



Vaasan yliopisto
UNIVERSITY OF VAASA

OSUVA Open
Science

This is a self-archived – parallel published version of this article in the publication archive of the University of Vaasa. It might differ from the original.

Developing Persona Analytics Towards Persona Science

Author(s): Salminen, Joni; Jung, Soon-Gyo; Jansen, Bernard

Title: Developing Persona Analytics Towards Persona Science

Year: 2022

Version: Accepted manuscript

Copyright © Authors | ACM 2022. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in Proceedings of IUI '22: 27th International Conference on Intelligent User Interfaces, <http://dx.doi.org/10.1145/3490099.3511144>.

Please cite the original version:

Salminen, J., Jung, S-G. & Jansen, B. (2022). Developing Persona Analytics Towards Persona Science. *Proceedings of IUI '22: 27th International Conference on Intelligent User Interfaces*, 323-344. New York: Association for Computing Machinery.
<https://doi.org/10.1145/3512891>

Developing Persona Analytics Towards Persona Science

Joni Salminen

Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar; and University of Vaasa, Vaasa, Finland, joni.salminen@uwasa.fi

Soon-gyo Jung

Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar, sjung@hbku.edu.qa

Bernard J. Jansen

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

Much of the reported work on personas suffers from the lack of empirical evidence. To address this issue, we introduce Persona Analytics (PA), a system that tracks how users interact with data-driven personas. PA captures users' mouse and gaze behavior to measure users' interaction with algorithmically generated personas and use of system features for an interactive persona system. Measuring these activities grants an understanding of the behaviors of a persona user, required for quantitative measurement of persona use to obtain scientifically valid evidence. Conducting a study with 144 participants, we demonstrate how PA can be deployed for remote user studies during exceptional times when physical user studies are difficult, if not impossible.

Keywords

Personas, User research, Persona science, Persona analytics, Remote user studies

1 Introduction

The current work discusses how personas can be effectively combined with the concept of analytics, i.e., the use of end-user data for drawing insights into human factors [53]. We start by defining the key concepts, and we then explain our approach to infusing personas with analytics, which we denote as **Persona Analytics (PA)**. We refer to **users** when we mean stakeholders that *use* personas for decision making (e.g., designers, software developers, marketers). For other terminology, we refer to **end-users** when we mean people on whose information personas are based.

Personas, commonly applied in human-computer interaction (HCI) [16], design [8], and business domains such as marketing and sales [74], are fictional depictions of end-users, patients, customers, or other groups of interest [16]. Personas convey end-users' needs and requirements [8], alleviate decision-makers' self-referential bias [6], and enable thinking of end-users even when none are physically present [66]. Also, personas give a human face to analytics data [29,35], humanize segments [11], give design inspiration [57], help compare end-users [34], and facilitate prioritizing end-user needs [69] for system development [67]. Persona profiles show relevant information about end-users [58] (see Figure 1). Nielsen summarizes a good deal of personas literature on their creation, assessment, and use [57]. Personas are easily digestible snapshots of end-users, audiences, or customers for use throughout an organization.



Figure 1: Example of a persona profile. Persona Analytics enables tracking users’ mouse and gaze interactions as the user engages with different information elements (A: Picture, name, and text description, B: Audience size, C: Sentiment, D: Social media quotes, E: Topics of interest, and F: Most viewed contents).

Personas are increasingly being enriched with quantitative data [72], or their creation is partially or completely carried out by algorithmic processes, which is referred to as *algorithmically generated persona development* [50]. When quantitative data becomes part of the created persona profiles, these profiles start approaching other analytics systems in terms of producing end-user metrics. In fact, algorithmically generated personas can be seen as an alternative method of end-user understanding, as alternatives to UX analytics tools (e.g., Google Analytics, Adobe Analytics, HubSpot, Mixpanel, Crazy Egg, and so on), when it comes to design tasks that require user insights (e.g., understanding website visitors for improving usability). As such, personas personify the numerical data on end-user characteristics and behaviors—*turning numerical reports into persona profiles* [29].

This transformation from “cold numbers” into “warm people” has been denoted as a benefit of personas, in that personified end-user data is treated more empathetically than nameless, faceless numbers [74]. Algorithmically generated personas are also supported by the automation of data science pipelines—i.e., the process of persona creation can now be automated—as well as web technologies [52] that enable serving the personas to users via web browsers—i.e., via *interactive persona systems* [33]. As such, PA can be defined as follows:

DEFINITION 1: *Persona analytics refers to decision-makers (i.e., persona users) in organizations using personas as analytical tools to better understand their end-users or other groups of interest.*

The above definition follows the conventional understanding of algorithmically generated personas using quantitative data [29,30,52]. As mentioned, we define a ‘user’ as someone who uses a persona for a professional task, which can relate to software development, design, marketing, or any other domain where personas are applied. Therefore, users can be software developers, designers, marketers, or other stakeholders involved in user-centric decision making. Because of this connection between personas and users, there exists another aspect to the concept of PA. Namely, PA can be seen as a research instrument to generate lasting knowledge about personas and how users interact with them. Its role is to grant HCI researchers a systematic approach for collecting data about persona user behavior and metrics for dealing with this data. In so doing, PA paves the way to more effective application of *persona science*, defined as ‘the use of empirical scientific methods, such as experiments, to produce robust and generalizable information about persona creation, evaluation, use, and impact’.

Indeed, in prior research [36,37], PA is defined as the systematic measurement of behaviors and interactions of persona users engaged with interactive persona systems. This is consistent with the second definition we put forth:

DEFINITION 2: *Persona analytics refers to how researchers investigate the behaviors of persona users.*

It is this second definition that motivates our current work, because examining persona users' engagement with personas can generate vital insights for *persona science* and the *design of personas* and persona systems that better serve stakeholders' information needs about end-users or customers. The persona research urgently needs a strong empirical orientation to produce knowledge that is believable and can truly push forward the boundaries of personas practice and theory and add up to a coherent understanding of the persona user. Advocates of the scientific method in persona research [4,5,11,12,23] have continuously mentioned the lack of empirical experiments and quantitative measurements as a bottleneck for progress in terms of theory and practice.

Our definition of persona science implies not only collecting data and conducting research on personas but also making an effort to devise theories that explain the data and guide further data collection. Persona science deals with real user behavior and formulating theories that are relevant to the design of personas. The focus in these efforts lies in the study of the persona users, which we demonstrate in this work by introducing new tools for measuring persona user behavior. To this end, the current work concerns itself with the development of a novel PA system embedded within a persona system, with the purpose of more effectively investigating/researching the behavior of persona users. Three research questions (RQs) are posed:

- **RQ1:** *How to implement PA in an interactive persona system?*
- **RQ2:** *What kind of research questions can PA address?*
- **RQ3:** *How can PA be used for understanding persona user behavior?*

The goal of the current work is to report efforts of building analytics features into an interactive persona system. We demonstrate the capabilities of this PA system and discuss its value for empirical persona research. In practice, PA can assist in designing layouts, features, and information content in algorithmically generated persona profiles. To achieve these benefits, it is necessary to incorporate analytics into personas, so that the interaction between the users and personas (and interaction features) can be captured. Previous efforts of this work appeared in [36,37] – relative to these, the current work adds a full-scale case study demonstrating the system capabilities with a real user study (previous research only tested the system with one pilot user). The PA system has implications for researchers and practitioners who are increasingly adopting web-based tools for remote testing since social distancing hampers in-person user studies [17]. This trend is likely to continue as tools and practices for remote user studies evolve.

2 Related Work

Algorithmically generated Personas. Although quantitative personas were first created within software requirements engineering [7,8], the concept of personas being data-driven was introduced by McGinn and Kotamraju [50] and later deployed by others [40,41,51,95,97]. Though, the idea of using “data” for personas dates to Cooper’s [16] concept that personas should be based on real user goals instead of fiction. While data orientation has remained a consistent theme in the persona literature [16,17,32,9,10,50,51], three trends contribute to the rise of algorithmically generated personas [29,72]: (1) *availability of user and customer data* from online analytics and social media platforms; (2) *democratization of data science tools and algorithms* that enable automated persona generation; and (3) *web technologies* that remove the limitations of static personas via interactive user interfaces. These trends denote a shift from unchanging “flat file” personas into dynamic “full-stack personas” that update automatically and are traceable to individual user-level data [29].

Interactive Persona Systems. From algorithmically generated personas, the next logical step of evolution is *interactive persona systems* [3,52,72], defined as interactive user interfaces (UI) that display persona profiles. This UI can, but not necessarily always, be accessed via web browsers [32,33,35,38]. The benefits of web technologies

are their broad applicability and accessibility. Personas served via the web can be accessed virtually from anywhere using any device that supports web browsing (see [Figure 2a](#)). Supporting technologies, such as user account management, can be integrated with relative ease using standard libraries and best practices. *Interactivity* refers to users performing various actions on the personas, such as analyzing information on gender distributions, refreshing the persona quotes, filtering the quotes by sentiment and topic ([80](#)), predicting a persona’s interest for a given topic ([2,3](#)), and engaging in dialogue ([45,47](#)). The interactive features are enabled by standard Web technologies, such as HTML, CSS, and JavaScript.

Emerging opportunities in Literature. Following these developments in algorithmically generated personas and interactive persona systems that have been described as transformational ([52](#)), multiple opportunities can be envisioned. We highlight five such opportunities. First, (i) interaction techniques and multimedia (e.g., persona chat/dialogue systems ([13](#)), video, AI agents ([87](#))...) could be incorporated into persona systems to serve various end-user needs ([75](#)). Second, (ii) new features for comparing personas by design goal metrics, such as diversity ([74](#)) and inclusivity ([21](#)), could be added. Third, (iii) personas could be integrated into an external system to enable persona-based recommendations ([46](#)), content management, and customer relationship management, as well as facilitating online advertising ([79](#)) via application programming interfaces (APIs) ([38](#)). Fourth, (iv) developers could provide explainability, transparency, and context, which are important when applying algorithms for persona creation ([80,88](#)), as illustrated in [Figure 2b](#). Finally, (v) interactive systems can be used to drill down to the persona information and make quantitative predictions ([3](#)).



