen Borgesius, and Sandra Wachter. 2024. Generative Discrimination: What Happens When Generative AI Exhibits Bias, and What Can Be Done About It. Retrieved October 27, 2024 from http://arxiv.org/abs/2407.10329.

Hansson, Lon Åke Erni Johannes, Pargman, Teresa Cerratto, Pargman, Daniel Sapiens, 2021. A Decade of Sustainable HCI: Connecting SHCI to the Sustainable Development Goals. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Yokohama Japan. ACM, pp. 1–19. https://doi.org/10.1145/3411764.3445069. May 06, 2021.

HLEG, A.I., 2019. Ethics Guidelines For Trustworthy AI. European Commission. Available at: https://ec.europa.eu/digital-single….

Matthew K. Hong, Shabnam Hakimi, Yan-Ying Chen, Heishiro Toyoda, Charlene Wu, and Matt Klenk. 2023. Generative AI for Product Design: Getting the Right Design and the Design Right. arXiv preprint (2023).

Howard, Charity, 2015. The dangers of personas. Interactions 22 (2), 42–45, 2015.

Howard, T.W., 2015. Are Personas Really Usable? Communication Design Quarterly Review 3 (2), 20–26. https://doi.org/10.1145/2752853.2752856, 2015.

Howard, Tharon W., 2015. Are personas really usable? Commun. Des. Q. Rev 3 (2), 20–26. https://doi.org/10.1145/2752853.2752856. March 2015.

Hsu, Pei-Fang, Lu, Yu-Han, Chen, Shih-Chu, Pei-Yi Kuo, Patricia, 2024. Creating and validating predictive personas for target marketing. International Journal of Human-Computer Studies 181, 103147. https://doi.org/10.1016/j.ijhcs.2023.103147. January 2024.

Tiancheng Hu and Nigel Collier. 2024. Quantifying the Persona Effect in LLM Simulations. https://doi.org/10.48550/arXiv.2402.10811.

Bernard J. Jansen, Soon-gyo Jung, Joni Salminen, Jisun An, and Haewoon Kwak. 2018. Leveraging Online Social Media Data for Persona Profiling. March 15, 2018. Hamad bin Khalifa University Press (HBKU Press), ICTPP78. https://doi.org/10.5339/qfarc.2018.ICTPP78.

Jansen, Bernard J., Salminen, Joni, Jung, Soon-gyo, 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. https://doi.org/10.2478/dim-2020-0005, 2020.

Jansen, Bernard, Jung, Soon-gyo, Salminen, Joni, 2021. The Effect of Hyperparameter Selection on the Personification of Customer Population Data. International Journal of Electrical and Computer Engineering Research 1 (2). https://doi.org/10.53375/ijecer.2021.31. December 2021.

Jansen, Bernard, Salminen, Joni, Jung, Soon-gyo, Guan, Kathleen, 2021. Data-Driven Personas, 1st ed. Morgan & Claypool Publishers.

Jung, Soon-gyo, Salminen, Joni, Jansen, Bernard J., 2021. Persona Analytics: Implementing Mouse-tracking for an Interactive Persona System. In: Extended Abstracts of ACM Human Factors in Computing Systems - CHI EA '21. ACM. https://doi.org/10.1145/3411763.3451773, 2021Virtual conference.

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

Jung, Soon-gyo, Salminen, Joni, Kwak, Haewoon, An, Jisun, Jansen, Bernard J., 2018. Automatic Persona Generation (APG): a Rationale and Demonstration. In: CHIIR '18: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval. New Jersey, USA. ACM, pp. 321–324. https://doi.org/10.1145/3176349.3176893. March 2018.

Kaate, Ilkka, Salminen, Joni, Jung, Soon-Gyo, Santos, João M., Häyhänen, Essi, Xuan, Trang, Azem, Jinan Y., Jansen, Bernard, 2025. The 'fourth wall' and other usability issues in AI-generated personas: comparing chat-based and profile personas. Behaviour & Information Technology (February 2025). Retrieved April 15, 2025 from. https://www.tandfonline.com/doi/abs/10.1080/0144929X.2025.2469659.

