Introducing Persona Ecosystem Playground (PEP): Generating Behaviorally Grounded Personas from Think-Aloud Sessions

Danial Amin
University of Vaasa
Vaasa, Finland
danialam@uwasa.fi

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

Bernard J. Jansen
Qatar Computing Research Institute
Hamad Bin Khalifa University
Doha, Qatar
bjansen@hbku.edu.qa

Abstract

Persona generation methods rely primarily on structured data like surveys and analytics, leaving rich qualitative sources such as think-aloud transcripts and interview records underutilized. We present Persona Ecosystem Playground (PEP), a system that uses LLMs with retrieval-augmented generation (RAG) to generate personas from unstructured qualitative data while maintaining systematic traceability to source evidence. Demonstrated on think-aloud transcripts from 67 participants in an educational chatbot study, PEP generates five personas with 77% of attributes directly supported by transcript evidence (cosine similarity ≥ 0.8), including 96% direct support for behavioral attributes and 90% for goal attributes. The integration of RAG-based grounding, cosine similarity validation, and Rao's Quadratic Entropy diversity measurement into a single automated pipeline distinguishes PEP from prior persona systems that lack systematic traceability. This work establishes a validated persona generation foundation for multi-group stakeholder modeling, where distinct personas representing different interdependent groups can be generated from heterogeneous qualitative sources.

CCS Concepts

• **Human-centered computing → HCI**.

Keywords

Personas, Large Language Models, Think-Aloud Protocols, RAG, Unstructured Data

ACM Reference Format:
Danial Amin, Joni Salminen, and Bernard J. Jansen. 2026. Introducing Persona Ecosystem Playground (PEP): Generating Behaviorally Grounded Personas from Think-Aloud Sessions. In *ACM Conversational User Interfaces 2026 (CUI '26), July 21–24, 2026, Bremen, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3816046.3816269

1 Introduction

Personas are fictitious but realistic user representations that help decision makers understand users by embodying actual user segments' goals, behaviors, and needs [10, 22]. They are most useful when the decisions at hand affect multiple interdependent groups, yet existing persona methods almost exclusively model a single user population [16]. In education, students, instructors, parents, and administrators form a system in which each group's behaviors influence the others; in healthcare, patients, clinicians, and administrators interact similarly. Wicked problems such as war, sustainability, and diplomacy require understanding how these groups relate, not just how one of them thinks [2, 20]. A persona method that only captures one group's perspective is, by design, an incomplete tool for these contexts.

Existing persona generation has made progress on data grounding, but has not addressed this limitation. Most data-driven approaches rely on structured data— surveys [18], web analytics [25], or social media metrics [19] — which describe users through predefined categories rather than capturing their verbalized experiences and thought processes [29]. Unstructured qualitative sources — interview records [22], observational notes [24], and think-aloud transcripts [14] — remain largely untapped because manual coding is labor-intensive [7]. More critically, even the systems that do process unstructured data [8, 11, 31] focus on single populations, and none combine RAG-based generation with systematic traceability and quantitative diversity measurement. The result is a gap between what qualitative data can reveal about multiple stakeholder groups and what persona systems can actually produce.

We address this gap with the *Persona Ecosystem Playground (PEP)* [4], an interactive persona system that uses LLMs with retrieval-augmented generation (RAG) to generate personas from unstructured qualitative data while maintaining traceability to source evidence (see Figure 1). PEP is designed toward multi-group persona generation, producing distinct grounded personas for different stakeholder groups from heterogeneous data sources. This paper validates PEP's single-group persona generation stage using think-aloud transcripts from 67 participants in an educational chatbot user study, while multi-group persona generation and validation remain topics for future work. Data-driven persona development is an applied area that adopts methods from adjacent fields to improve persona quality and scalability [25]. PEP follows this tradition. The contribution is the integration of RAG grounding, cosine similarity (CS) validation, and Rao's Quadratic Entropy (RQE) diversity measure into a single automated pipeline, which researchers and practitioners can use directly.

