PersonaBase: A Formative Evaluation Study of Computational Resources for Data-Driven Personas

Joni Salminen
School of Marketing and Communication
University of Vaasa
Vaasa, Finland
joolsa@utu.fi

Danial Amin
School of Marketing and Communication
University of Vaasa
Vaasa, Finland
danial.amin@uwasa.fi

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

Bernard J. Jansen
Department of Information Management
Peking University
Peking, China
jjansen@acm.org

Abstract

Drawing inspiration from the natural language processing community's definitions of tasks, datasets, metrics, and baselines, we present PersonaBase, a repository of computational resources for data-driven persona (DDP) research. PersonaBase contains (1) computational notebooks that present algorithmic approaches for persona generation and evaluation, (2) persona generation prompts, (3) persona systems, and (4) datasets with user data or personas. Formative evaluation with four domain experts suggests that PersonaBase addresses a real need and suggests that DDP benchmarking can be further developed through community building and learning from other computing sciences.

CCS Concepts

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

Keywords

Data-driven personas, benchmarking, computational resources

ACM Reference Format:
Joni Salminen, Danial Amin, Ilkka Kaate, and Bernard J. Jansen. 2026. PersonaBase: A Formative Evaluation Study of Computational Resources for Data-Driven Personas. In *37th ACM Conference on Hypertext (HT '26), September 14–18, 2026, London, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3800935.3830833

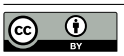
1 Introduction

In human-computer interaction (HCI), representing real user groups through fictional *personas* [10] is a methodology that stimulates empathy and guides decision making among developers, designers, and other stakeholders who need information about users [13, 29]. Personas aim to provide an empathetic perspective to user needs based on data rather than an abstract idea or vague assumptions about the users [1, 10, 25]. This perspective then supports the user-centered development of products, including hypertext systems [5, 31, 33]. Although personas were initially created using small qualitative data [14], scholars adopted large-scale data sets to develop *data-driven personas* (DDPs), which are based on quantitative data and analysis techniques [18, 20, 26, 27].

These DDPs are driven by the application of data science techniques, such as machine learning (ML) algorithms, to structured datasets about people [4, 26]. The progress in natural language processing (NLP) and generative artificial intelligence (GenAI) has nudged DDP development toward the analysis of unstructured text data, such as user study transcripts, interviews, and/or open ended surveys that can be processed using GenAI techniques to develop personas [11, 36, 38, 39]. Together, the "classical" ML algorithms and the modern GenAI techniques form a major part of the state-of-the-art in DDP development.

The dependence of DDP research on ML and GenAI techniques anchors it in the computational research domain. However, DDP development currently lacks systematically developed benchmarking resources and practices to establish quality standards and methodological rigor similar to other computational fields [41, 42]. Reviews have found that less than 10% of persona studies share resources such as code, datasets, or algorithms [34, 35], and researchers have repeatedly called for improved rigor in DDP research [7, 8, 24, 37]. The lack of computational resources undermines research reproducibility and impedes scientific progress in DDPs, which is acute because persona development is increasingly affected by technological progress, including large language models (LLMs) [21, 38, 39, 43] that require robust evaluation due to inherent risks, e.g., DDPs being misaligned with actual user groups [3, 17, 32].

At the same time, LLMs also present an opportunity for replicable and scientifically measurable DDP development; for example, by sharing prompts [2]. The shareability of prompts and code now makes a leap toward persona science [37] a possibility. *But what computational resources should be shared? What needs to be shared? How to share these resources?* These questions are fundamental to progress in DDPs but remain unaddressed in the current body of DDP literature. Against this backdrop, this research addresses the following research question: *How to develop computational resources for data-driven persona research?*



HT '26, London, United Kingdom

ACM ISBN 979-8-4007-2564-7/26/09
https://doi.org/10.1145/3800935.3830833

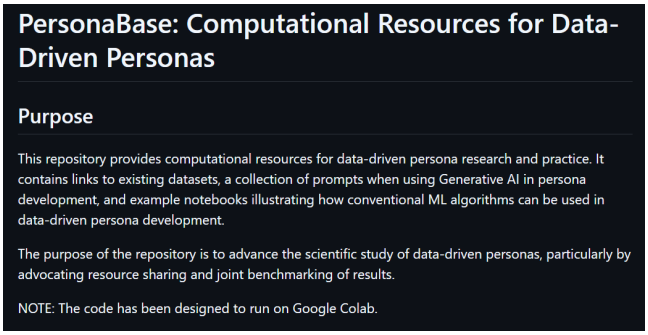


Figure 1: PersonaBase includes computational resources for learning and further developing DDP methods. The files of the repository are in https://github.com/joolsa/PersonaBase.

As a contribution of this work, we introduce `PersonaBase` (see Figure 1), a computational resource repository to advance DDP research by providing shared datasets, generative AI prompts, example computational notebooks for persona development and evaluation, and standardized resources to enable reproducible benchmarking and collaborative scientific study within the persona research community. To our knowledge, `PersonaBase` is the first attempt to systematically aggregate and organize computational resources specifically for DDP benchmarking purposes. Although individual resources exist in fragmented form (as we catalog in this work), no prior effort has provided an integrated benchmarking infrastructure for the DDP community. We note that `PersonaBase` constitutes an initial benchmarking *framework* and resource repository—not a fully established, community-adopted benchmark in the conventional sense—intended to enable and encourage standardized evaluation in DDP research.

We are targeting two "personas" with this contribution (see Figure 1): (1) novice persona researchers seeking guidance on how to generate and evaluate DDPs and (2) expert persona researchers seeking to improve the state-of-the-art.

For *novice researchers* getting started with DDPs, `PersonaBase` provides practical example notebooks that demonstrate the applicability of ML algorithms to DDP development, along with ready-to-use datasets that enable experimentation without first collecting their own data. Overall, the value proposition for the novice persona researcher is that `PersonaBase` helps them learn DDP development by applying computational resources to their own use case.

For *expert researchers* seeking to improve the state-of-the-art of DDPs, `PersonaBase` provides a shared foundation for benchmarking new methods against existing approaches using standardized datasets and metrics, for which we list and provide example code. The repository facilitates resource sharing, enables researchers to build on previous work, provides diverse datasets for testing novel techniques, and contributes to the development of common standards in DDP research.

2 Conceptual Underpinnings

Shared resources create a common vision and *scientific understanding* for comparing methods, measuring progress, promoting reproducibility, and providing a more structured entry point for newcomers to learn how to contribute. This vision is *collective* and *legitimate* within the NLP research community [23]: researchers are aware of it, accept it, and act upon it.