Kaate, Ilkka, Salminen, Joni, Jung, Soon-Gyo, Xuan, Trang Thi Thu, Häyhänen, Essi, Azem, Jinan Y., Jansen, Bernard J., 2025. You Always Get an Answer": Analyzing Users' Interaction with AI-Generated Personas Given Unanswerable Questions and Risk of Hallucination. In: Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25). New York, NY, USA. Association for Computing Machinery, pp. 1624–1638. https://doi.org/10.1145/3708359.3712160. March 24, 2025.

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

Kim, Jang Hyun, Kim, Yunhwan, 2019. Instagram user characteristics and the color of their photos: Colorfulness, color diversity, and color harmony. Information Processing & Management 56 (4), 1494–1505. https://doi.org/10.1016/j.ipm.2018.10.018, 2019.

Königs, Peter, 2022. What is Techno-Optimism? Philos. Technol. 35 (3), 63. https://doi.org/10.1007/s13347-022-00555-x. July 2022.

Kwak, Haewoon, An, Jisun, Jansen, Bernard J., 2017. Automatic Generation of Personas Using YouTube Social Media Data. In: Proceedins of the Hawaii International Conference on System Sciences (HICSS-50). Waikoloa, Hawaii, USA, pp. 833–842. January 04, 2017.

Laufer, Benjamin, Jain, Sameer, Cooper, A.Feder, Kleinberg, Jon, Heidari, Hoda, 2022. Four Years of FAccT: a Reflexive, Mixed-Methods Analysis of Research Contributions, Shortcomings, and Future Prospects. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 401–426. https://doi.org/10.1145/3531146.3533107. June 21, 2022.

Christopher Lazik, Christopher Katins, Charlotte Kauter, Jonas Jakob, Caroline Jay, Lars Grunske, and Thomas Kosch. 2025. The Impostor is Among Us: Can Large Language Models Capture the Complexity of Human Personas? https://doi.org/10.48550/arXiv.2501.04543.

Ang Li, Haozhe Chen, Hongseok Namkoong, and Tianyi Peng. 2025. LLM Generated Persona is a Promise with a Catch. https://doi.org/10.48550/arXiv.2503.16527.

Q. Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. April 01, 2023. . Retrieved February 20, 2025 from https://www.microsoft.com/en-us/research/publication/designerly-understanding-information-needs-for-model-transparency-to-support-design-ideation-for-ai-powered-user-experience/.

Liu, Mengjuan, Liu, Chenyang, Yang, Yunfan, Liu, Jiang, Jing, Mohan, 2023. Promoting Open-Domain Dialogue Generation Through Learning Pattern Information Between Contexts and Responses. In: Natural Language Processing and Chinese Computing: 12th National CCF Conference, NLPCC 2023, Foshan, China, October 12–15, 2023, Proceedings, Part II. Berlin, Heidelberg. Springer-Verlag, pp. 351–362. https://doi.org/10.1007/978-3-031-44696-2_28, 2023.

Yi Liu, Gelei Deng, Yuekang Li, Kailong Wang, Zihao Wang, Xiaofeng Wang, Tianwei Zhang, Yepang Liu, Haoyu Wang, Yan Zheng, and Yang Liu. 2024. Prompt Injection attack against LLM-integrated Applications. https://doi.org/10.48550/arXiv.2306.05499.

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

Lum, Kristian, Zhang, Yunfeng, Bower, Amanda, 2022. De-biasing "bias" measurement. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). New York, NY, USA. Association for Computing Machinery, pp. 379–389. https://doi.org/10.1145/3531146.3533105. June 20, 2022.

Ahmed Magooda, Alec Helyar, Kyle Jackson, David Sullivan, Chad Atalla, Emily Sheng, Dan Vann (HE/HIM), Richard Edgar, Hamid Palangi, Roman Lutz, Hongliang Kong, Vincent Yun, Eslam Kamal, Federico Zarfati, Hanna Wallach, Sarah Bird, and Mei Chen. 2023. A Framework for Automated Measurement of Responsible AI Harms in Generative AI Applications. abs/2310.17750, (October 2023). Retrieved February 20, 2025 from https://www.microsoft.com/en-us/research/publication/a-framework-for-automated-measurement-of-responsible-ai-harms-in-generative-ai-applications-2/.

Marsden, Nicola, Haag, Maren, 2016. Stereotypes and politics: reflections on personas. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4017–4031, 2016.