Our evaluation shows that 77% of persona attributes achieve direct support in source transcripts (CS ≥ 0.8), with 96% direct support for behavioral attributes and 90% for goal attributes, the attributes most critical for authentic user representation. These results demonstrate that PEP's pipeline reduces hallucination while capturing interaction patterns from the transcripts. Our contributions are: (1) a method for generating personas from unstructured

CUI '26, Bremen, Germany

ACM ISBN 979-8-4007-2741-2/26/07
https://doi.org/10.1145/3816046.3816269

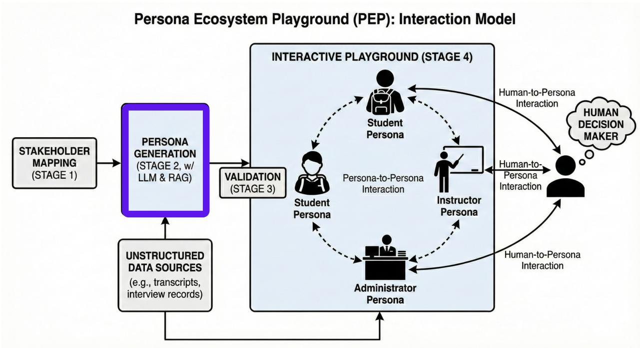

Figure 1: Conceptual framework of Persona Ecosystem Playground showing the four stages of (1) Stakeholder mapping, (2) Persona generation (which is the focus of this article), (3) Persona validation by the real users, and (4) Interactive playground enabling persona-to-persona and human-to-persona interactions.

qualitative data using LLMs with RAG, maintaining systematic traceability between persona attributes and source evidence; (2) an automated evaluation pipeline combining CS-based attribute validation and RQE-based diversity measurement, with concrete thresholds for practitioners; and (3) a validated generation foundation for multi-group persona modeling, with the multi-stakeholder interaction stage as the next phase of PEP's development.

2 Related Work

Using LLM technology has expanded automated persona generation possibilities [1, 28, 30], but no existing system achieves what PEP aims for (see Table 1). De Paoli [11] showed that LLMs can analyze interview data to compose persona narratives, but did not provide systematic traceability by mapping the persona attributes back to specific source evidence. *PersonaGen* [33] uses knowledge graphs and GPT-4 to generate personas from user feedback, but relies on structured data and does not address multi-stakeholder ecosystems. Interactive systems like *Survey2Persona* [26] enable conversational engagement with personas, but focus on single-population modeling from structured survey data. *Proxona* [8] processes unstructured audience comments with RAG-based interaction, but targets single-population (audience) representation for content creators. Sun et al. [31] developed *Persona-L* with interactive interfaces using RAG for users with complex needs, but from structured data without ecosystem modeling. Shin et al. [30] demonstrated effective human-AI collaborative workflows, but for structured survey data representing single populations.

None of these approaches addresses the requirements for persona ecosystem modeling: (1) generating validated personas from unstructured data while maintaining systematic traceability, and (2) modeling interactions with multiple interdependent stakeholder groups. This combination is, to our knowledge, the first to address both requirements in one pipeline.

3 Evaluation Study

3.1 Study Design

We tested PEP by conducting a user study, recording participant think-aloud utterances, and then feeding them to PEP for persona generation. The user study tested a system called Cipherbot [17], which is a Socratic educational AI chatbot. The user study participants' task was to learn about endangered species by using Cipherbot.

We recruited 67 participants (38 male, 29 female, mean age 36.7, range 18-70). Education status referred to currently enrolled or recently completed degrees and included 30 doctoral, 18 master's, 15 bachelor's, and 4 high school students. Nationalities included Pakistani, Jordanian, Indian, Lebanese, and Syrian. The study used a 2×2 within-subjects Latin square design crossing modality (text vs. audio) with species (butterfly vs. frog). Each participant completed two 50-60 minute sessions.