Figure 2: (a) Interactive persona features, such as [A] browsing the available personas, [B] searching and sorting by user-defined criteria, [C] explanatory tooltips, and [D] export of usage logs. (b) Adding transparency to algorithmically generated personas. The first layer shows Mamdouh, a young Egyptian. The second information layer, accessible by clicking a chart icon, shows that Mamdouh actually comprises many demographic groups, of which [Egypt, 25-34 Male] is deemed the most representative by the algorithm

Research Gap. While technology introduces novel opportunities for user-to-persona interaction, at the same time, these trends create an opportunity for better understanding of how persona users, such as designers, software developers, and marketers interact with personas. This better understanding of persona user behavior can lead to substantial advances in *persona science* (i.e., the academic study of personas and their usage), but it requires effective implementation of measurement. The lack of empirical persona user research has been noted

by several researchers [48,72,77]. The unifying factor behind these possibilities is the need for understanding the persona user behavior, which requires measurement. In our solution, this measurement capability is provided by PA.

3 Methodology for Persona Analytics

3.1 Requirements Journey

There is no standard method of building an analytics system. In our case, all the researchers had extensive experience of both Web analytics and personas research, which was instrumental in this process. This experience consisted of working with industry-leading analytics solutions, such as Google Analytics, for more than a decade in the case of two authors and half a decade in the case of one researcher. The persona research of the authors, when combined, also extends well beyond a decade and mostly consists of empirical work. Therefore, we had a vision of what we wanted to accomplish, what is missing from current research and practice, and what research questions in persona science should be addressed via empirical data.

We started out by “drinking our own Kool-Aid,” i.e., by defining the ideal user persona for the system. This “persona” of a user of the PA system is a *researcher* that wants to conduct persona user studies in order to address scientifically important questions. Measuring user behaviors helps researchers tackle open research questions in persona science, which is the goal of this persona. To support this persona, a few requirements are posed: (a) the data must be accurate so that proper conclusion can be drawn from it, (b) there should be the possibility to include several data types to enable the comparison of different end-user inputs, and (c) the dimensionality of the data should not be overwhelming for the analysis task, i.e., the data needs to be exportable in a format that is relatively easy to analyze.

These desiderata were considered in the design of PA by incorporating multiple data sources and by keeping the reporting data granularity at a user-friendly level – in other words, reports with different levels of detail and aggregation are provided, as explained later in this manuscript.

Second, we brainstormed the type of questions that the PA system would need to be able to address by providing data for the researcher persona. The following list of scientific questions of interest (SQ) was collaboratively obtained among the research team members, pertaining to various persona aspects (in parentheses):

- **SQa:** How do users interact with persona profiles? (interaction techniques)
- **SQb:** What interactive features facilitate users’ discovery of personas for a given task? (interaction techniques)
- **SQc:** What information of personas do users pay attention to? (information design)
- **SQd:** What persona information influences users’ design choices and how? (information design)
- **SQe:** What persona information influence users’ behaviors or attitudes about end-users? (persona perceptions)
- **SQf:** How do users compare personas for a task? (cognitive styles, information processing)
- **SQg:** How and why do users choose a persona for a given task? (cognitive styles, information processing)
- **SQh:** How do users or user groups differ by their persona use? For example, are persona users with less experience in personas using them differently? Are there gender differences? (demographic, cultural, and social factors)

Addressing these and other vital questions can provide much needed direction for persona science, addressing aspects of persona creation, validation, use, and value in use. Empirical, scientific inquiry is not only needed to produce valid knowledge for practitioners using personas, but it is also required to create robust *theories* on personas and their users. Aligned with principles of scientific inquiry, persona research can benefit from adopting more rigorous research designs, including hypothesis formulation based on theories in HCI, information science, social psychology, and other fields tangential to personas; and followed by systematic testing of those hypotheses,

then revising the theory to adapt to persona context. Addressing these questions can help persona creators understand which aspects of human-to-human interactions apply to human-to-persona interactions, so design decisions can be made to mitigate unwanted effects (e.g., stereotyping [48] and seeing personas as irrelevant, abstract, or misleading [49]) as much as possible.

3.2 Defining Metrics for Persona User Behavior

In practice, to address these questions, the PA system needs to track various measures and use these measures to compute metrics. Therefore, we needed to devise PA system metrics. These metrics were defined based on their ability to address the types of questions posed earlier. These metrics can be divided into (a) *persona-based metrics* and (b) *user-based metrics*. The persona-based metrics include, for example, the following (with potential use mentioned after definition):

- **Time spent per persona:** the duration users interacted with a given persona. *Purpose:* Proxy measure for users' interest – it is likely that users spend more time with personas they find more interesting.
- **Number of visits per persona:** the number of a given persona was visited by the users. *Purpose:* Proxy measure for users' interest – it is likely that users visit personas they find interesting more often.
- **Persona bi- and trigrams:** the number of times users visited specific two or three personas during a session or time period. *Purpose:* Bi- and trigrams can be indicative of comparative behavior, i.e., how users compare personas. (Technically, this is not a metric but a measure; however, we mention it here due to its nature of being computed.)

In other words, these metrics communicate aggregate information about how one persona did relative to another – i.e., was one more popular than another, in what order where they visited, and so on. User-based metrics, in turn, communicate about a user or a group of users. These include:

- **Number of personas visited:** the number of personas a user visited during a session. *Purpose:* to understand how thoroughly a user viewed the personas. For example, a user that only visits a small number of personas either quickly found what they were looking for, satisfied with the “first acceptable choice” [93], or was not engaged with the system and/or personas.
- **Persona coverage:** the relative share of personas a user visited out of the personas available. *Purpose:* The same as previous, but as a ratio metric of the visited personas / the number of available personas. The higher the number of personas becomes, the more likely it is that the persona coverage per user decreases, as users would be unlikely to browse a very high number of personas for their professional tasks.
- **Average visit duration:** the average time spent per persona for a given user. *Purpose:* can reveal if the user was more or less engaged relative to other users.
- **Persona rank correlation:** the degree to which the order of a user visiting the personas corresponded with the personas' order of presentation in the system (can be computed based on visit duration as well). *Purpose:* to test if there are order effects that affect persona use.

Many of these metrics are inspired by similar metrics used in information theory [92], eye-tracking studies [25,42,71], and Web analytics [1,14,31]. While similar metrics are well established in said fields, these are metrics are *not* established in persona science and research. In fact, we are aware of no previous study that discusses metrics for persona user behavior – again, this hampers scientific progress in this domain. Coupled with participant data, the metrics can help analyze how users view different personas, if there is selection bias based on demographic factors of personas and their users, and so on. While these basic metrics provide a useful starting point for persona science, a lot more development in this domain is needed. It is also important to modify and adapt known metrics for the persona context, because they could be computed or interpreted differently when studying personas. We discuss this matter later in [Section 6.8](#).

3.3 Determining the Data Collection Modes

The data collection modes were largely pre-determined by what is possible using the current Web technologies. Mouse-tracking is the obvious choice due to its commonality in online analytics and support provided by all Web browsers [39]. The advantages of mouse-tracking are three-fold: it (i) offers an unobtrusive form of tracking of natural user behavior, (ii) does not require calibration, and (iii) has perfect accuracy—i.e., there is virtually no measurement error, but the users' movement of the mouse is perfectly traceable to specific pixels and UI elements. On the negative side, mouse-tracking is considered a weaker proxy for attention than gaze movement, i.e., eye-tracking [9,55], mainly because users might not always move their mouse when processing information on the screen. As processing of persona information requires eye-sight, eye-tracking is a useful data source to complement mouse-tracking in interactive systems [18]. The challenge of webcam-based tracking is that error margins can pose challenges for data quality, as there are differences in terms of hardware quality, lighting conditions, distance to screen and device, and a myriad of other conditions that can decrease online eye-tracking data quality [98]. While these issues do concern both separate hardware trackers (e.g., Tobii, MyGaze, GazePoint) and webcam-based eye-tracking, for the latter, the challenges in data quality are much higher because webcams do not provide access to infrared frequencies that the professional trackers use.

Therefore, because mouse- and eye-tracking in a remote user study context each involve their unique advantages and challenges, it is appropriate to integrate both data collection modes into the PA system, which simultaneously completes the scope of the requirements. Figure 3 offers a conceptual overview of the PA system. Overall, the overall algorithmically generated persona process is as follows: end-users' data => personas => persona users (e.g., marketers) => persona users' data (collected via PA) => researchers and analysts (studying persona user behavior). In other words, personas serve the information needs of stakeholders, and PA serves the information needs of researchers interested in persona user behavior.



Figure 3: Conceptual diagram of Persona Analytics. Multiple users can simultaneously interact with the persona system. The user's interaction with persona profiles is captured via mouse- and eye-tracking, recorded in a central database (DB), and outputted via reporting interface. By analyzing the reports, researchers can make important discoveries for persona science.

4 System Implementation

4.1 Overview

The defined questions acted as a guiding idea for requirements and implementation. In software development projects, requirements detail what is needed from a system [20]. Engineers or developers tend to implement features and functionalities according to the requirements to create the system. Two of the authors collaborated on creating the requirements, and one of the authors with the necessary skills implemented them. The system was tested internally and with a pilot user (reported in [37]), and we found it to log the data correctly.

The implementation of PA was carried out for an interactive system called Automatic Persona Generation (APG) [2,3], which is a state-of-the-art system for algorithmically generated persona development. The system is available at <https://persona.qcri.org>. While we defer the reader to related work for a complete description of APG's system functionalities and associated algorithms, the following subsection provides a brief explanation of the APG system.

4.2 Algorithmic Approach for Persona Generation

APG generates personas from online analytics data—e.g., from YouTube audience statistics or Google Analytics log data on end-users. APG infers demographically and behaviorally distinct patterns from user datasets [27,28]. APG has been previously applied to datasets on social media users [89], ad target groups [84], online news audiences [2,3], and video game players [90]. The APG persona creation relies on three main steps [30]: (a) identify unique user behavioral patterns using non-negative matrix factorization (NMF) [43], (b) associate these behavioral patterns with representative demographics (age, gender, country) to form “skeletal personas”, and (c) enrich the skeletal personas with personified information that matches the demographics (name, picture, job, education level, relationship status, topics of interest). Figure 4 summarizes the algorithmic persona generation process.