Benchmarking in NLP tends to be based on a few core components that establish a framework to objectively evaluate scientific contributions (see Table 1). In general, benchmark datasets define on what *tasks* models are tested, and *metrics* define how model performance is measured on those tasks [40]. In NLP benchmarking, a task is a specific, well-defined natural language problem with associated data and evaluation metrics [28], designed to test a particular capability of an NLP model (e.g., translation, question answering, sentiment analysis). Specific tasks, benchmark datasets, and metrics are *common knowledge* within the community. For example, every professional NLP researcher is expected to know the field's basic tasks, benchmark datasets, and metrics. This helps create a common conceptualization of the field and aligns the researchers' efforts to contribute to well-defined problems with measurable success. These problems are commonly decided to be worthy of pursuit.

Table 1: Benchmarking resources and practices in computational research.

Component	What it is
Task	The specific problem the core contribution is intended to solve.
Datasets	The data used for testing the core contribution on the defined task. Typically, this would be the data for DDP development.
Metrics	The quantitative measure(s) used to assess how well the core contribution performs the task on the test dataset.
Baselines	Existing, well-understood models, algorithms, or simpler methods used as a point of comparison for the given task.

Furthermore, it is vital to observe that computational resources can also be combined into *benchmark suites*, which are based on the logic that evaluating a range of tasks provides a more robust measure of general language understanding [41]. Computational techniques contribute to several stages of DDP development [18], so performance should be evaluated separately for each task using appropriate metrics, after which an overall performance interpretation of the suggested method(s) should be provided. However, most prior work reports only *one* part of the DDP generation process, typically the segmentation part, such as clustering [6], with few articles testing multiple DDP development tasks [21].

The increasing involvement of researchers with NLP and ML backgrounds in DDP efforts is exerting more pressure toward more rigorous approaches in DDP development and evaluation. For example, Li et al. [24] propose four pathways to a more rigorous approach to persona development: (1) building frameworks to determine standard persona attributes, (2) establishing theoretical

and empirical groundwork for the design of persona profiles, (3) generating datasets and evaluation standards to test and improve persona creation techniques, and (4) promoting cross-disciplinary collaboration with application-specific assessment of potential risks and biases inherent in persona generation processes.

A prerequisite for meaningful benchmarking is a conceptual understanding of what constitutes a "good" persona—that is, the desirable traits or *desiderata* that personas should exhibit. Prior literature suggests several quality dimensions, including consistency (internal coherence of persona attributes), diversity (representation of varied user segments), coverage (how well the persona set accounts for the underlying user population), and fairness (equitable representation of demographic groups) [18, 37]. As long as these desiderata can be quantified and personas measured along these dimensions, different persona sets obtained through different creation processes can be objectively compared. PersonaBase's evaluation notebooks (PE01a–PE06b) operationalize these quality dimensions, and we return to this discussion in Section 3.2.

3 PersonaBase

3.1 Taxonomy

We developed a taxonomy for organizing DDP resources through a process grounded in NLP benchmarking conventions. Specifically, we mapped the established NLP benchmarking components—tasks, datasets, metrics, baselines (see Table 1)—onto the DDP domain, identifying Persona Development (PD) and Persona Evaluation (PE) as two primary task types. We then extended this framework with resource types specific to DDP research that emerged from our systematic review of HCI venues and computational repositories: Datasets (DS), Prompts (PP), Systems (PS), and Repositories (PR). In this taxonomy, the first two letters indicate the resource type, followed by a numerical identifier and a letter suffix denoting the data source. The suffix 'a' signifies that the resource uses simulated or synthetic data, while 'b' indicates it employs real, anonymized data. For example, `PD01a` refers to a Persona development resource that utilizes simulated data, whereas `PE03b` is a Persona evaluation resource that uses real anonymized data.

3.2 Computational Notebooks

At the time of writing, `PersonaBase` contains 11 notebooks divided into two types: (1) five persona development (`PD01a-PD05a`) and (2) six persona evaluation notebooks (`PE01a-PE06b`). The notebooks were generated using Claude 4-Sonnet, with prompting provided by the lead researcher (Associate Professor), who has multiple years of experience with computational techniques and DDPs. The lead author first verified the code and its outputs, and in case of any inconsistencies, Claude was asked to provide fixes. Second, the code was verified by a PhD researcher with multiple years of industry experience in programming and data science. As such, the code in the notebooks we provide was validated by a human expert. We acknowledge that formal error rates were not tracked during this process, which is a limitation; however, the two-stage human verification ensures that the released code is functionally correct and produces the expected outputs.

In terms of content, the persona development notebooks (`PD01a to PD05a`) demonstrate five algorithmic approaches to persona generation, including non-negative matrix factorization for YouTube audience behavior (`PD01a`), clustering analysis (`PD02a`), principal component analysis (`PD03a`), Gaussian mixture models (`PD04a`), and LLM-based methods (`PD05a`). The evaluation notebooks demonstrate multiple ways to assess persona quality by computing and analyzing metrics, including measuring user perception on the persona perception scale (`PE01a`), quantifying diversity of personas (`PE02a`), analyzing the relationship between persona set size and dataset coverage (`PE03a`), and tracking behavioral metrics in interactive persona systems (`PE04a`). The evaluation notebooks benchmark different segmentation techniques (`PE05b`) and a unified evaluation approach (`PE06b`) that combines fairness, diversity, coverage, and consistency metrics to compare persona generation methods across varying set sizes, demonstrating how to select persona creation methods based on empirical evidence.

Overall, the notebooks address the gap in standardized baselines and reproducible methods by demonstrating five algorithmic approaches to persona generation (`PD01a-PD05a`) and six evaluation frameworks (`PE01a-PE06b`) with executable code that researchers can directly replicate and adapt.

3.3 Datasets

To find DDP datasets, we searched computational repositories (GitHub, Kaggle, HuggingFace) for datasets that mention personas. We included all the HCI-related datasets we could find and a sample of the NLP-related datasets[1], as involving all of them would not serve this study's purpose for building HCI-oriented resources for DDP development. As such, `PersonaBase` contains 12 datasets that can be used for developing and evaluating personas, as well as for tasks like updating persona sets. The metadata about these datasets includes (1) ID, (2) Dataset name, (3) Year released, (4) Description Realness (synthetic or real), (5) Domain type (HCI, NLP/ML, Other), (6) Content type, and (6) URL to dataset.

Inspecting the datasets in `PersonaBase` suggests a change from real, human-generated persona data in the early period (2011-2018) to predominantly synthetic, AI-generated personas from 2023 onward, with 9 of the 12 datasets (75%) released in the most recent two years (2023-2025), suggesting an accelerating trend toward synthetic persona generation coinciding with advances in LLMs. Consistent with this, the distribution of datasets indicates a divide between real (n = 3, 25%) and synthetic (n = 9, 75%) data, with all real datasets published before 2023 and all synthetic datasets emerging from 2023 onward. Moreover, the datasets are predominantly oriented toward NLP/ML applications (n = 8, 67%) compared to HCI-focused uses (n = 4, 33%), reflecting the focus of the NLP field on conversational AI and language model alignment rather than user experience or usability.