The learning task asked participants to learn about an endangered species through chatbot interactions using two different learning modalities (text and audio), and then to compose a social media post asking their audience to support conservation efforts. Participants verbalized their thoughts continuously following the standard think-aloud protocol [14]. When a participant fell silent for some time, the moderator prompted neutrally (e.g., "What are you thinking?"). The full protocol is in the supplementary material. Sessions were audio-recorded. Auto-transcription via Microsoft Word produced roughly 311k words, including moderator utterances, with an average of 2,700 words per session. After removing moderator utterances, the participant corpus contained approximately 257k words (per-participant range: 2,100–6,400; median: 3,700), which PEP used for persona generation. The five generated personas represent learner archetypes from this participant pool and are intended for chatbot designers' to decide among two learning modalities (text and audio) as to which is more suitable for different learners. We encourage the readers to find the full details of the user study and the results in Xuan et al [32].

3.2 Persona Generation and Validation

PEP transforms unstructured experiential data into validated personas through four stages (see Figure 2): preprocessing of transcripts into a retrievable corpus, persona set generation with diversity measurement, attribute expansion grounded in retrieved evidence, and validation through reverse-query CS. The RAG layer combines a vector database for embedded chunks, a search-and-ranking layer for retrieval, and an LLM for generation. Based on preliminary testing, we use `Pinecone` [23], `Cohere` [9], and `Gemini` [15], respectively. The following paragraphs describe how these technologies are used. The complete prompt templates and validation criteria are available in the online supplementary material[1].

Data Preprocessing. We applied recursive chunking to the think-aloud transcripts, which segments text while maintaining contextual continuity from previous chunks, that is critical for preserving the connected nature of users' verbal protocols [14]. Chunks

[1]https://osf.io/n2gzt/overview?view_only=d7188964739e40a5b03ffa38f2bacdf2

Table 1: Positioning PEP against existing LLM-generated persona systems and approaches.

Approach	Unstruct. Data	Struct. Data	Ecosystem Model	Persona Interact.	RAG	Traceability
System approaches						
Persona-L [31]	✓	✗	✗	✓	✓	✗
PersonaCraft [18]	✗	✓	✗	✓	✗	✓
PersonaFlow [21]	✗	✓	✗	✓	N/A	✓
PersonaGen [33]	✗	✓	✗	✗	✗	✗
Proxona [8]	✓	✗	✗	✓	✓	N/A
Survey2Persona [26]	✗	✓	✗	✓	✓	N/A
PEP	✓	✓	✓	✓	✓	✓
Non-system approaches						
De Paoli [11, 12]	✓	✗	✗	✗	✗	✗
Schuller et al. [28]	✗	✓	✗	✗	✗	✗
Shin et al. [30]	✗	✓	✗	✗	✗	✗