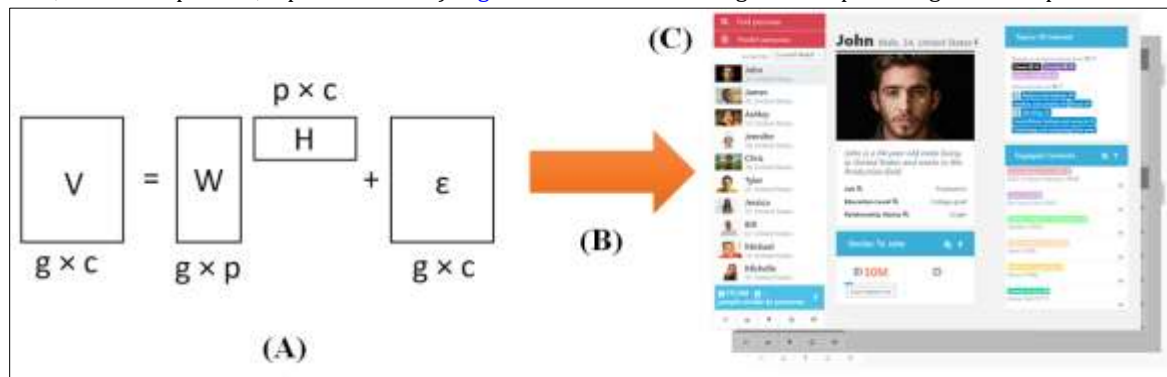


Figure 4: (A) applying the NMF algorithm [43] to the user dataset V that consists of demographic groups (g) and content (c). This matrix is decomposed to W and H , both involving the hyperparameter p that indicates the number of personas. Epsilon describes the error term. Through enrichment process (B), explained in the body text, APG produces a set of p personas (C) that have personified information, conceptually known as “personification of big data” [96].

4.3 Measurement Paradigm

The key insight going from APG (interactive persona system) to PA (persona analytics) is that, when personas are provided through a web browser, PA takes place via mouse- (and eye-)tracking that records the persona users’ mouse (or gaze) movements and clicks (eye fixations) on the persona profiles and their information elements.

This enables empirical persona research, such as building click paths, persona visit sequences, dwell time analyses, and so on. In PA, we track both the information usage behavior within the persona profiles and the transitions between the personas. One can think of personas as “pages” in the conventional Web analytics terms, and then information reports are the pages’ content. To answer research questions that advance persona science, we need both levels of tracking.

To this end, the PA system records the user’s mouse and gaze movements simultaneously during the session. The data is stored in a backend database. To support analysis, the screen coordinates are automatically converted to the corresponding information elements in the logs using JavaScript. In other words, the PA system records that a given mouse hovering or fixation was targeting, e.g., “Persona picture”. The duration of hover or fixation is calculated based on the “in” and “out” timestamps. In total, the PA system has 120 predefined HTML elements describing all information in the persona profile page (see [Figure 5](#)). The main elements include *Headline* (name, gender, age, country), *About* (picture, text description, job, education level, relationship status), *Sentiment*, *Topics of Interest*, *Viewed Conversations* (Quotes), *Viewed Contents*, and *Audience Size*, corresponding to typical information in persona templates [\[58\]](#).



Figure 5: Because interactive persona systems serve the personas via a web browser, web technologies, such as HTML and JavaScript, can be used for tracking how users interact with the personas. The frames in the figure illustrate how information elements in the persona profile are tagged for user tracking. In total, PA tracks 120 elements in the persona profile.

4.4 Online Eye-Tracking

The eye-tracking is implemented using WebGazer.js¹ [\[62–64\]](#), a webcam-based gaze tracker developed at Brown University. WebGazer is available in JavaScript as an open-source library². Multiple alternative frameworks were compared, but we chose WebGazer based on four reasons (a) it provides a relatively good accuracy based on our pilot testing, (b) is actively developed based on the update frequencies in the GitHub repository, (c) the source

¹ <https://webgazer.cs.brown.edu/>

² <https://github.com/brownhci/WebGazer>

code is publicly available and can be integrated into systems such as APG, and (d) the software is provided free of charge. These properties make WebGazer a feasible online eye-tracker for research-based systems such as APG.

4.5 Administrative Features

In the APG's UI, system administrators can *enable* either mouse tracking, eye tracking, or both for all users or for a subset of users. From a user's point of view, the only difference is that, when eye-tracking is enabled, every session starts with calibration (see [Figure 6a](#) and [b](#)). The PA system processes mouse- and eye-tracking data identically, which means that both types of interaction are recorded in the database and can be exported in a single file in order to improve usability for the researcher using the PA system. When the coordinates of hovering or gazing correspond to a predefined element, this event, along with timestamp (in/out) and meta-data (User and Session ID), is sent by the client browser via Ajax (Asynchronous JavaScript and XML) to the backend database. The PA system maps the coordinates to the persona information element the user is interacting with. Administrators can download the log files for data analysis (see [Figure 6d](#)). They can also create new user studies from the persona system's backend.

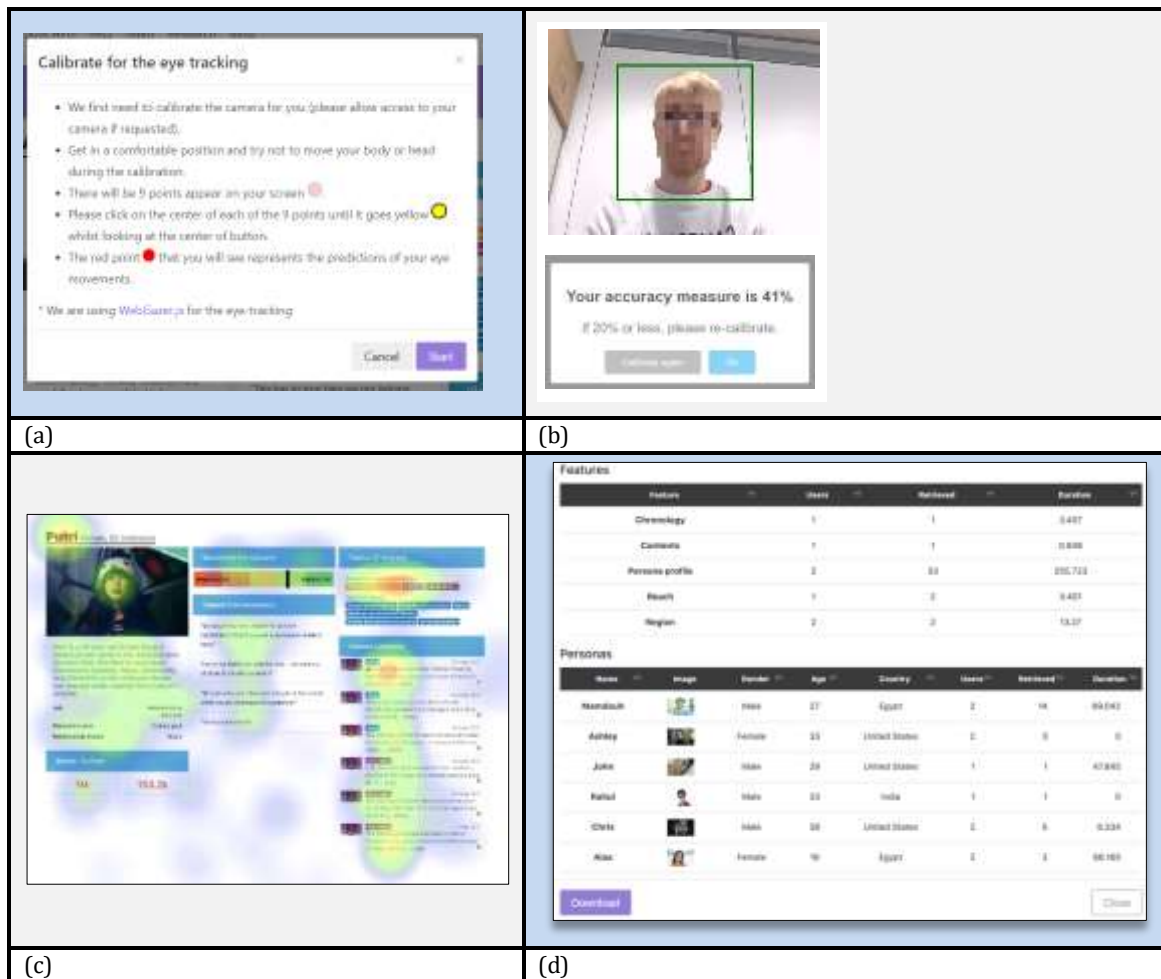


Figure 6: (a) Eye-tracking calibration dialogue and (c) how it shows to users; (c) example of a user’s eye-tracking pattern before the data is converted to each specific information element in the persona profile (denser color indicates more gaze fixations in a given area), and (d) data export dialogue shown for researchers to export user logs. The logs are provided in CSV files which can be downloaded via the Download button.

To prepare the data exports, the system uses Pandas (i.e., Python Data Analysis Library) after retrieving the logs from the backend database. It computes the duration of each interaction based on “in” and “out” timestamps and generates a comprehensive data report, in which each mouse and eye fixation event, its timestamp, its target information element in the persona profile, and meta-information (Session ID and User ID), are saved into a file. The logs can be downloaded for further analysis. Information about the variables logged by the PA is included in Supplementary Material³. These variables were determined based on the metrics and questions detailed in the previous sections. As a result, the data recorded by the PA system enables the calculation of various metrics of persona user behavior. We now illustrate, though an example user study, how PA can serve persona researchers.