Matthews, Tara, Judge, Tejinder, Whittaker, Steve, 2012. How do designers and user experience professionals actually perceive and use personas?. In: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12. ACM Press, p. 1219. https://doi.org/10.1145/2207676.2208573, 2012.

McGinn, Jennifer (Jen), Kotamraju, Nalini, 2008. Data-driven persona development. In: Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08. Florence, Italy. ACM Press, p. 1521. https://doi.org/10.1145/1357054.1357292, 2008.

Sarah Mercer, Samuel Spillard, and Daniel P. Martin. 2025. Brief analysis of DeepSeek R1 and its implications for Generative AI. https://doi.org/10.48550/arXiv.2502.02523.

Mesgari, Mostafa, Okoli, Chitu, Ortiz de Guinea, Ana, 2015. Affordance-based User Personas : a mixed-method Approach to Persona Development. In: AMCIS 2015 Proceedings. June 26, 2015. . Retrieved from. https://aisel.aisnet.org/amcis2015/HCI/GeneralPresentations/1.

Midjourney. Midjourney. Retrieved November 11, 2024 from https://www.midjourney.com/website.

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

Mitrokhov, Konstantin, 2024. Between world models and model worlds: on generality, agency, and worlding in machine learning. AI & Soc. https://doi.org/10.1007/s00146-024-02086-9. October 2024.

Moreno, Felipe Romero, 2024. Generative AI and deepfakes: a human rights approach to tackling harmful content. International Review of Law, Computers & Technology 38 (3), 297–326. https://doi.org/10.1080/13600869.2024.2324540. September 2024.

Mulder, Steve, Yaar, Ziv, 2006. The User is Always Right: A Practical Guide to Creating and Using Personas For the Web. New Riders.

Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. (July 2023).

Pranav Narayanan Venkit, Jiayi Li, Yingfan Zhou, Sarah Rajtmajer, and Shomir Wilson. 2025. A Tale of Two Identities: An Ethical Audit of Human and AI-Crafted Personas. https://doi.org/10.48550/arXiv.2505.07850.

Nielsen, Lene, 2018. Design personas - New ways, new contexts. Persona Studies 4 (2), 1–4. https://doi.org/10.21153/psj2018vol4no2art799 (November 2018).

Nielsen, Lene, 2019. Personas - User Focused Design, 2nd ed. Springer, New York, NY, USA. 2019 edition ed.

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

OpenAI. 2023. GPT-4 Technical Report. OpenAI. Retrieved from https://openai.com/research/gpt-4.

Papagiannidis, Emmanouil, Mikalef, Patrick, Conboy, Kieran, 2025. Responsible artificial intelligence governance: a review and research framework. The Journal of Strategic Information Systems 34 (2), 101885. https://doi.org/10.1016/j.jsis.2024.101885. June 2025.

David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. https://doi.org/10.48550/arXiv.2104.10350.

Dan Perkel, Angela Kochoska, Will Notini, and Beth Holzer. 2025. The case against AI-generated users. IDEO. Retrieved August 18, 2025 from https://www.ideo.com/journal/the-case-against-ai-generated-users.

Portigal, Steve, 2023. Interviewing Users: How to Uncover Compelling Insights. Rosenfeld Media.

Prpa, Mirjana, Troiano, Giovanni Maria, Wood, Matthew, Coady, Yvonne, 2024. Challenges and Opportunities of LLM-Based Synthetic Personae and Data in HCI. In: Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24). New York, NY, USA. Association for Computing Machinery, pp. 1–5. https://doi.org/10.1145/3613905.3636293. May 11, 2024.

Pruitt, John, Adlin, Tamara, 2006. The Persona Lifecycle: Keeping People in Mind Throughout Product Design, 1 edition ed. Morgan Kaufmann, Boston.

Rapp, Amon, Lodovico, Chiara Di, Torrielli, Federico, Caro, Luigi Di, 2025. How do people experience the images created by generative artificial intelligence? An exploration of people's perceptions, appraisals, and emotions related to a Gen-AI text-to-image model and its creations. International Journal of Human-Computer Studies 193, 103375. https://doi.org/10.1016/j.ijhcs.2024.103375. January 2025.