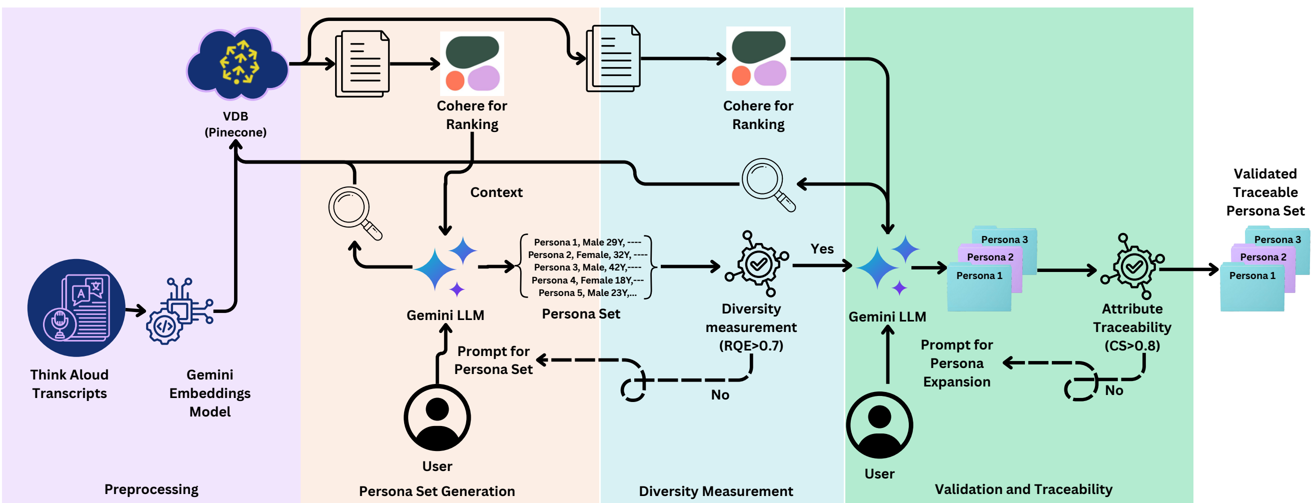

Figure 2: Persona generation process from unstructured data using PEP. From left to right, the process starts by preprocessing of the transcripts, followed by generation of persona set, which is checked for diversity using Rao's Quadratic Entropy (RQE), expanded by the LLM to include attributes for each persona, that is finally validated by reverse query using Cosine Similarity (CS).

were created at natural task boundaries (e.g., survey section completions, task phase transitions) to maintain semantic coherence. This yielded 2,571 chunks averaging 100 words each, a chunk size based on experience. Each chunk retained metadata tags for key study variables: modality (text vs voice interaction with Cipherbot), participant ID, species (butterfly vs frog—the two knowledge domains participants learned), and task phase. We generated embeddings using Gemini and populated Pinecone, configured with a cosine distance metric.

Persona Set Generation. Given that modality was a key variable in our study design examining how interaction mode affects learning behaviors, we generated a persona set for that. We created a set of five personas, aligning with persona research indicating that 4-6 personas optimize coverage while maintaining manageability [24, 25]. While PEP's framework supports generating fewer or more personas based on data complexity and stakeholder needs, five provided sufficient coverage of observed interaction patterns in our dataset.

Persona Attributes and Structure. Each persona contains roughly ten attributes following established persona templates [10, 22, 24]. Goals, interaction behaviors, frustrations, technology usage patterns, and supporting quotes are derived from the transcripts, with each attribute linked to specific chunks through RAG-retrieved evidence. Demographic attributes (gender, age, education, profession) are assigned through a fairness rule that requires the persona set to reflect the demographic composition of the recruited participant pool. We adopt this approach because demographic profiles are a standard persona component, and a fairness-driven assignment grounded in the actual sample is more principled than treating demographic cues in unstructured text as evidence. The JSON structure preserves traceability for the data-grounded attributes and records the fairness-rule basis for the assigned demographics.

Attribute Extraction. For each behavioral dimension surfaced from the transcripts, the LLM drafted skeletal profiles using the top-20 most semantically similar chunks reranked by Cohere. Each generated attribute was paired with its supporting chunk IDs. When chunks contained conflicting evidence for the same dimension, the LLM was prompted to attribute the patterns to different personas in the set rather than merge them, using the conflict as a diversification signal. Attributes that could not be linked to a supporting chunk above the CS threshold were removed during validation.

Diversity Measurement and Iteration. We measured persona set diversity using Rao's Quadratic Entropy (RQE) [3], which quantifies the overall differentiation across all personas in the set on a scale from 0 (identical personas) to 1 (maximally diverse). Based on validation studies [3], RQE $\geq$ 0.75 indicates good diversity for a persona set, RQE = 0.60-0.74 indicates moderate diversity, and $<$ 0.60 indicates insufficient differentiation. Our initial persona set achieved RQE = 0.48, indicating substantial overlap among the five personas. Following systematic prompt refinement using the Prompt Lab framework [6], we iteratively adjusted generation prompts to emphasize distinct interaction patterns that emerged from the data: e.g., re-reading behaviors, frustration expression styles, and question engagement patterns. After three iterations, the final persona set achieved moderate to good diversity (RQE = 0.77). As PEP is an automated system, different generation runs may produce persona sets with varying RQE values; we established RQE = 0.77 as our acceptance threshold, applying regeneration for sets below this value.