For this same reason, many of the datasets actually contain brief text snippets that are presented as "personas", or a set of structured data containing attributes like demographics, hobbies, etc. There

[1]We are aware of the fact that our inclusion of NLP/ML datasets mentioning personas is not comprehensive; however, it is adequate for making the analytical reasoning that we do. For a more thorough review of NLP/ML persona datasets, we direct the reader to Chen and colleagues' review of 22 datasets used in personalized dialogue research [9].

are only two datasets (DS01 and DS03) focused on persona development; that is, containing individual or grouped data about people toward the purpose of using that data to develop personas, with the two other HCI-focused datasets developed for analyzing immersive deepfake personas (DS06)—that is, personas that use AI-generated synthetic faces (deepfakes) as profile images or video presentation of the persona rather than stock photos or illustrations—and demographic and other biases in LLM-generated personas (DS09). The rest of the datasets contain personas in three types: (1) persona descriptions (n = 5, 42%), (2) persona dialogues (n = 4, 33%), and (3) deepfake personas.

The datasets also exhibit a clear evolution in scale and complexity, from early small-scale collections with simple biographical facts (e.g., Persona-Chat's (DS02) 4-5 sentence profiles across 10K dialogues) to recent massive repositories containing millions of multidimensional personas with rich demographic grounding, psychological attributes, and domain-specific expertise (e.g., PersonaHub's 1 billion personas (DS11)), Nemotron's 100K personas with 16+ attributes (DS10)).

Third, coupled with the synthetic nature, there is often a pronounced strive for diversity, with many of the datasets explicitly mentioning the idea of "being aligned with real populations" (DS08) and representing diverse perspectives (DS07, DS09, DS11)—this is a clear departure from the "handful of personas" that "describe the average users" paradigm in classic persona development (DS01, DS03). Fourth, while previous persona profiles consisted of approximately 5-15 attributes [30], these novel computational personas include a clearly higher number of attributes (e.g., 32 (DS10) and 90 (DS12)), which are typically provided as structured data, not as persona profiles "of old". This, again, is a consequence of novel persona generations' strive to represent diversity better than the handful of personas commonly generated previously [19]. From an HCI perspective, the increase in the number and complexity of personas imposes new interaction challenges because persona users can easily struggle with the information overload related to large persona sets [22]. To this end, DDP researchers emphasize the need for making personas more accessible and navigable by enabling search, filtering, and categorization of personas [20, 37].

We also note additional recent datasets relevant to DDP research, such as the PersonaGen dataset by Gugliotta et al. [15], which provides persona-driven machine-generated text. Such datasets further illustrate the rapidly evolving landscape and the need for a repository that is continuously updated. Overall, the persona datasets collected for PersonaBase address the gap in shared, standardized data for benchmarking by providing both synthetic and real anonymized datasets that enable researchers to test and compare persona generation methods without proprietary data.

3.4 Systems

We reviewed prior DDP research from high-ranking HCI conferences, including ACM CHI, IUI, UIST, UMAP, and DIS, to identify articles that develop persona systems. Through this process, we identified 19 systems. In the PersonaBase repository, we recorded the following information about each system: (1) ID, (2) Name of system, (3) Year introduced, (4) Description, and (5) URL. These systems can serve as baselines for comparison or inspiration for further DDP development.

Investigating these systems, we can observe a dramatic acceleration in development since 2023, with only two systems (11%) introduced before 2023 compared to 17 systems (89%) emerging from 2023 onward. This is likely associated with the widespread availability of LLMs and the resulting explosion of automated persona generation capabilities across diverse application domains. The systems exhibit a trajectory from early specialized data-processing tools handling specific input types–such as APG's PS01 social media analytics (2018) and S2P's PS02 survey data focus (2022)–to the post-2023 explosion of LLM-powered frameworks that include interactive simulations (e.g., PS04), multimodal persona creation combining text, image, and video (PS05), domain-specific applications (e.g., PS06 for requirements engineering, PS17 for accessibility needs, PS19 for audience simulation), and collaborative workflows (PS13, PS14, PS16).

Five of the systems (26%) make their source code (or parts of it) publicly available (PS04, PS08, PS10, PS12, and PS15) (links available in PersonaBase). Overall, as our contribution, PersonaBase lists, describes, and links to 19 DDP systems. This addresses the gap in accessible persona development tools by cataloging existing systems that span from basic data processing to LLM frameworks, enabling researchers to learn from, benchmark against, and build upon existing work.

3.5 Prompts

Due to the apparent influence that LLM technologies have had on DDP research and practice, we decided to include prompts (i.e., instructions for automated persona generation using LLMs) as part of PersonaBase. For this, we extensively reviewed articles published in HCI top venues since the introduction of LLMs, and we also included arXiv publications in this search because some of the articles reporting persona prompts were only available as preprints at the time of our investigation. We found 27 articles that share their prompts for persona development or evaluation. Using the PersonaBase, the user can easily copy a given prompt for their own experiments.

The *Persona Generation Methods* (n = 20, 74.1%) prompts focus on approaches for generating personas. We manually categorized these into five categories, which are (1) direct generation prompts such as PP02's skeletal persona creation followed by expansion, (2) demographic-driven synthesis (e.g., PP05), (3) attribute-based generation (e.g., PP10), (4) two-stage summary-to-detailed approach (e.g., PP27), and (5) iterative workflow-based generation methods like PP11's multi-stage process combining interview analysis with trait extraction, PP21's rule-based validation workflow, and PP25's thematic analysis pipeline. The *Persona Evaluation and Application* prompts (n = 7, 25.9%) focus on using, evaluating, or interacting with already-generated personas. This includes evaluation frameworks such as coherence and consistency rating prompts (e.g., PP04), quality assessment metrics (e.g., PP12), and persona validation based on user perceptions of personas (e.g., PP14). Some prompts are integrated into complete interactive persona systems like PP18's audience understanding tool with chat-based persona interaction,

PP23's emotion-driven conversational personas, and PP24's writer-defined feedback generation. Overall, researchers can use prompts in PersonaBase as starting points in their own projects rather than developing prompts from scratch.

3.6 Repositories

During the development of PersonaBase, we identified other computational resources on GitHub concerning DDPs. We decided to list these, along with their metadata, including (1) ID, (2) Repository name, (3) Year, (4) Description, (5) Contains code, (6) Contains data, (7) Associated with a research paper, and (8) URL. These repositories illustrate how to set up computational repositories for DDP projects. Inspecting the repositories, we can observe a progressive maturation from 2019 to 2024, evolving from documentation-focused efforts (PR01 providing descriptive role profiles) and foundational resources (PR02 survey templates and image libraries in 2021) toward increasingly sophisticated computational approaches, including algorithmic comparison frameworks (PR03, PR04), AI-automated generation tools (PR05, PR06), and recent advanced integrations of LLMs with human-AI collaborative workflows (PR07) and specialized analytical methods (PR08). There is a notable trend showing that 75% (6 of 8) repositories include executable code, 63% (5 of 8) provide datasets, and 75% are associated with research publications, reflecting sound practices in providing reproducible, data-driven, and rigorous persona development methodologies.