³ <https://www.dropbox.com/s/yeu1jrohs6lbpg8/central%20variables.docx?dl=0>

5 Validation Study

A remote user study was conducted in which 114 participants used an interactive persona system with PA enabled, to browse a set of 10 personas, created from a dataset of a tourism-promoting organization with 1.8M (1,795,115) user likes over 5,312 Instagram posts. The participants' task was to choose a persona to target for tourism marketing (i.e., promoting a specific destination, in this case, a country). Basic demographics of participants are provided in [Table 1](#).

Table 1: Participant demographics.

Age			Male	Female	Non-binary	N = 118 (81.9%*)
M	SD		62	55	1	
35.35	9.08		52.5%	46.6%	0.01%	

*For 18.1% of the participants, we did not have demographic information.

The participants were recruited using an online data collection service called *Prolific* [61]; the same service has been used in various other persona user studies [82,83,88]. We used the platform's industry categories as a sampling criterion, including "Art/Design", "Graphic Design", and "Market Research", in order to reach people that work in industries where personas are relevant. Students were excluded. All participants were provided with a definition of personas, and a task description prior to their use of the system. The study flow is illustrated in [Figure 7](#). The study dealt with testing the effect of simple and complex explanations on user behavior and perceptions, served via a product walk-through (i.e., a process that sequentially shows different parts of the system to a user that is logging in for the first time). The participants were directed from the online service to the system, which randomly allocated each participant into one of the three experimental conditions. The participants were randomly assigned by the APG system to one of the three experimental conditions (simple explanations about the system, complex explanations, and no explanations at all).

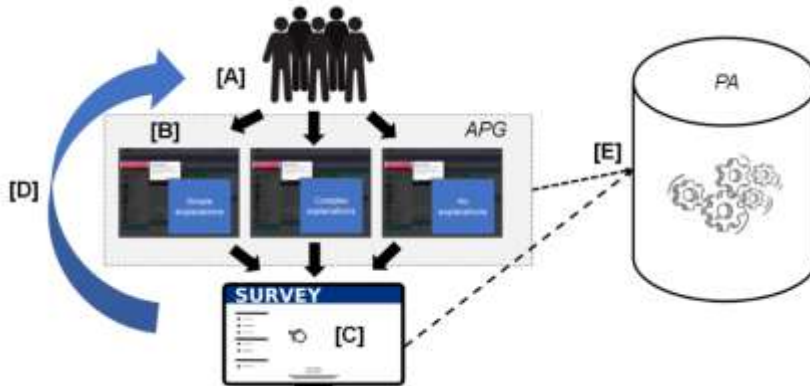


Figure 7: An example of how the APG and PA systems can be synchronized to conduct a full remote user study. In this case, we used an online platform to recruit participants to the study [A]. After using the system to browse the personas [B] (no time limit was imposed), the participants completed a survey [C] and were redirected back to the data collection platform [D] that logged successful study completions. PA recorded both the survey and behavioral data [E].

After using the persona system, which consistent viewing as many personas as they wanted out of the ones created by the algorithm (see [Table 2](#)) for as long as they wanted, the users could click on a banner to indicate they are ready to complete their task, after which the participant is transferred to a survey platform that collects

data about their task completion, perceptions, and demographic variables. Upon completing the survey, the participant is automatically redirected back to the online data collection platform, where their participation is marked complete. Both APG and the survey platform record the UTM parameters⁴ that identify (in an anonymous way) the participant so that participants that pass the data validation stage (i.e., researchers validating that their responses were genuine) can be easily compensated, and their system usage data is linked with their survey responses for further analysis. During system usage, PA was enabled, and data was collected on the users' interactions with the persona system.

Table 2: Personas the algorithm created for the user study.

Persona	Age	Gender	Country
Mamdouh	28	Male	Egypt
Rahul	34	Male	India
Ashley	25	Female	United States
Muhammad	34	Male	Pakistan
Alaa	18	Female	Egypt
John	26	Male	United States
Abdalaziz	23	Male	Egypt
Rizky	18	Male	Indonesia
Putri	20	Female	Indonesia
Chris	40	Male	United States

6 Explorations of Persona User Behavior

6.1 Overview

In this section, we analyze the collected data to demonstrate how PA can serve empirical persona user research. We do not explicitly test any hypotheses, although the data obtained from PA could be used for that, but for the sake of demonstration, we inductively analyze the data and provide exploratory findings about persona user behavior. We then synthesize these findings in the form of propositions that future research could test, with complementary theorization, as hypotheses. In other words, we illustrate how PA can be of service towards persona science.

The results of the effect of the three conditions will be reported in a future publication; here, we focus on demonstrating how PA can be used to analyze the data obtained from user experiments. For parsimony, the following analyses focus on the mouse-tracking data. Because the eye-tracking data is logged in the precisely same data structure as the mouse-tracking data, the exact same analyses and metrics can be obtained from eye-tracking. Based on our piloting of the eye-tracking module, the accuracy strongly varies (from ~16% to ~80% in our testing) by the user, condition, and equipment. This is also why it is more reliable to carry out this demonstration with the mouse-tracking data.

6.2 Descriptive Statistics

Descriptive statistics about participants' engagement with the system (see [Table 3](#)) indicate that, on average, participants spent around 8.5 minutes browsing the personas for their task and visited persona profiles on average 14 times. What is striking is the high dispersion among the participants – the standard deviations are high for both the dwell time (SD = 8.3 minutes) and visit counts (SD = 11). The participant with the shortest dwell time only used the system for 20 seconds, while the participant with the longest dwell time used the system for

⁴ Urchin Tracking Module parameters are a standard technique for tracking source and meta-data of Web traffic.

more than an hour (62.6 minutes). The shortest visit path only included visiting one persona, whereas the longest path consisted of visiting the persona profiles 60 times, which equals 6 visits per profile on average. These results indicate a major dispersion in engagement, with some participants being “persona power users” while others lack significant engagement with the system. Future analyses could investigate how these two extreme user types differ (e.g., demographic or industry variables that might explain the differences) and why (e.g., low task motivation, not perceiving personas as relevant or useful).

Table 3: Dwell time (in seconds) and number of visits to the persona profiles by the participants.

	Mean	SD	Min	Max	Median
dwell time (i.e., system usage time)	514.3	496.8	19.5	3756.8	410.8
visits (i.e., number of times loading a persona profile)	14.0	11.0	1	60	12

As a whole, exactly half (50.0%) of the users viewed at least 9 of the personas (see Figure 8a), i.e., achieving a persona coverage of 90%, while less than a third (31.9%) viewed three or fewer personas. The fact that close to half (47.2%) viewed all 10 personas implies that users have a need for viewing a variety of personas—10 was the highest number in this study, but it seems likely that given the choice, users would have viewed more than 10 personas. Concerning information viewing patterns (Figure 8b), the persona profiles contain 8 parent information elements (About, Audience Size, Headline, Sentiment, Timeline, Topics of Interest, Viewed Contents, and Viewed Conversations). Only a minority of persona visits contained viewing all 8 information elements (0.5%). While some visits included only viewing one information element (10.5%) – perhaps an indication of rapid verification of a recalled detail – more than half of the persona visits (52.3%) contained the viewing of at least four parent information elements (i.e., at least half of the main information in persona profiles). Section 6.5 investigates further what information was most viewed.

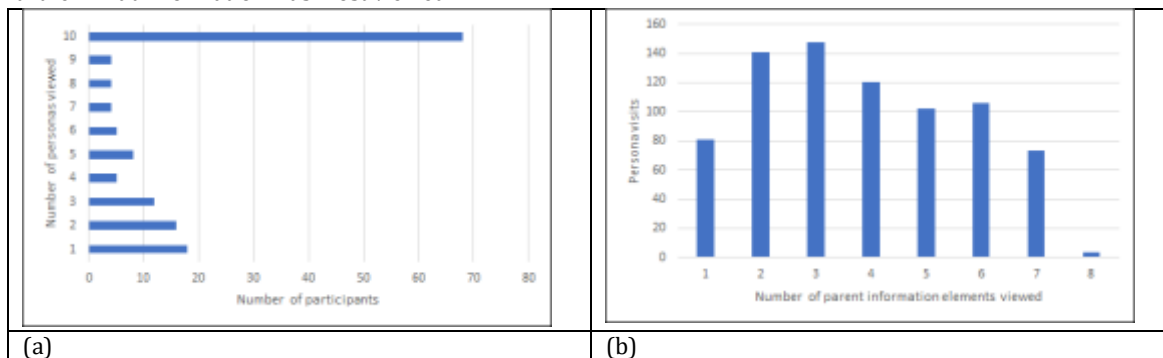


Figure 8: (a) Persona coverage (i.e., the number of personas a participant viewed during their whole sessions), (b) Information coverage (i.e., how many parent information elements were viewed by the users during each viewing of a persona).

6.3 Correlations and Gender Effects

There was no notable correlation between participant age and their system usage time ($r = 0.08$) or age and number of visits ($r = -0.16$). In terms of dwell times, results from a t-test indicate that females were using the system longer ($M = 634.8$ seconds) than males ($M = 493.4$ seconds), $t(114) = -1.55$, $p = 0.06$. In terms of visit count, there was no significance difference, with females ($M = 14.8$) and males ($M = 16.3$) visiting a roughly even number of personas, $t(114) = 0.79$, $p = 0.21$. (Both these tests were based on the 118 participants for which we

had gender information; there was one participant who indicated non-binary gender and who was therefore excluded from the analysis.) For males, with 95% confidence, the population mean for persona visit duration is between 33.4 and 59.4 seconds, based on 61 samples. For females, with 95% confidence, the population mean for persona visit duration is between 46.5 and 80.3, based on 55 samples. Finally, there is also no significant effect based on persona gender, with both male ($M=37.8$ seconds) and female ($M=35.3$ seconds) personas being frequented roughly an even amount of time, $t(1996) = 0.50, p = 0.31$.

6.4 Persona Viewing Patterns

One of the researchers conducted an exploratory data analysis (EDA) on 21 participants' patterns of viewing the personas - by pattern, we mean how long a participant viewed a persona in their sequence of browsing the personas. This EDA revealed several different patterns of viewing the personas (see [Figure 9](#)), including (a) shark fin, (b) u-shape, (c) stabilizing, (d) sporadic, (e) linear declining, and (f) triangle shapes.