Validation and Traceability. Each persona attribute was validated through reverse RAG queries mapping claims to source evidence, with cosine similarity values (CS) $\geq$ 0.8 corroborating support from transcript data. Cosine similarity is a standard metric used for measuring semantic similarity between text entries [13]. Persona attributes with CS $<$ 0.8 were flagged for expert review or removal, ensuring that all personas maintained systematic traceability to source data. This validation yielded CS $\geq$ 70% across all personas, meeting the grounding threshold.

Figure 3 shows one of the five generated personas as a concrete example. We validated each persona's data-grounded attributes by mapping them back to transcript evidence through reverse-lookup queries calculating CS. Table 2 presents validation examples across attribute types.

Of 48 data-grounded attributes across five personas (excluding the 10 demographic attributes assigned via the fairness rule), 37 (77%) achieved direct support with CS values above 0.80. A further 8 (17%) showed indirect support (CS = 0.6–0.8), and 3 (6%) were removed automatically during iteration for lacking sufficient support. Behavioral attributes achieved 24/25 (96%) direct support and goal attributes 9/10 (90%), which are the most important attribute types for a persona. **The validation shows that PEP maintains traceability between persona attributes and source evidence, reducing hallucination that would otherwise undermine user representation.**

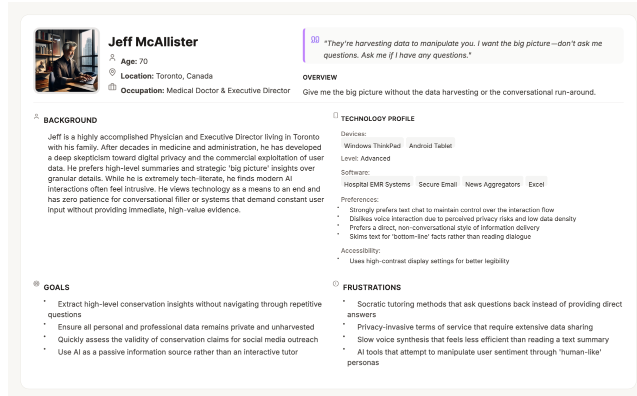

Figure 3: Example persona generated by PEP. Behavioral, goal, and frustration attributes are validated via reverse-query CS against transcript chunks. Demographic attributes are assigned via the fairness rule (Section 3.2). The full set of five personas with all supporting chunks is available in the supplementary material.

4 Discussion

For our educational chatbot study, PEP generated personas capturing distinct learning interaction patterns across 67 participants who represented diverse backgrounds and roles — spanning professors, academicians, doctoral researchers, master's students, bachelor's students, and high school students from multiple nationalities. This diversity matters as instead of drawing from a homogeneous population, PEP produced grounded personas from a participant pool whose members occupied different positions relative to the learning task, suggesting that the pipeline can accommodate varied user profiles within a single unstructured data source. The traceability results are worth noting: 96% direct support for behavioral attributes and 90% for goal attributes demonstrates that RAG-based generation can effectively constrain the LLM to source evidence, reducing the hallucination risk that makes unconstrained LLM persona generation problematic [5, 27]. Cosine similarity should be read as a proxy, as it only captures semantic proximity between an attribute statement and a retrieved chunk, which is informative. However, it does not measure factual correctness or alignment with participant intent. A principled next step would be a sensitivity analysis in which human raters score attribute-evidence matches sampled from low, mid, and high CS bands, with the expectation that higher bands receive higher goodness ratings. We report this as a direction for future work rather than a check completed here.

The current study demonstrates PEP on a single dataset from one educational context, and generalizability to noisier, culturally varied, or multilingual corpora remains to be tested. The evaluation also reports internal pipeline metrics only. Human assessment of persona quality, baseline comparison against alternative persona generation systems, and demonstration of downstream design utility are not included, and the short paper format does not accommodate them. However they are necessary for the overall validation of the system and remain work for future considerations. Moreover, choices in embedding and prompting may further introduce systematic biases that CS-based traceability fails to detect. Together, these points lead to a clear path for extending the work toward

Table 2: Validation examples for persona attributes. Behavioral and goal attributes are validated via reverse-query CS against transcript chunks. Demographic attributes are assigned via the fairness rule (Section 3.2) and are therefore not CS-validated. CS = cosine similarity.