4 Formative Evaluation Study

4.1 Approach

To evaluate PersonaBase, we conducted semi-structured Zoom interviews of four persona experts with different levels of experience with DDPs (see Table 2). We emphasize that this evaluation is explicitly *formative*—its purpose is to identify improvement directions for PersonaBase rather than to validate community-wide standards. Formative evaluations with small expert samples are an established practice in HCI to obtain early-stage feedback on research prototypes [29]. The four experts were purposefully selected to represent our two target user groups: two are experts in DDPs specifically (E01 and E03, corresponding to the "expert persona" we had defined), while two are less versed in DDPs and more geared toward traditional design personas (E02 and E04, corresponding to the "novice persona" we had defined).

Despite the small sample, the experts exhibit thematic convergence in five themes, providing consistent feedback on issues such as entry barriers, evaluation challenges, and community building needs. All had computer science or HCI degrees and both industry and research experience, with current focus on academic research. The identification of the experts was based on published research on personas, and they were recruited by direct messages requesting feedback on developing computational resources for persona development. We first asked the experts about their persona experience and computational resource experience; we then walked the experts through PersonaBase by screen sharing and asked their feedback, especially on how the repository could be further improved.

The sessions lasted about half an hour each (M = 35.5, SD = 14.8 minutes), and they were moderated by the lead author. The sessions were recorded with the participants' permission and transcribed for further analysis. The study moderator took notes during the sessions and analyzed the transcripts using Claude 4.5-Sonnet, a process that included asking Claude to summarize the sessions and then manually verifying the summaries by reading them line-by-line and editing them for factual accuracy. We acknowledge that this pragmatic analytical approach—Claude-assisted summarization followed by manual verification—does not constitute formal thematic analysis with inter-rater reliability. However, we consider it appropriate for the exploratory, formative purpose of this study, where the goal is to identify expert concerns and improvement suggestions rather than to produce a definitive coding of themes. We encourage future work to conduct larger-scale evaluations with structured coding frameworks. The summary findings, based on the session recording transcripts and moderator notes, are presented in the following subsections.

ID	Nationality	Age	Gender	DDP[a]	CRE[b]
E01	South Korea	30s	Female	High	High
E02	China	20s	Female	High	High
E03	USA	60s	Male	High	Medium
E04	Turkey	30s	Male	Medium	Medium

Table 2: Expert demographics and expertise. [a]DDP Experience. [b]Computational Resource Experience.

4.2 Entry Barriers and Guided Discovery

Experts identified the challenge of helping users get started with PersonaBase, particularly when thinking of the experience from the perspective of those new to DDPs or from non-computational backgrounds. E01 recalled that when first learning about DDPs, "it was very hard to imagine about what kind of persona outcome there should be," pointing out the cold start problem that novices face when approaching an unfamiliar resource landscape. E04 echoed this concern, asking "What is the pipeline? Where do I start, right? And what should I expect at the end?" and suggesting that prerequisites be made explicit—"If you never use Python, don't bother, right?"—to set appropriate expectations. E03 observed that the repository "is definitely set up for the computational persona researchers" and that many persona researchers would find it "beyond their capability and not really even in their paradigm," implying that PersonaBase is currently more oriented for technical DDP research community.

To increase the accessibility of PersonaBase, experts proposed various forms of interactive navigation. E02 suggested search and filtering features so that "the user can come in and even their intent is very vague" and still receive guidance, warning that without such mechanisms users "will feel a little bit overwhelmed because I assume the list of everything will still like growing and growing." E01 similarly recommended "visualizations or more interactive ways" of providing information beyond the static GitHub format. E04 proposed a chatbot interface and FAQ section addressing questions like "I have this type of data, what can I do?" to help users match their specific situations to appropriate resources.

In response to this feedback, we improved PersonaBase with a FAQ file containing step-by-step information, a "Getting Started" guide mapping user backgrounds to appropriate entry points, and clear documentation of prerequisites. While some computational literacy is inherently required for computational DDP resources, these additions aim to lower the barrier for researchers transitioning from qualitative to quantitative persona methods.

4.3 Contextual Configurability

Experts emphasized that persona attributes, methods, and evaluation criteria must vary based on domain, purpose, and organizational context, complicating standardization efforts. E01 noted that their own work "deliberately excluded traditional demographics like gender in favor of focusing on users' potential jobs and interests," illustrating how purpose shapes attribute selection. They proposed persona templates as a solution to both "the attribute selection (i.e., what information to select for personas) and the information presentation (i.e., how many attributes to select) problems," suggesting that domain-specific starting points like e-commerce or writer-focused templates would help researchers overcome the "blank page problem." They recommended that these aspects be improved in PersonaBase.

However, E01 also acknowledged fundamental limits to standardization: "even within a specific domain, datasets and use cases vary, which hinders standardization." The experts expressed uncertainty about whether comparative tables mapping personas to goals would be "actually meaningful for reproducing them" given the complexity introduced "not only the method, but also data part." This suggests that the experts think there is a trade-off between structured guidance and the reality of contextual variation, which implies that PersonaBase must balance standardization with some form of flexibility or configuration, perhaps by documenting the scope of purposes for each resource rather than prescribing universal approaches.

4.4 Support for Evaluation

Experts repeatedly referred to DDP evaluation as a key challenge, so that there is uncertainty about appropriate metrics, baselines, and judging approaches. E01 noted that despite including notebooks for diversity, coverage, and perception analysis, fundamental questions remain about appropriate baselines: "For persona task, it's really hard to have the correct baseline. So because like, is it a human reflecting the real world or is it using the different method with generation task, like computation method, or is it designers should design for personas?" E02 similarly expressed difficulty navigating evaluation options: "it's very hard to find a kind of metrics... and also it's very context dependent," suggesting that PersonaBase could better guide users toward appropriate evaluation approaches for their specific contexts.

Regarding automated evaluation, E01 shared experiences with LLM-based assessment that achieved only "approximately sixty percent agreement with human evaluators," leading them to advocate for "human AI collaboration approach for evaluation" where LLMs serve as first-stage filters followed by human judgment. They suggested PersonaBase's evaluation resources could incorporate such hybrid approaches, noting that "persona will be more, much more subjective" than other evaluation tasks. E03 proposed that PersonaBase could help establish shared evaluation standards through computational competitions, suggesting metrics focused on "efficiency metrics like computational cycles, processing speed, or coverage measures such as the proportion of outliers not captured by generated personas." However, they acknowledged that formalizing such benchmarks requires careful attention to context-setting, which PersonaBase could help standardize.