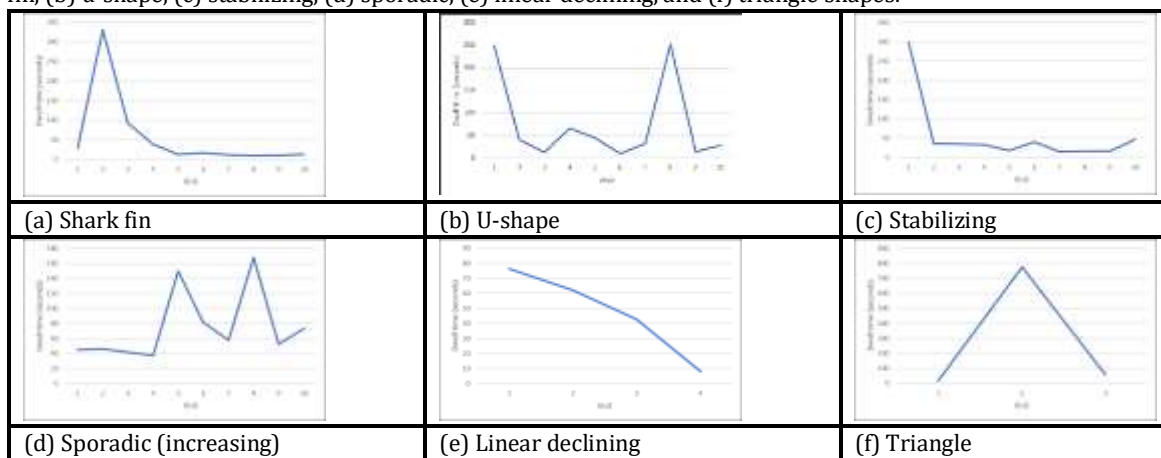


Figure 9: Different persona viewing duration patterns based on an exploratory analysis.

The variety of patterns indicates that it might be difficult to find general “laws” that would govern how individual users explore a set of personas. However, among the manually reviewed samples, we observed that the dwell times tend to decrease over the number of visits - 16 out of 21 had such a trend (76.2%) (e.g., *a* and *c* in [Figure 9](#)), while only three participants (14.3%) had an increasing dwell time trend (e.g., *d*). The two remaining had trends that could not be categorized as either decreasing or increasing (*b* and *f*). Thus, it appears that the time spent reviewing persona profiles decreases over the number of visits. In cases where the dwell time appears to “resurge” (e.g., *b*), the participant may be returning to a persona they found interesting earlier in order to verify, learn more, or compare information. Some patterns remain highly sporadic till the end of the session (e.g., *d*), while others seem to stabilize early (e.g., *c*).

6.5 Persona Information Viewing Behavior

We investigated where users focused their attention as a proxy for attention and interest. Results in [Figure 10a](#) show that the users were most interested in social media quotes in the persona profile (68.4% of the total dwell time), followed by the personas' basic information (“About”, 14.1%) and audience size (8.1%) that indicates how many people there are on Facebook and Twitter similar to the persona. Plotting the data shows that users' attention is unevenly distributed, with the quotes garnering over five times more dwell time than the second most popular information, i.e., the persona's basic information. Unlike for persona visits, dwell time and visit count is strongly correlated ($r = 0.63$) for persona information. It is known from previous persona studies that quotes are

very impactful for users' perceptions of personas [86], but it is interesting that the mouse-tracking shows the quotes overshadow other information this strongly.

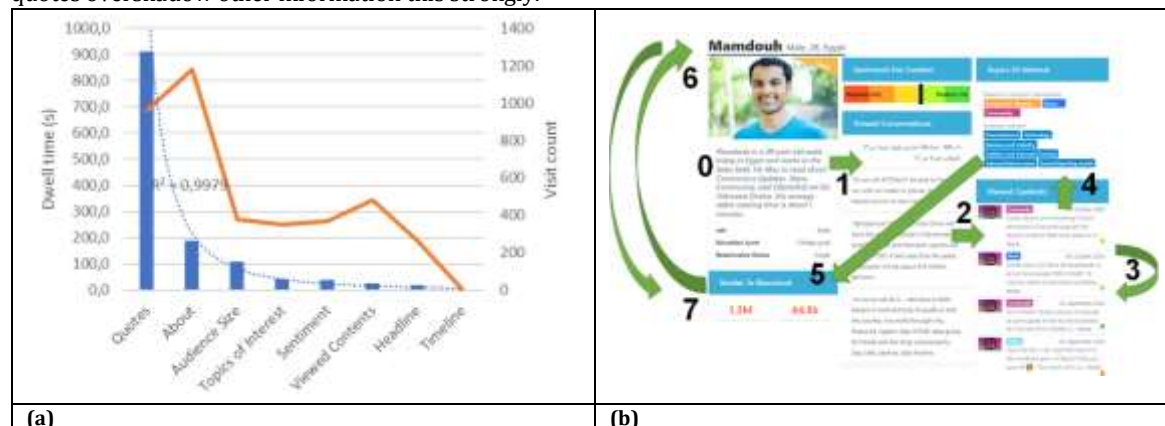


Figure 10: Information viewing behavior of the participants. (a) Dwell time (bars) and visit counts (line). (b) Most common transitions between the parent information elements across all personas (Mamdouh is used for illustration). Users start browsing the persona's basic information, including picture, text description and sociographics (State 0). They then move to comments (S1), most viewed content (S2), which is viewed repeatedly (S3), before moving to topics of interest (S4), audience size (S5), persona's name and demographics (S6), and back to audience size (S7).

On the other hand, if we instead of dwell time focus on the number of visits, personas' basic information ("About," as indicated by the orange line in Figure 10a) becomes the most important information element. This element contains the text description and picture of the persona that, again, previous research has been found influential for persona profiles [58]. The information viewing sequence (see Figure 10b) seems to move diagonally from top left to bottom right, then up, then bottom left, and back up and finally down ($\searrow \uparrow \swarrow \uparrow \downarrow$). This sequence was obtained by calculating the most common parent elements in states $S_0 \dots S_7$ across all participant-persona pairs.

Two takeaways can be elicited from these findings: (a) that personas' quotes, text description, and picture are among the most impactful information based on users' mouse engagement, and (b) measuring dwell time and visit counts can give different results, which is why measuring these two separately makes sense – an information element with a high visit count but low dwell time is frequented in short bursts, whereas an information element with a high dwell time is focused on for a longer time; it is logical that quotes interest people because they are seen to reflect the persona's attitudes and are information-rich for various user tasks.

6.6 Effect of Order of Personas

Order of presentation has been shown to affect how the information is accessed, used, and recalled [19]. Among the notable effects in this line of work are, e.g., the primacy effect that implies first seen information is the most impactful [68] and serial effect, implying that first and last items in a list are given special attention [54]. In persona system context, these effects can matter, because the personas are shown in a list, and such effects can therefore cause that, e.g., the needs of the first and last personas in the list would be considered more strongly than those of other personas.

When plotting the data to investigate, three observations can be made from Figure 11: (1) there is a first persona effect, i.e., the first shown persona (Mamdouh) gets substantially more attention than the others, (2) there is no strong pattern of primacy effect in terms of declining dwell time based on persona's order of being displayed in the system. However, (3) the fact that the last persona (Chris) is the second most viewed implies

serial effect in which the first and last items of a list garner the most attention. Spearman rank correlation between persona order and dwell time is negligible ($r = 0.176$). Correlation between system order and number of visits is moderate ($r = 0.576$). Correlation between dwell time and visits is also moderate ($r = 0.455$). However, when we compute the most common personas visited (i.e., S_1 = the first persona the user visits, S_2 = the second persona they visit, ... S_{10} = the tenth persona they visit), we find that the TOP-10 path is *precisely* identical to the order of presenting the personas in the system. A further check reveals that 31 users (21.5%) follow this sequence when using the system. That is, about one fifth of the users browse all the personas in the order in which they were presented.

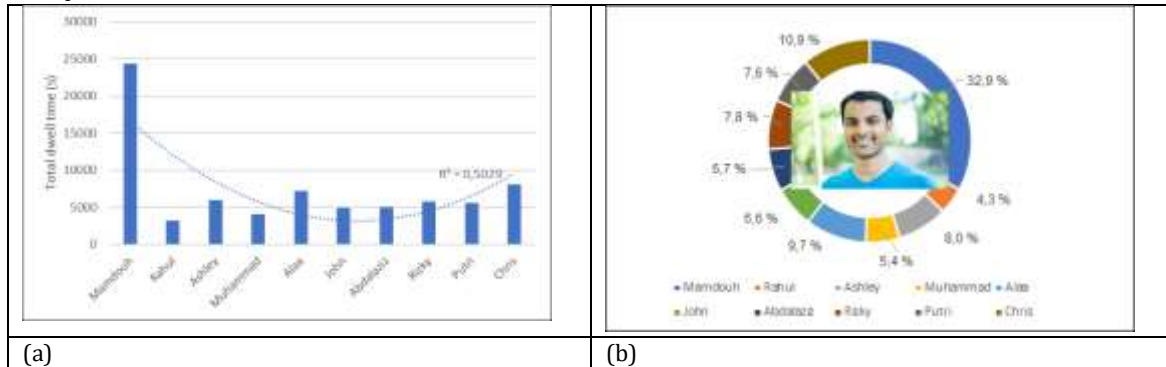


Figure 11: (a) Dwell time distribution among personas. (b) First persona effect. Mamdouh (the first persona shown in the system) received almost one third of all dwell time from the users. If dwell time was distributed evenly, he should only receive 10%, which means an excess of 229% from this equal baseline, some of which likely stems him being the default persona in the system.