Ronkko, K., 2005. An Empirical Study Demonstrating How Different Design Constraints, Project Organization and Contexts Limited the Utility of Personas. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Big Island, HI, USA. IEEE. https://doi.org/10.1109/HICSS.2005.85, 2005220a–220a.

Rönkkö, Kari, Hellman, Mats, Kilander, Britta, Dittrich, Yvonne, 2004. Personas is Not Applicable: Local Remedies Interpreted in a Wider Context. In: Proceedings of the

Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1 (PDC 04). Toronto, Ontario, Canada. ACM, pp. 112–120. https://doi.org/10.1145/1011870.1011884, 2004.

Joni Salminen, Danial Amin, and Bernard J Jansen. Using AI for User Representation: An Analysis of 83 Persona Prompts.

Salminen, Joni, Chhirang, Kamal, Jung, Soon-Gyo, Thirumuruganathan, Saravanan, Guan, Kathleen W., Jansen, Bernard J., 2022. Big Data, Small Personas: How Algorithms Shape the Demographic Representation of Data-Driven User Segments. Big Data. https://doi.org/10.1089/big.2021.0177 (August 2022).

Salminen, Joni, Froneman, Willemien, Jung, Soon-gyo, Chowdhury, Shammur, Jansen, Bernard J., 2020. The Ethics of Data-Driven Personas. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '20). Honolulu, HI, USA. Association for Computing Machinery, pp. 1–9. https://doi.org/10.1145/3334480.3382790. April 25, 2020.

Salminen, Joni, Guan, Kathleen Wenyun, Jung, Soon-Gyo, Jansen, Bernard, 2022. Use Cases for Design Personas: a Systematic Review and New Frontiers. In: CHI Conference on Human Factors in Computing Systems (CHI '22). New York, NY, USA. Association for Computing Machinery, pp. 1–21. https://doi.org/10.1145/3491102.3517589. April 29, 2022.

Salminen, Joni, Guan, Kathleen, Jung, Soon-gyo, Absar Chowdhury, Shammur, Jansen, Bernard J., 2020. A Literature Review of Quantitative Persona Creation. In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, Hawaii, USA. ACM, pp. 1–14. https://doi.org/10.1145/3313831.3376502. April 25, 2020.

Salminen, Joni, Guan, Kathleen, Jung, Soon-Gyo, Jansen, Bernard J., 2021. A Survey of 15 Years of Data-Driven Persona Development. International Journal of Human–Computer Interaction 37 (18), 1685–1708. https://doi.org/10.1080/10447318.2021.1908670. November 2021.

Salminen, Joni, Jansen, Bernard J., An, Jisun, Jung, Soon-gyo, Nielsen, Lene, Kwak, Haewoon, 2018. Fixation and Confusion – Investigating Eye-tracking Participants' Exposure to Information in Personas. In: Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR 2018). New Jersey, USA. ACM, pp. 110–119. https://doi.org/10.1145/3176349.3176391, 2018.

Salminen, Joni, Jansen, Bernard J., An, Jisun, Kwak, Haewoon, Jung, Soon-gyo, 2018. Are personas done? Evaluating their usefulness in the age of digital analytics. Persona Studies 4 (2), 47–65. https://doi.org/10.21153/psj2018vol4no2art737. November 2018.

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

Salminen, Joni, Jung, Soon-gyo, Jansen, Bernard J., 2021. Are Data-Driven Personas Considered Harmful? : Diversifying User Understandings With More Than Algorithms. Persona Studies 7 (1), 48–63. https://doi.org/10.21153/psj2021vol7no1art1236. December 2021.

Salminen, Joni, Liu, Chang, Pian, Wenjing, Chi, Jianxing, Häyhänen, Essi, Jansen, Bernard J, 2024. Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions. In: Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). New York, NY, USA. Association for Computing Machinery, pp. 1–20. https://doi.org/10.1145/3613904.3642036. May 11, 2024.

Salminen, Joni, Santos, Joao M., Jung, Soon-gyo, Eslami, Motahhare, Jansen, Bernard J., 2019. Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. International Journal of Human–Computer Interaction 1–13. https://doi.org/10.1080/10447318.2019.1688946, 0, 0November 2019.