Attribute Type	Example Attribute	Supporting Evidence from Transcripts	CS
Behavioral Attributes (Direct Support)			
Jameela	Re-reading questions multiple times	"I'm reading this question again to make sure I understand what they're asking"	0.87
Jeff	Rejects Socratic methods	"Don't ask me questions. Ask me if I have any questions"	0.92
Maya	Prompt anxiety	"I wasn't sure on what to write" + requests for tutorials	0.84
Bashir	Scans for key data points	Verbalizations about extracting summaries quickly, visual structure importance	0.85
Goal Attributes (Direct Support)			
Jameela	Specialist-level depth	Chunks discussing technical terminology preferences, comparing perspectives	0.81
Hana	Speed and minimal friction	"I need instant responses", frustration with slow voice synthesis	0.83
Demographic Attributes (Fairness-Rule Assignment, not CS-validated)			
Hana	Age 16, high school student	Assigned to span younger end of participant age range	N/A
Jeff	Profession: medical doctor	Assigned to span professional roles in participant pool	N/A
Personality Descriptors (Removed)			
Initial generation	"Meticulous" descriptor	No direct transcript support; replaced with behavioral description	0.42
Initial generation	"Cautious" descriptor	No direct transcript support; replaced with "reviews multiple times"	0.38

a full paper that addresses external validation, bias auditing, and replication in other domains.

The motivating ambition, however, is multi-group persona generation as producing distinct, grounded personas for multiple interdependent stakeholder groups from heterogeneous data sources. Consider healthcare policy design, where patients, clinicians, administrators, and advocacy groups each hold perspectives that shape outcomes and where decisions made without understanding all groups risk unintended consequences [20]. PEP's pipeline could generate patient personas from interview transcripts, clinician personas from observational notes, and administrator personas from meeting recordings, giving decision makers a grounded view of the full stakeholder picture. Similar applications apply to inclusive UX research, where underrepresented communities are routinely absent from structured analytics, and to policy co-design, where citizen voices exist primarily in qualitative consultation records.

The next phase of PEP's development is the interactive playground, which we term PEP-Sim. In PEP-Sim, validated personas serve as conversational agents, enabling decision-makers to interact with multiple stakeholder personas simultaneously and observe where perspectives converge or conflict. The planned architecture builds on agent-based modeling (LangChain, LangGraph), with personas grounded in their source-validated attributes to ensure that simulated interactions reflect documented behavioral patterns rather than model defaults. The evaluation protocol will recruit HCI researchers and domain practitioners to interact with PEP-Sim personas and rate realism, usefulness, and coverage against their domain knowledge. A comparative condition will assess whether multi-stakeholder persona interactions yield insights that sequential single-population analyses miss, with a baseline comparison against existing LLM persona tools on traceability and diversity metrics.

5 Conclusion

This paper shows that PEP generates validated personas from unstructured qualitative data while maintaining traceability to source evidence. By combining LLMs with RAG, cosine similarity validation, and RQE-based diversity measurement into a single automated pipeline, PEP addresses a gap, since rich experiential data from think-aloud protocols, interviews, and observational records has been largely unavailable to data-driven persona methods. The validation results with 77% direct attribute support overall, 96% for behavioral attributes, show that the pipeline grounds persona generation in what participants actually said rather than what LLMs tend to assume. The concrete CS and RQE thresholds we established give practitioners actionable benchmarks for assessing persona quality from unstructured sources. Multi-group persona generation, with validated stakeholder personas interacting in PEP-Sim, is the next step.