Overall, these results indicate that (a) the system’s default persona garners the most attention, and (b) a sizeable portion of the users visit the personas in the exact order that the system shows them. Perhaps the system should evenly fluctuate the default position among the personas to mitigate for “discrimination” arising from this effect. The current logic is that the system loads the persona by default that has the highest audience representation, i.e., the most engagements in the baseline user data the personas are created from. Testing various rationale for the default persona and the effect of rotation are excellent ideas for future work. We also observed a potential cultural effect, which is a possible manifestation of the users better identifying with the personas from their own cultural sphere (see [Table 4](#)). Cultural aspects in personas remain an important area of future work, such as the special role of default and exit personas, i.e., those with whom users finish their browsing session (see [Figure 12](#)).

Table 4: Ethnic bias? The only three Western personas rank the highest in terms of average view time (apart from Mamdouh that is the default persona). 89.8% of the participants were from Western countries (Europe and United States). Users may feel more comfortable identifying with personas from their own culture and ethnicity. This also implies persona studies should employ culturally and ethnically diverse samples to obtain internationally valid results.

Persona	Avg time per visit
Mamdouh (rank = 1)	69.10
Chris (rank = 10)	37.78
John (rank = 6)	33.78
Ashley (rank = 3)	33.49
Alaa (rank = 5)	32.98



Figure 12: Exit personas. The number of times a user stopped their session after viewing a given persona. Because the task dealt with considering a specific persona, it is possible that the exit indicates a higher likelihood of the persona being chosen for the task.

6.7 Modeling Persona User Behavior

Because the PA system records the user’s transition from one persona information element to another (based on mouse and gaze movements), as well as capturing transitions from one persona to another, there are important opportunities for modeling user behavior, some of which we illustrate here. [Figure 13](#) describes these opportunities through the concept of *persona-gram*, which refers to a string of letters depicting a user’s path of visiting either the personas or the information elements within a persona profile. This information can be stored as a state transition matrix (see bottom of [Figure 13](#)) which can be further used for computing the probability of a user transition from one state to another.

In [Figure 13](#), the names on the left illustrate a user’s path of visiting the personas. On the right-hand side (User 1 = U_1), the same path is transformed into a string. The string format enables the comparison of different users using Levenshtein’s edit distance (ED) [73]. For example, User 2 (U_2) differs from User 1 (U_1) in only two string states (bolded in [Figure 13](#)), yielding $ED_{1,2} = 1$. Users that have a low edit distance are similar to each other in terms of their persona use behavior, whereas users with a high edit distance are behaviorally more different. When computing the distances of all users, it becomes possible to identify “average” behaviors and distinct outlier behaviors. (Moreover, it is possible to consider the duration of each visit to get a higher dimensional representation of the user’s viewing behavior.)

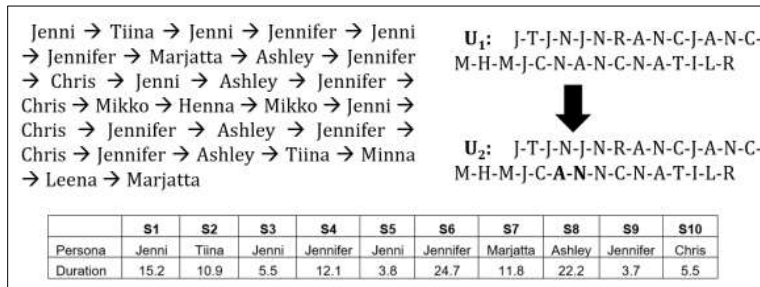


Figure 13: Illustration of persona-grams using imaginary data. The series on the left describes a user’s transition from one persona to another. User 1 on the right is the same sequence transformed into a string. User 2 has the same sequence except for two differences (A-N, bolded), which means the edit distance is 2, i.e., one needs to make two edits to make the strings identical. The fewer changes one needs to make, the more similar the sequence of viewing a person is between two users.

We computed the average edit distance across the dataset obtained and found the number highly dispersed. In other words, two users would rarely view the personas (or the information elements within the personas) in the same or similar order. This finding is interesting in itself – it implies users’ processing of persona information is more idiosyncratic than anticipated. To give a simple example, [Table 5](#) shows three randomly chosen participants that each have a path length of 10, i.e., they visited persona profiles 10 times during their session.

Table 5: Persona-grams of three randomly chosen users who each visited persona profiles 10 times. The color codes indicate the same persona being visited in the same sequence: yellow is shared by all three users, green is shared by Users 1 and 3, and turquoise by User 2 and 3. User 3 has a more similar browsing behavior with User 1 than with User 2, and User 1 and User 2 share the least similarity.

User 1	User 2	User 3
Mamdouh	Mamdouh	Mamdouh
Alaa	Rizky	Alaa
Putri	Chris	Putri
Alaa	Rahul	Alaa
Rizky	Alaa	Ashley
Abdalaziz	Muhammad	Chris
Alaa	Abdalaziz	Abdalaziz
Putri	Ashley	John
Rizky	John	Mamdouh
Alaa	Putri	Ashley

As can be seen from [Table 5](#), the behaviors are almost completely unique—for example, the only shared visit among the three is to the first persona, which is the default shown by the system. Hence, edit distance is a troublesome metric, because in this case, we would need 9 edits to make User 1 and User 2 — ratio-wise, this is 9/10, so 90% change rate (i.e., 9 out of 10 paths are different). Due to high uniqueness demonstrated by this example (which is also accentuated by the fact that the strings are of different lengths across the dataset!), the similarity of behaviors could perhaps be measured using other options. For example, User 2 and User 3 viewed 7 same personas (Mamdouh, Chris, Alaa, Abdalaziz, Ashley, John, and Putri) and 3 different personas (Rizky, Rahul, Muhammad). So, even though their exact viewing sequences are very different, the users actually view more the same than the different personas, i.e., there is likeness in their browsing behavior. To quantify this likeness, we can apply *set theory* to form an intersection (i.e., an overlap of paths). The intuition is that if two users visited

more same personas than two other users, their persona browsing behavior was more similar. To can quantify this by calculating the *Jaccard coefficient* (J), which simply indicates the overlap between two sets. This metric is commonly used in information theory to compare sets [44].

Applying J to our examples from Table 5, we can observe that User 2 and User 3 are more similar to each other ($J=0.7$) than User 1 and User 2 ($J=0.5$) or User 1 and User 3 ($J=0.5$). (For replication, the sets are: User 1 – M, A, P, A, R, B, A, P, R, A; User 2 – M, R, C, H, A, U, B, S, J, P; and User 3 – M, A, P, A, S, C, B, J, M, S.) Unlike ED, which is only applicable to pairwise comparison, sets can be expanded from pairwise comparisons to multiple sets (see Figure 14).

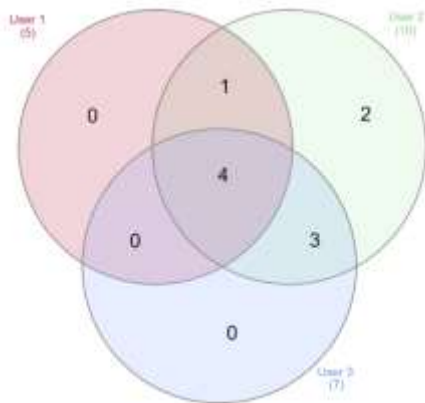


Figure 14: Among the 10 available personas, 4 (40%) were viewed by all three users. Users 2 and 3 viewed 3 personas that User 1 did not view, whereas Users 1 and 2 viewed 1 persona that User 3 did not view. User 2 viewed 2 personas that neither of the other users viewed, which indicates this user had the most diverse viewing pattern.

Obtaining the number of intersecting elements (i.e., shared personas that any number of participants viewed) is trivial and easy using basic functions in scripting languages like R and Python, which increases the practical appeal of modeling persona behavior using sets. One can also use sets to compare the behaviors of different groups. We illustrate some of these cases in Figure 15. For example, (a) **union** can be used for identifying all personas that a group of users viewed, which can be beneficial when the number of personas exceeds a handful, as would be the case for large and heterogeneous online audiences. (b) **Intersection** shows common elements of two or more users or groups. Intersection can reveal common personas of interest, i.e., that most participants engaged with. (c) **Difference** can show personas that one group viewed exclusively, e.g., those that were unique to more experienced persona users. Finally, (d) **subset and superset** can help make comparisons on the variety of personas visited. For example, in our previous case, User 2 is a superset of User 1 (i.e., User 2 visited all the personas that User 1 did and more).

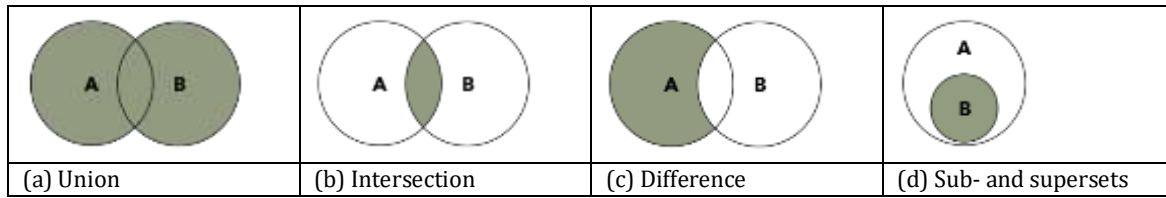


Figure 15: Examples of basic set operations and how they can be used for investigating persona user behavior.

While approaching the analysis of persona user behavior using set theory seems fruitful, we can include even more information in the comparison. Namely, the set representation ignores that the visits are typically unevenly distributed among the personas, both by count and duration. Some personas are viewed more than once; some are viewed considerably longer than others. A set would not consider this information at all.

For a representation that considers such information, we can turn to empirical distributions or probability distributions. These indicate how the time or number of visits is allocated between different personas during a user session. For example, if a user visits Alaa 5 times, Putri 5 times, and John 2 times, the empirical distribution is $5 / 12; 5 / 12; 2 / 12$; or $[0.42, 0.42, 0.16]$. For a finite set of personas and users, we can compute a complete probability distribution for each user, where complete implies that each persona-user pair will have a value. Then, from information theory, we can use several metrics to compare the obtained probability distributions. These are known as statistical distance metrics (e.g., Kullback–Leibler (KL) divergence or Jensen-Shannon distance). The smaller the distance between two users, the closer their behavior is in terms of how they divide their time among the available personas. Applying this logic to our dataset, we obtain the probability distributions indicated in [Table 6](#).