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

Sattele, Vanessa, Ortiz, Juan Carlos, 2024. Generating user personas with AI: Reflecting on its implications for design. DRS Biennial Conference Series (June 2024). Retrieved from. https://dl.designresearchsociety.org/drs-conference-papers/drs2024/researchpapers/319.

Schneier, Bruce, 2024. LLMs' Data-Control Path Insecurity. Commun. ACM 67 (9), 31–32. https://doi.org/10.1145/3660674 (September 2024).

Schön, Donald A., 1992. The Theory of Inquiry: dewey's Legacy to Education. Curriculum Inquiry 22 (2), 119–139. https://doi.org/10.2307/1180029, 1992.

Schuller, Andreas, Janssen, Doris, Blumenröther, Julian, Probst, Theresa Maria, Schmidt, Michael, Kumar, Chandan, 2024. Generating personas using LLMs and assessing their viability. In: Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24). New York, NY, USA. Association for Computing Machinery, pp. 1–7. https://doi.org/10.1145/3613905.3650860. May 11, 2024.

Shams, Rifat Ara, Zowghi, Didar, Bano, Muneera, 2023. AI and the quest for diversity and inclusion: a systematic literature review. AI Ethics. https://doi.org/10.1007/s43681-023-00362-w. November 2023.

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

Shneiderman, Ben, 2022. Human-Centered AI. Oxford University Press, Oxford, New York.

Siegel, David A., 2010. The Mystique of Numbers: Belief in Quantitative Approaches to Segmentation and Persona Development. In: CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10). New York, NY, USA. ACM, pp. 4721–4732. https://doi.org/10.1145/1753846.1754221, 2010.

Smrke, Urška, Rehberger, Ana, Plohl, Nejc, Mlakar, Izidor, 2025. Exploring the Feasibility of Generative AI in Persona Research: a Comparative Analysis of Large Language Model-Generated and Human-Crafted Personas in Obesity Research. Applied Sciences 15, 1937. https://doi.org/10.3390/app15041937 (February 2025).

Turner, Phil, Turner, Susan, 2011. Is stereotyping inevitable when designing with personas? Design studies 32 (1), 30–44. https://doi.org/10.1016/j.destud.2010.06.002, 2011.

Nick De Voil. 2010. Personas considered harmful. De Voil Consulting, London, UK. Retrieved November 4, 2019 from http://www.devoil.com/papers/PersonasConsideredHarmful.pdf.

Wan, Yixin, Pu, George, Sun, Jiao, Garimella, Aparna, Chang, Kai-Wei, Peng, Nanyun, 2023. Kelly is a Warm Person, Joseph is a Role Model": Gender Biases in LLM-Generated Reference Letters. In: Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore. Association for Computational Linguistics, pp. 3730–3748. https://doi.org/10.18653/v1/2023.findings-emnlp.243. December 2023.

Wolfram, Stephen, 2023. What Is ChatGPT Doing:... and Why Does It Work? Wolfram Media.

Ziwei Xu, Sanjay Jain, and Mohan Kankanhalli. 2025. Hallucination is Inevitable: An Innate Limitation of Large Language Models. https://doi.org/10.48550/arXiv.2401.11817.

Zhang, Xiang, Brown, Hans-Frederick, Shankar, Anil, 2016. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). San Jose, California, USA. ACM, pp. 5350–5359, 2016.

Zhang, Xishuo, Liu, Lin, Wang, Yi, Liu, Xiao, Wang, Hailong, Arora, Chetan, Liu, Haichao, Wang, Weijia, Hoang, Thuong, 2024. Auto-Generated Personas: Enhancing User-centered Design Practices among University Students. In: Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24). New York, NY, USA. Association for Computing Machinery, pp. 1–7. https://doi.org/10.1145/3613905.3651043. May 11, 2024.

Fufangchen Zhao, Guoqiang Jin, Jiaheng Huang, Rui Zhao, and Fei Tan. 2024. Consistency Matters: Explore LLMs Consistency From a Black-Box Perspective. https://doi.org/10.48550/arXiv.2402.17411.

Mi Zhou, Vibhanshu Abhishek, Timothy Derdenger, Jaymo Kim, and Kannan Srinivasan. 2024. Bias in Generative AI. Retrieved October 27, 2024 from http://arxiv.org/abs/2403.02726.