Acknowledgments

The think-aloud data analyzed in this paper were originally collected as part of the Cipherbot user study reported in [32]. We thank Trang Xuan, who led that study as principal investigator, together with Ilkka Kaate, Farhan Ahmed, Rajat Patil, Soon-Gyo Jung, and Jinan Y. Azem, for their work in designing and conducting the sessions. We are also grateful to the 67 participants whose verbalizations form the empirical basis of the personas generated and validated here.

References

[1] Danial Amin, Joni Salminen, Farhan Ahmed, M.H. Sonja Tervola, Sankalp Sethi, and Bernard J. Jansen. 2026. Creating and Evaluating Personas Using Generative AI: A Scoping Review of 81 Articles. In *CHI Conference on Human Factors in Computing Systems (CHI'26)*. doi:10.1145/3772318.3790608

[2] Danial Amin, Joni Salminen, Farhan Ahmed, Emillia Zemskova, and Bernard J. Jansen. 2026. How Can Tensions in Sustainability Transitions Be Made Visible?: Challenging the Actor-Centered Ecosystem Approach through Persona-Based Modeling. In *Ecosystems and Networks for Sustainable Business*. Palgrave Macmillan. In review.

[3] Danial Amin, Joni Salminen, and Bernard Jansen. 2026. Proposing a Quantitative Framework for Persona Set Diversity Measurement as a Pathway to Inclusive User Representation. *IEEE Access* (2026). In review.

[4] Danial Amin, Joni Salminen, Bernard Jansen, Ilkka Kaate, and Waleed Akhtar. 2026. Large language models (LLMs) in human-computer interaction: Using LLM-generated personas to model everything from minority views to entire

ecosystems. In *Artificial Intelligence & Large Language Models: A Scientific Perspective*. CRC Press. doi:10.1201/9781003492252-10

[5] Danial Amin, Joni Salminen, Bernard J. Jansen, Joongi Shin, and Dong Hun Kim. 2025. Generative AI Personas Considered Harmful? Examining the Risks of Algorithmic User Representations in HCI Design. *International Journal of Human-Computer Studies* (2025). doi:10.1016/j.ijhcs.2025.103657 Published.

[6] Danial Amin, Joni Salminen, Ilkka Kaate, and Bernard J. Jansen. 2026. "Designing Prompts to Design Personas": A Review of Prompt Design for AI-Generated Personas from Techno-Fairness Perspective. (2026). In review.

[7] Chris Chapman and Russell P. Milham. 2006. The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. 634–636. doi:10.1177/154193120605000503

[8] Yoonseo Choi, Eun Jeong Kang, Seulgi Choi, Min Kyung Lee, and Juho Kim. 2025. Proxona: Supporting creators' sensemaking and ideation with LLM-powered audience personas. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–32. doi:10.1145/3706598.3714034

[9] Cohere. 2024. Cohere Rerank: Semantic Search Reranking. https://cohere.com/rerank.

[10] Alan Cooper. 1999. *The Inmates Are Running the Asylum*. Sams Publishing.

[11] Stefano De Paoli. 2023. Writing user personas with large language models: Testing phase 6 of a thematic analysis of semi-structured interviews. *arXiv preprint arXiv:2305.18099* (2023).

[12] Stefano De Paoli. 2026. User personas, ideation and large language models: A post-hoc study. *International Journal of Human-Computer Studies* 208 (Jan. 2026), 103690. doi:10.1016/j.ijhcs.2025.103690

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186. doi:10.18653/v1/N19-1423

[14] K. Anders Ericsson and Herbert A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. MIT Press.

[15] Google DeepMind. 2024. Gemini: Google's Most Capable AI Model. https://deepmind.google/technologies/gemini/.

[16] Jonathan Grudin and John Pruitt. 2006. Personas, participatory design and product development: An infrastructure for engagement. In *Proceedings of the Participatory Design Conference*, Vol. 11. 144–161.