Table 6: Distribution of example users' visits among the shown personas. Non-zero values are highlighted.

	Mamdo uh	Rah ul	Ashl ey	Muham mad	Ala a	Joh n	Abdala ziz	Riz ky	Put ri	Chr is
User 1	10%	0%	0%	0%	40%	0%	10%	20%	20%	0%
User 2	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
User 3	20%	0%	20%	0%	20%	10%	10%	0%	10%	10%

The distance D between two distributions can be computed using the following equation,

$$D_{p,q} = \frac{1}{N} \sum_{i=1}^N |p_i(x) - q_i(x)|,$$

where p and q are the distributions to compare (e.g., User 1 and User 2). The formula calculates the absolute difference for both users for each persona X_i and then takes the average as the distance number. Unlike KL divergence, which is non-symmetrical (i.e., the distance between User 1 and User 2 might not be the same), this formula gives symmetrical results (i.e., $D_{pq} = D_{qp}$). (As a sidenote, symmetry is, of course, desirable for our purpose, because there is no reason to assume that the results should differ when comparing User 1 to User 2 or *vice versa*; in both cases, the sequences are the same.) When inputting the fractions from [Table 6](#) to this formula, we can see the results aligning with the J comparison, so that Users 2 and 3 are the most similar ($D = 0.06$), whereas User 1 is equally distant to User 2 ($D = 0.10$) and User 3 ($D = 0.10$). This example illustrates how concepts from information science can be leveraged for persona science, namely, by understanding persona viewing

behaviors as probability distributions and then computing distance. Essentially, the smaller these distance values are, the closer two persona viewing behaviors would be (as the behaviors are represented as probability distributions).

6.8 Interpreting the Metrics

It is well known that general rules about whether a metric value is “good” or “bad” are difficult to draw – for example, for an entertainment website or social media service, it is desirable that users spend a lot of time on the site, because time is positively correlated with revenue models [100]. However, the opposite applies for government information websites or search engines – the user is expected to find the information as quickly as possible and then leave the site. So, in some cases, small engagement time is optimal; in other cases, it is not optimal. For personas, the same applies, with even more nuance – in many professional tasks, users are time-pressed; they have deadlines, they want the information immediately, and so on. For these scenarios, low engagement time with a persona system would be considered a good sign (given that the user was able to complete their task successfully or that the personas helped the user). However, in scenarios where the user is conducting end-user research (e.g., market research), they might be interested in dwelling deep into the persona details: a low engagement time would therefore not indicate that the personas were useful for the user.

We provide further examples, as the matter of defining and computing various metrics is not trivial.

Example 1: Consider that a user visits 5 personas during the session and does overall 20 visits. Now, if we apply the average, we get $20 / 5 = 4$. However, the comparative aspect is if user visiting these 5 personas during their first 5 visits, versus visiting the 5 personas during their first 10 visits, the behavioral pattern is different. In the former case, the user is first visiting many personas, and then spending the rest of the time comparing them. In the latter case, the user is engaging in comparative behavior already during visiting the first personas. These two strategies would be different, but a simple average would miss the nuance. To quantify such patterns, we can compute x = how many personas the user visited and y = how many visits it took the user to visit each of the x personas at least once. This metric can be called “persona scanning tendency” or PST. Using the above example, the scores would be different:

$$\text{PST}_5 = 5 / 5 = 1 \text{ versus } \text{PST}_{10} = 5 / 10 = 0.5.$$

A lower PST score would indicate a smaller tendency to scan all or many personas first and then dwell into their information, and vice versa. A high PST score could also be associated with a linear user behavior, in which the user sequentially visits the available personas. In contrast, a low PST score could be associated with non-linear user behavior. For example, if it took the user ten visits to see all the five personas, they were likely doing comparisons along the way. Therefore, this PST metric can reveal insights about different users’ tendency to compare personas against one another and proceed in an organized manner when browsing the available personas.

Example 2: How to measure how diversely a user browses the personas? Highly diverse behavior would be one that looks at many personas. For example, if there were a total of 10 personas available, 5 out of 10 personas visited is less diverse than 10 out of 10 personas visited. That is, persona coverage would seem like a good metric (being $5 / 10 = 0.5$ and $10 / 10 = 1$ for the above cases). But, if the user only made 5 visits and visited 5 personas (out of the 10), then that is more diverse than making 12 visits and visiting 6 personas. In other words, when assessing the diversity of visit behavior, understood in this manner, the denominator should be the number of visits the particular user made, not the number of available personas. So, even if persona coverage would be higher for the user visiting 6 personas ($6 / 10 = 0.6$), their visit diversity would still be lower than that of the user visiting one persona less (i.e., $6 / 12 = 0.5$ vs. $5 / 5 = 1$).

There are also cases where different granularity would be needed, for example, considering all information elements (of which there are 120) versus only parent information elements (of which there are 8), or considering forward-only movement versus both forward and backward movements (i.e., across information elements and/or personas). Depending on these choices, the dimensionality of the analyzed dataset can greatly increase or decrease. Overall, these examples highlight the non-trivial nature of measuring and understanding persona user

behavior. Interpreting PA metrics follows the general patterns in UX research: a high value can indicate either a desirable or undesirable effect depending on the task type and user goals: for example, longer dwelling time may be a sign of more interest, but it could just as well be a sign of confusion and disorientation. Aligning the metrics with expressed user perceptions (e.g., “I was confused when using the personas”) can be useful in this regard. Overall, case-dependent interpretation is required.

7 Discussion

7.1 Highlights and Novelty

Even though there is some empirical evidence showing the effectiveness of personas for specific tasks [78], as a whole, the body of knowledge remains underdeveloped, even after 20 years of personas being part of HCI research and practice. The purpose of PA is to help generate new knowledge on persona user behavior towards the advancement of persona science. Examples include behaviors related to order of presenting the personas, revisit frequency, users’ styles of browsing and comparing personas, and persona choice – i.e., how and why people choose a specific persona for their decision-making task. Analyzing users’ typical viewing patterns and dwell time per persona information can inform persona information design (i.e., what information users most interact with), and help deduce understanding of persona usage based on a real system and real or realistic use cases and scenarios. Combined with algorithmically generated persona systems, this opens the possibility of updating persona profiles in real-time based on expressed user needs.

It is well known that the body of knowledge on personas relies on a high number of case studies as opposed to repeated experiments with independent samples (i.e., different organizations, locations, users, etc.). Even when a case study focuses on a specific context, there is a need for multiple case studies to establish something more generalizable than the one case can do [99]. Without having more rigor, persona research cannot proceed much, and it risks going in circles instead of establishing evidence-based empirical phenomena. Examples of this “going on circles” include conflicting findings about personas being applicable and not being applicable (e.g., [70] vs. [60]) – to date, nobody has explained *when* personas are applicable and when not. The only way to achieve plausible explanations is to carry out repeated testing and measure the results empirically, for which PA provides a way.

Users’ interaction with personas can be measured in many ways. Here, we focused on two commonly used technologies: mouse tracking and eye tracking, as these two technologies have unique strengths and weaknesses, and both can be implemented in a Web-based interactive persona system. As acknowledged in many HCI studies, eye- and mouse-tracking are helpful techniques for studying user engagement with interactive systems [15,91], and personas in particular [26,76,78,85,86]. When integrated directly into an interactive persona system, these data collection methods can deliver rich datasets describing persona use and modeling for complex user behaviors [18,22].

Moreover, integrating these techniques also enable user studies during exceptionally difficult times (e.g., during a global pandemic) when it is either not possible or very difficult to conduct in-person user studies.

Highlights of this work include the following:

- **Providing conceptual underpinnings of persona analytics and persona science**, two emerging and promising concepts for HCI and user segmentation.
- **Developing a novel persona analytics system embedded within a persona system**, by integrating mouse- and eye-tracking functionalities. Measuring users’ mouse and gaze activities grants an understanding of the behaviors of a persona user. Studying the behavior of the persona user is essential to generate empirical knowledge and theories of personas and human-persona interaction.
- **Demonstrating an end-to-end experimental loop for empirical persona studies**, using an interactive persona system and data collection platforms with a user study of 144 participants. Remote user studies such as this help scale persona user studies from the conventional 30-or-so to hundreds of participants.

- **Providing exploratory findings of persona user behavior**, based on the use of PA, its metrics, and various statistical techniques.

The trends contributing to interactive persona systems are likely to continue, including the evolution of (1) digital user data from online analytics platforms through APIs, (2) data science algorithms and libraries that integrate into quantitative persona creation process, and (3) web medium that surpasses the limitations of paper for persona delivery and user engagement. Interactive persona systems commonly rely on web standards [33] that enable users to access personas from any device with an internet connection and make it possible to record persona users' behavior and analyze with using PA, a customized solution for tracking users' interaction with persona profiles. This system, therefore, has value and potential for advancing persona science and the development of interactive persona systems for years to come.

The current work puts forward a new artifact for studying the behavior of persona users. While previous persona user studies [73,76,78] have analyzed dwell time and users' information viewing sequence, these metrics have, as far as we know, not been captured directly from an interactive persona system in any previous research. Thus, the presented solution has novelty in its field. As far as we know, the study presented the first "online laboratory" solution for interactive persona systems during a period in history where there is a need for remote user study solutions.

7.2 Novel Opportunities to Push Forward Persona Science

Persona science needs progress on all fronts, eyeing on long-term theory formulation but also investing in short-term returns through the use of empirical methods. Persona science can contribute to a much-needed transition beyond the general claims that "personas work" or "personas do not work", or the repetition of their "benefits" and "problems," into systematically examining the conditions where the effects emerge.

Many human factors have not been investigated empirically. Based on case studies, variables of special interest include at least (a) Experience [81], (b) Task type [4], (c) Job role [59], and (d) Culture [56]. Investigation of combination of these human factors would be based on specific research objectives.