4.5 Target Audience Definition

Experts raised questions about the repository's target audience and strategies for ensuring adoption. E03 distinguished between researchers and practitioners, noting that the repository "appears more geared toward researchers because practitioners typically have their own approaches, contexts, and proprietary data, making shared datasets less relevant to their needs." E01 observed a potential mismatch between positioning and content, suggesting that "the materials seem oriented toward novices learning about DDPs rather than experienced researchers seeking advanced resources."

For dissemination, E03 emphasized proactive outreach: "I really would encourage the marketing aspect of it to make, to get it, get it out there in front of researchers and practitioners, maybe a lot of computational practitioners that may be interested in the code internally for their companies." They recommended establishing a dedicated website rather than relying solely on GitHub and suggested that publications describing the resource would prove essential for building community awareness. E02 responded positively to learning about human curation, stating that manual verification made the repository "feel more trustworthy compared to automated paper collectors that lack quality control."

4.6 Community Building

Experts envisioned extending the repository's value through interactivity, competition, education, and generative applications. E03 advocated for competition-based mechanisms, drawing parallels to SemEval workshops: "doing something like the Semaval folks did might be a good approach, where, kind of incur, running workshops and competitions in these computational persona areas to get persona folks to use it." Such competitions would transform PersonaBase from a passive collection into an active research community with shared evaluation standards.

Educational applications also emerged as a promising direction. E03 noted that "if I was teaching, this would be a great benefit. I mean, all the data is already collected, the algorithms are there. I could assign every student or a team of students a different algorithmic approach," suggesting the repository could "jumpstart the creation of such courses" by reducing preparatory overhead. E04 proposed video-based workflow demonstrations showing "the full pipeline, starting with articulating aims and purposes, evaluating whether available data suits the intended analysis, considering algorithmic choices, and (...) producing finished personas."

Finally, E03 introduced generative applications, suggesting that validated datasets could help researchers "supplement their own data collection when they discover gaps" and that "researchers could potentially use LLMs in combination with existing PersonaBase

datasets to generate synthetic data for missing variables." This positions repository resources not merely as benchmarks but as building blocks for data augmentation in new research projects.

5 Discussion

5.1 Research Implications

Our findings suggest that DDP research has benefited a great deal from computational resources. For example, researchers typically use pre-existing data science algorithms for persona generation, including k-means, PCA, LDA, LSA, and so on (see review in [35]). All of these algorithms have been developed and made publicly available by researchers outside the persona development domain, stemming from problem spaces other than user representation, segmentation, or persona generation. Therefore, DDP research has *effectively appropriated* research and development in open science; yet, it has contributed little back and often applies methodologies that were not tailored for the specific purpose of persona development. It can be argued that persona generation, as a set of computational tasks, differs from *both* general data segmentation that generic data science algorithms typically support *and* even from other types of user segmentation because personas are more comprehensive than contextually limited user segments.

Therefore, integrating computational approaches into persona development seems inevitable for the maturity of the DDP field. Computational approaches offer potential gains in scalability and replicability that manual processes cannot achieve, particularly when addressing diverse, global user bases or rapidly evolving product ecosystems [18, 27]. In addition to supporting scientific progress, development of computational resources also helps *communicate* state-of-the-art results in a field [16]. For example, given the lack of established benchmarks for DDP development, we cannot say which methods are better than others; and we cannot determine how far we are from reaching the best results. We cannot even determine, with objective measures, whether the field has evolved since DDPs were first introduced as a concept [26], because *we have no standards to compare against!* `PersonaBase` (and contributions of similar nature) aims to nudge the research community toward the development of such standards. Without shared standards, researchers risk being locked in perpetual reinvention of the wheel, i.e., each developing their own persona development method that ultimately achieves the same result than previous methods.

A key challenge is that computational contributions in DDP research are limited not only in terms of *availability* but also in terms of *mindset*, meaning that researchers in this space are not collectively pursuing performance improvements using shared tasks, datasets, and metrics. This fragmentation likely hampers scientific progress in DDPs, although proving that is rather difficult. It is more easily observable that DDP researchers often rely on their own task definitions, datasets, and metrics rather than trying to find common ground.

Our formative evaluation with four persona experts suggests that the current absence of computational resources, especially regarding benchmarks and examples, actually hinders both novice and expert researchers, making it difficult to learn from previous work or determine whether new methods represent genuine improvements. On the other hand, the experts articulated that thoughtfully designed computational resources can make DDP methods more accessible while preserving—or even improving—the possibility for contextual, qualitative interpretation. To this end, `PersonaBase` aims to support robust benchmarking practices while explicitly providing diverse evaluation frameworks. Practitioners and researchers are encouraged to visit and contribute to the `PersonaBase` resources in the online repository[2].

Finally, in 2018, Duda posed the question, "who owns personas?" [12]. She was referring to the differences between how HCI and marketing professionals use personas. At present, this question can be posed from perspective of computational techniques dominating the conceptual space of personas, which can yield a rift between computationally oriented ("technical") and qualitatively oriented ("traditional") persona researchers[3]. We believe that a solution to this rift is designing persona research in a way that incorporates qualitative ideas and research traditions when implementing new technologies in persona development. In turn, qualitative approaches could also benefit from adopting ideas from the computational side of personas, for example, using LLMs to facilitate the analysis of unstructured data for persona generation [11]. Resolving the (largely unnecessary) rift between quantitatively and qualitative oriented persona researchers could help to build a stronger community for persona research in HCI.

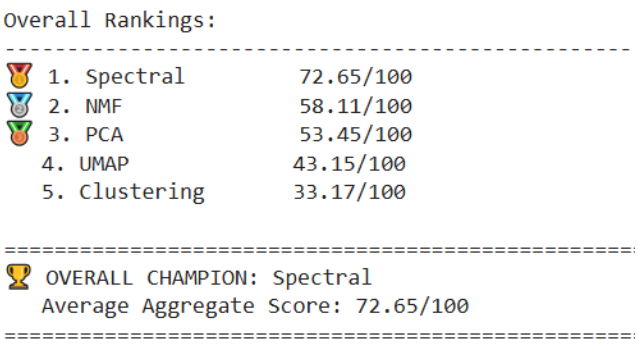
5.2 Risks of Standardization in Persona Research

An important consideration in developing computational standards for DDPs concerns the ethical and societal implications of formalizing such standards. Benchmarking standards can shape research practices and, consequently, the representations of users that personas embody. For example, there is a risk that standardized metrics may inadvertently privilege certain user representations, such as optimizing for demographic coverage leading to exclusion of marginalized groups that are underrepresented in training data. Similarly, the heavy reliance on synthetic persona datasets (75% of datasets in `PersonaBase`) raises concerns about bias amplification, where biases present in LLM training data propagate into benchmark datasets and, through computational methods, into deployed personas [3]. `PersonaBase` partially addresses this risk through the inclusion of fairness as a benchmarking dimension in the computational and by cataloging datasets that explicitly address bias. Nevertheless, the DDP community should remain vigilant about the risk that computational resources inadvertently encode normative assumptions about user representation, and the field should incorporate ongoing ethical review of its standards and resources.