[17] Soon-Gyo Jung, Johanne Medina, Kholoud Aldous, Jinan Azem, Joni Salminen, and Bernard J. Jansen. 2025. Cipherbot: A Learning Platform for AI-Augmented Education. In *Proceedings of the Augmented Humans International Conference 2025 (AHs '25)*. Association for Computing Machinery, New York, NY, USA, 478–481. doi:10.1145/3745900.3746106

[18] Soon-Gyo Jung, Joni Salminen, Kholoud Khalil Aldous, and Bernard J Jansen. 2025. PersonaCraft: Leveraging language models for data-driven persona development. *International Journal of Human-Computer Studies* 197 (2025), 103445. doi:10.1016/j.ijhcs.2025.103445

[19] Soon-gyo Jung, Joni Salminen, Haewoon Kwak, Jisun An, and Bernard J. Jansen. 2018. Automatic Persona Generation (APG): A Rationale and Demonstration. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 321–324. doi:10.1145/3176349.3176893

[20] Stacey Kuznetsov and Martin Tomitsch. 2018. A Study of Urban Heat: Understanding the Challenges and Opportunities for Addressing Wicked Problems in HCI. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174137

[21] Yiren Liu, Pranav Sharma, Mehul Oswal, Haijun Xia, and Yun Huang. 2025. PersonaFlow: Designing LLM-Simulated Expert Perspectives for Enhanced Research Ideation. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 506–534. doi:10.1145/3715336.3735789

[22] Lene Nielsen. 2019. *Personas - User Focused Design* (2nd ed.). Springer.

[23] Pinecone Systems. 2023. Pinecone: Vector Database for Machine Learning. https://www.pinecone.io/.

[24] John Pruitt and Tamara Adlin. 2006. Personas: practice and theory. In *Proceedings of the Conference on Designing for User eXperience*.

[25] Joni Salminen, Kathleen Guan, Soon-gyo Jung, Shammur Absar Chowdhury, and Bernard J Jansen. 2020. A literature review of quantitative persona creation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14. doi:10.1145/3313831.3376502

[26] Joni Salminen, Bernard Jansen, and Soon-Gyo Jung. 2022. Survey2Persona: Rendering survey responses as personas. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 67–73. doi:10.1145/3511047.3536403

[27] Joni Salminen, Soon-Gyo Jung, and Bernard J. Jansen. 2021. Are data-driven personas considered harmful?: Diversifying user understandings with more than algorithms. *Persona Studies* 7, 1 (July 2021), 48–63. doi:10.3316/informit.352977339951659

[28] Andreas Schuller, Doris Janssen, Julian Blumenröther, Theresa Maria Probst, Michael Schmidt, and Chandan Kumar. 2024. Generating personas using LLMs and assessing their viability. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7. doi:10.1145/3613905.3650860

[29] Norbert Schwarz. 1999. Self-reports: How the questions shape the answers. *American Psychologist* 54, 2 (1999), 93.

[30] Joongi Shin, Michael A Hedderich, Bartłomiej Jakub Rey, Andrés Lucero, and Antti Oulasvirta. 2024. Understanding human–AI workflows for generating personas. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 757–781. doi:10.1145/3643834.3660729

[31] Lipeipei Sun, Tianzi Qin, Anran Hu, Jiale Zhang, Shuojia Lin, Jianyan Chen, Mona Ali, and Mirjana Prpa. 2025. Persona-L has entered the chat: Leveraging LLMs and ability-based framework for personas of people with complex needs. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–31. doi:10.1145/3706598.3713445

[32] Trang Xuan, Joni Salminen, Ilkka Kaate, Farhan Ahmed, Danial Amin, Rajat Patil, Soon-Gyo Jung, Jinan Y. Azem, and Bernard J. Jansen. 2026. From Priming Modalities to Educational AI Chatbot Engagement: A Study of 67 Learners. *International Journal of Human–Computer Interaction* (2026). doi:10.1080/10447318.2026.2654818 Published.

[33] Xishuo Zhang, Lin Liu, Yi Wang, Xiao Liu, Hailong Wang, Anqi Ren, and Chetan Arora. 2023. PersonaGen: A tool for generating personas from user feedback. In *2023 IEEE 31st International Requirements Engineering Conference (RE)*. IEEE, 353–354. doi:10.1109/RE57278.2023.00048