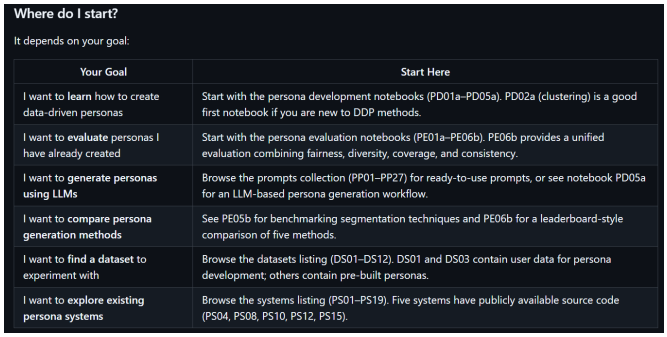
There is also a systemic risk that excessive 'benchmarkification' could lead to metric gaming rather than genuine progress if implemented carelessly (owing to the old adage, 'a measure seizes to be good when becoming a target'). However, our findings suggest

[2] https://github.com/joolsa/PersonaBase

[3] There is also another way to frame the "who own personas" question, which is HCI versus NLP. Interestingly, some HCI researchers consider the NLP usage of the term misleading because NLP personas are too "thin". For example, E04 referred to this as concept abuse and voiced a strong opinion, "So maybe they [NLP researchers] need to call it something else, not persona."



(a)



Where do I start?

It depends on your goal:

Your Goal	Start Here
I want to **learn** how to create data-driven personas	Start with the persona development notebooks (PD01a–PD05a). PD02a (clustering) is a good first notebook if you are new to DDP methods.
I want to **evaluate** personas I have already created	Start with the persona evaluation notebooks (PE01a–PE06b). PE06b provides a unified evaluation combining fairness, diversity, coverage, and consistency.
I want to **generate personas using LLMs**	Browse the prompts collection (PP01–PP27) for ready-to-use prompts, or see notebook PD05a for an LLM-based persona generation workflow.
I want to **compare persona generation methods**	See PE05b for benchmarking segmentation techniques and PE06b for a leaderboard-style comparison of five methods.
I want to **find a dataset** to experiment with	Browse the datasets listing (DS01–DS12). DS01 and DS03 contain user data for persona development; others contain pre-built personas.
I want to **explore existing persona systems**	Browse the systems listing (PS01–PS19). Five systems have publicly available source code (PS04, PS08, PS10, PS12, PS15).

(b)

Figure 2: Artifacts developed based on the formative evaluation study: (a) Leaderboard example from `PersonaBase` (PE06b) that we developed following the formative evaluation study. This notebook tests five different algorithms in persona generation task on multiple evaluation criteria, including accuracy, consistency, diversity, and coverage, and reports the average performance score. (b) Excerpt from the *Frequently Asked Questions* added based on the formative evaluation study.

that the greater risk currently facing DDP research is not overstandardization but rather the absence of shared evaluation standards, which prevents cumulative knowledge building and makes it impossible to distinguish methodological progress from mere proliferation. The solution is not to completely avoid benchmarking, but to develop benchmarking practices that preserve holistic evaluation of personas and their effects on real-world outcomes while facilitating systematic comparison.

5.3 Future Research Directions

Hardy et al. [16] argue that effective benchmarks must offer realistic evaluations (1) grounded in real-world scenarios, (2) integrate relevant domain knowledge, (3) clearly communicate their scope and objectives, (4) assess varied tasks, (5) provide sufficient difficulty to prevent rapid expiration, and (6) consider multiple performance dimensions instead of depending on singular metrics. Although `PersonaBase` addresses some of these criteria, more work remains to be done. In the following, we propose future research directions (RDs) that we believe would benefit the research community.

RD01: Successful resource analysis. The availability of resources can accelerate research and development of DDPs by simplifying implementation and experimentation, as seen in NLP. Therefore, investigating the history of NLP resources could yield important insights for computational resource development in DDPs (and HCI, more broadly). This could include, for example, analyzing BLEU (33,115 citations in Google Scholar as of April 2025) and ROUGE (19,393 citations in Google Scholar as of April 2025) to investigate what made them successful in NLP and what could be learned from their success when building DDP resources. Perhaps we should interview the creators of these resources to try and understand why the resources were successful, and suggest some broader "critical success factors" that could be applied by DDPs and beyond. At the same time, it is also worth acknowledging that benchmarking in other disciplines is not "perfect, either. The reader might have gotten the impression that everything is ideal in NLP, but in reality, benchmarking practices are under debate and periodically flux in NLP, too. For example, many benchmark datasets became redundant or require updating in the wake of radical improvements in GenAI technologies. Therefore, benchmarking requires constant attention and corrective measures from the research community.

RD02: Gaps and opportunities in resource dissemination. The experts emphasized that example implementations of DDPs remain "out of reach for a lot of researchers, especially persona researchers who have been primarily qualitative" (E03). This implies an accessibility barrier in which computational specialists can leverage cutting-edge techniques in DDP development, while qualitative researchers cannot participate as easily. `PersonaBase`'s notebooks provide documented, executable code that lowers technical barriers for researchers from non-computational backgrounds. Nevertheless, the effort to make computational resources more accessible for DDP researchers and practitioners requires ongoing maintenance as the DDP field continues its evolution and production of new computational resources.

There are at least three knowledge dissemination gaps that require in-depth investigation. First, the knowledge from DDP research to persona practice does not seem to transfer effectively. One can observe practitioners creating personas using poor methodologies and developing antipathy toward the persona technique, as a consequence. For example, using DDPs in companies involves several organizational roles, including designers, researchers, developers, and managers who have different entry points and levels of methodological understanding. Ensuring the participation of multiple stakeholders needed for computational resource development requires a "buy in," which in turn requires communicating the need for these resources (and DDPs) in plain language. Second, the knowledge between DDP research and supporting fields (for example, NLP and ML) does not transfer effectively. For example, DDP development overlaps with practices in specific computational tasks, for example, clustering is well studied in computational benchmarking. What can be learned from those studies to make persona segmentation a benchmarkable task? The third aspect here is the broader concept of usability of computational resource repositories;

in platforms like GitHub, navigability and usability of computational resource repository are not often considered or optimized for. Yet, our experts indicated a clear need for usability factors in computational software repositories, which highlights an important avenue for future research, not only applicable to `PersonaBase` but to other broad computational resource repositories, too.

RD03: Boundaries for benchmarking in DDP research. Future research should also critically examine any conflicts between benchmarking activities and the "nature" of HCI and DDPs in terms of providing insights for decision making through the mechanism of user representation. Possible conflict points are, among others, standardization versus contextualization, quantitative versus qualitative assessment, and immediate versus longitudinal evaluation. Furthermore, the experts' concern about researchers "re-trudging ground that earlier qualitative researchers have already done" (E03) verbalizes knowledge transfer challenges when computational researchers enter the DDP domain without knowledge of its history. This can result in inefficiency in which computational innovations rediscover already established best practices, or fail to do so, thereby losing the opportunity to incorporate hard-won domain knowledge into new DDP approaches. Repositories like `PersonaBase` could address this challenge by including curated references to the foundational persona literature, perhaps with annotations explaining how established principles relate to computational methods. For example, linking algorithmic diversity metrics to discussions of persona set completeness would help computational researchers better motivate technical choices in the light of persona theory.

RD04: Community building around tooling. Resource development should be approached as a community effort rather than a straightforward technical implementation. This is because the underlying "glue" that holds effective benchmarking together is the community. Without a community with a shared understanding of tasks, datasets, metrics, and baselines, there cannot be a joint effort in progress that utilizes those resources. This means that "getting benchmarking going" in DDPs is largely a *mission* that requires social actions, such as events (e.g., workshops, challenges), publication opportunities (e.g., special issues, mini-tracks), and joint coordination (e.g., research projects, labs). But how are research communities built around computational resources? Again, this remains an interesting and impactful question for future work, in which experiences from senior members of the computational sciences community can be of great value.

6 Conclusion

Our investigation indicates that DDP research faces several challenges in its development into a mature field. The field's computational tasks have not been formalized sufficiently to enable systematic investigation approaches. Metrics remain fragmented with few widely adopted standards. Researchers frequently apply similar methods that could function as baselines, but these require explicit, systematic usage that is currently missing. Research tends to favor interpretative studies over experimental designs, limiting methodological rigor. Code and data sharing practices, albeit present in the literature, are not actively leveraged to reproduce results. Finally, the community lacks norms that would encourage benchmarking practices. To address these matters, in this work, we have proposed adopting practices from other computational sciences. We have suggested specific areas of computational resources that the field should pursue and outlined directions for future efforts, including (1) clearly defining the computational tasks involved, (2) establishing evaluation frameworks and metrics, (3) developing tools and platforms, (4) curating benchmark datasets, and (5) formulating principles for the responsible use of "unpredictable" technologies such as GenAI. Consolidating computational resources can accelerate research, improve practical application, and ensure responsible development of next-generation data-driven personas in HCI and adjacent fields, including the development of hypertext systems.

Declaration of Generative AI

The authors used Claude 3.7 Sonnet and Gemini 2.5 Pro to improve the content and its presentation style. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for it.

Acknowledgments

We would like to thank Baki Kocaballi, Yoonseo Choi, and Jiangnan Xu for their valuable feedback that enabled this research.

References

[1] Moneerah Almeshari, John Dowell, and Julianne Nyhan. 2019. Using Personas to Model Museum Visitors. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP'19 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 401–405. doi:10.1145/3314183.3323867

[2] Danial Amin, Joni Salminen, Farhan Ahmed, Sonja M.H. Tervola, Sankalp Sethi, and Bernard J. Jansen. 2026. Creating and Evaluating Personas Using Generative AI: A Scoping Review of 81 Articles. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3772318.3790608

[3] Danial Amin, Joni Salminen, Bernard J. Jansen, Joongi Shin, and Dae Hyun Kim. 2025. Generative AI personas considered harmful? Putting forth twenty challenges of algorithmic user representation in human-computer interaction. *International Journal of Human-Computer Studies* 205 (Nov. 2025), 103657. doi:10.1016/j.ijhcs.2025.103657

[4] Jisun An, Haewoon Kwak, Soon-gyo Jung, Joni Salminen, M. Admad, and B. Jansen. 2018. Imaginary people representing real numbers: Generating personas from online social media data. *ACM Transactions on the Web (TWEB)* 12, 4 (2018), 1–26. ISBN: 1559-1131 Publisher: ACM New York, NY, USA.

[5] Tom Blount, David E. Millard, and Mark J. Weal. 2015. An Investigation into the Use of Logical and Rhetorical Tactics within Eristic Argumentation on the Social Web. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media (HT '15)*. Association for Computing Machinery, New York, NY, USA, 195–199. doi:10.1145/2700171.2791052

[6] Jonalan Brickey, Steven Walczak, and Tony Burgess. 2012. Comparing semi-automated clustering methods for persona development. *IEEE Transactions on Software Engineering* 38, 3 (2012), 537–546.

[7] Chris Chapman, Edwin Love, Russell P. Milham, Paul ElRif, and James L. Alford. 2008. Quantitative Evaluation of Personas as Information. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 52, 16 (Sept. 2008), 1107–1111. doi:10.1177/154193120805201602

[8] Chris Chapman and Russell P. Milham. 2006. The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 5 (Oct. 2006), 634–636. doi:10.1177/154193120605000503

[9] Yi-Pei Chen, Noriki Nishida, Hideki Nakayama, and Yuji Matsumoto. 2024. Recent Trends in Personalized Dialogue Generation: A Review of Datasets, Methodologies, and Evaluations. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. ELRA and ICCL, Torino, Italia, 13650–13665. https://aclanthology.org/2024.lrec-main.1192/

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

[11] 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

[12] Sabrina Duda. 2018. Personas—Who Owns Them. In *Omnichannel Branding: Digitalisierung als Basis erlebnis- und beziehungsorientierter Markenführung*, Vittoria von Gizycki and Carola Anna Elias (Eds.). Springer Fachmedien Wiesbaden, Wiesbaden, 173–191. doi:10.1007/978-3-658-21450-0_8

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

[14] Jonathan Grudin and John Pruitt. 2002. Personas, participatory design and product development: An infrastructure for engagement. In *Proc. PDC* (2002), Vol. 2. 144–152.

[15] Claudio Gugliotta, Lucio La Cava, and Andrea Tagarelli. 2025. PersonaGen: A Persona-Driven Open-Ended Machine-Generated Text Dataset. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. ACM, 6397–6401.

[16] Amelia Hardy, Anka Reuel, Kiana Jafari Meimandi, Lisa Soder, Allie Griffith, Dylan M Asmar, Sanmi Koyejo, Michael S. Bernstein, and Mykel John Kochenderfer. 2025. More than Marketing? On the Information Value of AI Benchmarks for Practitioners. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 1032–1047. doi:10.1145/3708359.3712152

[17] Helena A. Haxvig, Vincenzo D'Andrea, and Maurizio Teli. 2025. "I've never seen a glass ceiling better represented": Bias and gendering in LLM-generated synthetic personas from a participatory design perspective. *International Journal of Human-Computer Studies* 205 (Nov. 2025), 103651. doi:10.1016/j.ijhcs.2025.103651

[18] Bernard Jansen, Joni Salminen, Soon-gyo Jung, and Kathleen Guan. 2021. Data-driven personas. *Synthesis Lectures on Human-Centered Informatics* 14, 1 (2021), i–317. ISBN: 1946-7680 Publisher: Morgan & Claypool Publishers.

[19] Bernard J. Jansen, Soon-gyo Jung, and Joni Salminen. 2019. Creating Manageable Persona Sets from Large User Populations. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, LBW2713:1–LBW2713:6. event-place: Glasgow, Scotland Uk. doi:10.1145/3290607.3313006

[20] Bernard J. Jansen, Joni O. Salminen, and Soon-Gyo Jung. 2020. Data-driven personas for enhanced user understanding: Combining empathy with rationality for better insights to analytics. *Data and Information Management* 4, 1 (2020), 1–17. ISBN: 2543-9251 Publisher: De Gruyter Poland.

[21] S.-G. Jung, J. Salminen, K. K. Aldous, and B. 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

[22] I. Kaate, J. Salminen, S.-G. Jung, T. T. T. Xuan, E. Häyhänen, J. Y. Azem, and B. J. Jansen. 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* (2025). 1624–1638. doi:10.1145/3708359.3712160

[23] Douwe Kiela, Max Bartolo, Yixin Nie, Divyansh Kaushik, Atticus Geiger, Zhengxuan Wu, Bertie Vidgen, Grusha Prasad, Amanpreet Singh, and Pratik Ringshia. 2021. Dynabench: Rethinking benchmarking in NLP. In *Proceedings of the 2021 conference of the North American chapter of the Association for Computational Linguistics: human language technologies*. 4110–4124. https://aclanthology.org/2021.naacl-main.324/

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

[25] Igor Matias, Farhat-ul-Ain, Darina Akhmetzyanova, and Vladimir Tomberg. 2024. Modelling Users for User Modelling: Dynamic Personas for Improved Personalisation in Digital Behaviour Change. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*. ACM, Cagliari Italy, 445–451. doi:10.1145/3631700.3665241

[26] Jennifer (Jen) McGinn and Nalini Kotamraju. 2008. Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008). ACM Press, Florence, Italy, 1521. doi:10.1145/1357054.1357292

[27] T. Mijač, M. Jadrić, and M. Čukušić. 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)* (2018). IEEE, 1237–1242. doi:10.23919/MIPRO.2018.8400224

[28] Preslav Nakov, Giovanni Da San Martino, Tamer Elsayed, Alberto Barrón-Cedeño, Rubén Míguez, Shaden Shaar, Firoj Alam, Fatima Haouari, Maram Hasanain, Watheq Mansour, Bayan Hamdan, Zien Sheikh Ali, Nikolay Babulkov, Alex Nikolov, Gautam Kishore Shahi, Julia Maria Struß, Thomas Mandl, Mucahid Kutlu, and Yavuz Selim Kartal. 2021. Overview of the CLEF–2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, K. Selçuk Candan, Bogdan Ionescu, Lorraine Goeuriot, Birger Larsen, Henning Müller, Alexis Joly, Maria Maistro, Florina Piroi, Guglielmo Faggioli, and Nicola Ferro (Eds.). Vol. 12880. Springer International Publishing, Cham, 264–291. Series Title: Lecture Notes in Computer Science. doi:10.1007/978-3-030-85251-1_19

[29] Lene Nielsen. 2019. *Personas - User Focused Design* (2nd ed. 2019 edition ed.). Springer, New York, NY.

[30] Lene Nielsen, Kira Storgaard Hansen, Jan Stage, and Jane Billestrup. 2015. A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations. *Int. J. Sociotechnology Knowl. Dev.* 7, 1 (Jan. 2015), 45–61. doi:10.4018/ijskd.2015010104

[31] Vinicius Carvalho Pereira and Cristiano Maciel. 2024. Decolonizing and peripheralizing e-lit in Latin America: the case of the Acervo de Literatura Digital Mato-Grossense, Brazil. In *Proceedings of the 35th ACM Conference on Hypertext and Social Media (HT '24)*. Association for Computing Machinery, New York, NY, USA, 113–125. doi:10.1145/3648188.3675131

[32] M. Prpa, G. Troiano, M. Wood, and Y. Coady. 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* (2024) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3613905.3636293

[33] Giulia Renda, Marilena Daquino, and Valentina Presutti. 2023. Melody: A Platform for Linked Open Data Visualisation and Curated Storytelling. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media (HT '23)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3603163.3609035

[34] Joni Salminen, Kathleen Guan, Soon-gyo Jung, Shammur A. 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* (2020). 1–14.

[35] Joni Salminen, Kathleen Guan, Soon-Gyo Jung, and Bernard J. Jansen. 2021. A Survey of 15 Years of Data-Driven Persona Development. *International Journal of Human–Computer Interaction* (2021), 1–24. ISBN: 1044-7318 Publisher: Taylor & Francis.

[36] Joni Salminen, Soon-gyo Jung, Hind Almerekhi, Erik Cambria, and Bernard J. Jansen. 2023. How Can Natural Language Processing and Generative AI Address Grand Challenges of Quantitative User Personas?. In *HCI International 2023 – Late Breaking Papers* (2023) *(Lecture Notes in Computer Science, Vol. 14059)*, Helmut Degen, Stavroula Ntoa, and Abbas Moallem (Eds.). Springer Nature Switzerland, 211–231. doi:10.1007/978-3-031-48057-7_14

[37] Joni Salminen, Soon-Gyo Jung, and Bernard J. Jansen. 2022. Developing Persona Analytics Towards Persona Science. In *27th International Conference on Intelligent User Interfaces* (2022). ACM, 323–344. doi:10.1145/3490099.3511144

[38] J. Salminen, C. Liu, W. Pian, J. Chi, E. Häyhänen, and B. J. Jansen. 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* (2024) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3613904.3642036

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

[40] J. Thiyagalingam, M. Shankar, G. Fox, and T. Hey. 2022. Scientific machine learning benchmarks. *Nature Reviews Physics* 4, 6 (2022), 413–420.

[41] A. Wang, Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. Bowman. 2019. SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems. In *Advances in Neural Information Processing Systems* (2019), Vol. 32. https://proceedings.neurips.cc/paper/2019/hash/4496bf24afe7fab6f046bf4923da8de6-Abstract.html

[42] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (2018), Tal Linzen, Grzegorz Chrupała, and Afra Alishahi (Eds.). Association for Computational Linguistics, Brussels, Belgium, 353–355. doi:10.18653/v1/W18-5446

[43] 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 Computer Society, Los Alamitos, CA, USA, 353–354. doi:10.1109/RE57278.2023.00048