First, the effect of **users' experience with personas** on behaviors; this tends to be reported in persona studies but not included as a variable. How does novice persona users' use of personas differ from more experienced users? Can the behaviors of more experienced users be used for guiding the novice users to learn to use personas more efficiently? The task and **task type**, more specifically. This is reported but rarely controlled – most typically, only one task type is deployed and in only one empirical setting, without repetition to achieve robustness. As a result, the current body of literature has, for example, no comparison between different task types – e.g., design, content creation, ad targeting, etc. Personas can be deployed for a range of professional tasks, but no study compares what kind of personas are ideal for the various task types and if users approach the personas differently based on the task type.

The users' **job roles** – again, reported, but comparison among different roles and organizational units are rarely conducted, even though it is common sense that a person's job position would greatly affect how they use personas to support their work. Studies tend to mention "designers" but looking deeper into these users' job positions, it is revealed that they work in multiple departments, have multiple different perspectives to the end-user, and require much different information for their decision making. Overall, systematic analysis of these variables in experimental studies can produce long-lasting, consistent, and robust knowledge on personas and their users, extending the boundaries of persona research. Finally, **culture**. Previous research has hinted at cultural effects related to personas use [83], but there is no adequate understanding of how the cultural match between the shown personas and the users mediate the interaction and whether personas themselves can help bridge cultural gaps for design.

[Table 7](#) proposes a preliminary research roadmap. Naturally, due to the enormous scope of potential research topics, this proposal is not a complete one. However, it maps some relevant questions that can be addressed via PA.

Table 7: Research roadmap for empirical persona research. This roadmap includes research questions that PA can help address. Through solving multiple research questions, researchers can start formulating a unified theory of personas that would explain, with robust empirical when and when not personas work, what factors govern their successful use, and how can persona creators and champions increase the likelihood persona projects.

	Open Research Question (ORQ) and Sub-Questions (SQ)	Useful for...*	Variation by...
ORQ1	<p><i>What types of personas are most/least viewed?</i></p> <ul style="list-style-type: none"> • SQ1a: Are there differences based on personas' gender, age, nationality, ethnic background? • SQ1b: Are some types of personas systematically disadvantaged in terms of how frequently and how long users interact with them? If so, can system features, e.g., rearranging order of showing, mitigate such disadvantages? 	Persona Creation and Use ☒ How do users interact with persona profiles?	<p><i>experience</i> <i>task type</i> <i>job role</i> <i>culture</i></p>
ORQ2	<p><i>What persona information was most/least viewed?</i></p> <ul style="list-style-type: none"> • SQ2a: How is users' consumption of persona information affected by position and screen size of the information? • SQ2b: Is some information considered redundant regardless of its position and screen size? 	Persona Information Design ☒ What information do persona users pay attention to?	
ORQ3	<p><i>How does the user transition between (a) the personas and (b) the information elements in the persona profiles?</i></p> <ul style="list-style-type: none"> • SQ3a: What is the degree of linearity / predictability? • SQ3b: Are the dwell times consistently increasing / decreasing? 	Persona Information Design ☒ How can we model users' cognitive styles of using personas?	
ORQ4	<p><i>How does increasing/decreasing the number of personas affect user behavior?</i></p> <ul style="list-style-type: none"> • SQ4a: What interaction techniques can help users cope with more than a handful of personas? • SQ4b: What is the extra cognitive cost of adding a persona? 	Persona User Behavior ☒ What is the optimal number of personas?	
ORQ5	<p><i>How do the viewed (a) personas and (b) persona information influence users' design choices?</i></p> <ul style="list-style-type: none"> • SQ5a: What is the effect of persona use of desirable outcomes such as increased usability, satisfaction, profitability? • SQ25b: How can these outcomes be measured? 	Persona Impact, Value of Personas ☒ What is the real value of personas?	

	Open Research Question (ORQ) and Sub-Questions (SQ)	Useful for...*	Variation by...
*NOTES: The categories of Persona Creation, Validation, Use, and Impact originate from [77]. ⇒ indicates a higher abstraction of the ORQ in question.			

7.3 Practical Implications

Benefits for Researchers. Persona science is the application of scientific methods to persona research, with the goal of producing empirically valid knowledge about persona creation, validation, use, and impact for design outcomes, individual decision making, and organizations. PA supports persona science by aiding in the scientific process, such as *generation and testing of hypotheses: data exploration → inductive analysis → propositions → hypotheses → independent data collection → hypothesis testing → theory generation of personas and their users*. PA can be deployed for both inductive and deductive research. Deductive research typically formulates hypotheses and relies on experiments to test those hypotheses. The hypotheses can be inspired by works in HCI, social psychology, economics, etc. Manipulating variables in the interactive persona system, such studies can be conducted in-person or remotely (see [Figure 7](#) illustrating an end-to-end loop for remote participation). Inductive studies rely on freely exploring user behaviors to formulate propositions that can later be tested in controlled settings. Here, we conduct a series of exploratory analyses in order to demonstrate the type of information, metrics, and analyses that the PA can afford to researchers.

Benefits Relative to Pre-Existing Solutions. From a practical point of view, the reader might pose the question, “Why not just use Google Analytics for this purpose? Why develop a new system?”. There are various reasons for that. Compared to pre-existing industry solutions, such as Google Analytics (GA), which is the dominant Web analytics service [65], PA has four major benefits, summarized in [Table 8](#) and explained thereafter:

Table 8: PA vs. GA – a practical comparison.

	Persona Analytics	Google Analytics
Data ownership	x	-
Tracking customization	x	Partial
Clickstream logs	x	Requires Premium version
Persona metric exports	x	-

- *Data ownership:* the data is recorded to our servers, not to third party servers such as Google’s.
- *Customized event and information tagging:* As we know the exact functionality of the APG system, we can label the information elements and events with suitable names from the onset.
- *Clickstream data:* Unlike in the standard GA installation, we are able to obtain raw log data of all actions taken by a user in the APG system. The standard GA installation only provides aggregate data export.
- *Tailored reports for persona user analysis:* GA reports are designed for websites, not for custom-built systems like interactive persona systems. Therefore, the available metrics and the reports are not suitable for analytical questions related to personas. The development of our reports was inspired by analytical tasks for which the data would be used. Thus, the reports serve persona research better than reports in GA.

With these benefits, the PA system is more equipped for tracking persona user behavior than GA, and therefore has the potential to serve researchers more adequately.

Benefits for end-users. Personas are an important tool for professionals in various domains, which is why understanding how people use them is instrumental for the creation of better end-user insights tools and therefore advocate more customer- and user-centric thinking in organizations. The practical implementation of PA also requires consideration for the users’ privacy, to let them know their usage of the system is being tracked.

Therefore, we notify users of this tracking in APG's terms of service, which is similar to any other website tracking user behavior. Additional ethical considerations include acquiring consent from users when conducting user studies to track their system usage behavior, as well as acceptance from an institutional review board (IRB) when there is a reason to suspect that the research deals with topics that warrant ethical scrutiny (in persona user studies, harmful scenarios tend to be rare but nonetheless could exist for some study topics).

7.4 Specific Application Areas for Future Research and Development

There are several areas to further deepen the level of analysis. The data produced by the PA system can be examined using many computationally advanced approaches, for example, by creating a persona state-transition matrix and applying Markov Chain modeling or neural networks to the constructed matrix to model historical dependencies (similar to [94]). Similarly, future work could look into predicting user outputs (e.g., task success) using behavioral features. There are several architectures to deal with sequential data, including recurrent neural networks (RNNs) [10], that could be used for modeling purposes that can result in persona recommenders or user behavior classification. While we leave these additional considerations for future work, we do want to point out two prominent opportunities:

- **Prediction.** By varying the personas, their information, or interactive features, the effect of such variables on outcome variables such as persona choice, task completion success, quality, time, perceived usefulness, or design task impact (e.g., usability improvement) can be measured. For example, when there are many personas available, users often choose a specific persona for their task. This choice matters because selecting one persona over another can influence how different user groups' needs are considered (or not considered).
- **Persona recommenders.** For large and heterogeneous populations, the appropriate number of personas can be in the hundreds or more to correctly represent the diversity of the user population [11]. Therefore, there is a need, on the one hand, to show more personas to users, and on the other hand, to build tools and methods for users or systems to narrow down the number of candidate personas for a given task, without taking away users' choice of reaching beyond this candidate pool to browse marginalized or fringe user groups [21]. Approaches for achieving this may include recommending users a specific set of personas based on task criteria or user modeling. Simple examples that have been already implemented in interactive persona systems include sorting the personas based on their segment size (i.e., how many people they represent in the baseline data), either in *decreasing* order (when wanting to see the most representative personas) or in *increasing* order (when wanting to see outlier personas). However, more elaborate persona recommenders are missing to date. The recommenders' value increases as the number of personas increases because large persona sets with hundreds or even thousands of personas [27] pose a manageability problem for a human efficiently to deal with.

Finally, the PA approach could be applied to other forms of profile systems, including social media profiles, gaming avatar profiles, and so on. Although the system was designed with personas in mind, if the use case of wanting to learn how people interact with profiles is adequately similar, then PA-type of analytics can be deployed.

8 Conclusion

In this work, we demonstrated a fully functional persona analytics system embedded within a fully functional interactive persona system. We demonstrated ways in which persona analytics can reveal structural patterns in user interaction with personas. These patterns can lead to advancement of persona science and theoretical propositions for persona-user interaction. Persona research lacks empirical studies, so there is plenty of room for contributions by keen researchers. To this end, we hope our research encourages others to pursue empirical research questions in the persona domain. The techniques we demonstrated have special value during

exceptional times when physical user studies are hindered by social distancing. Finally, we proposed seven metrics that can be computed from the persona analytics data. In addition to these metrics, devising new quantitative metrics for persona studies is a valuable research direction.

References

- < bib id="bib1">< number>[1]</ number> Mohamed Aboelmaged and Samar Mouakket. 2020. Influencing models and determinants in big data analytics research: A bibliometric analysis. *Information Processing & Management* 57, 4 (July 2020), 102234. DOI:https://doi.org/10.1016/j.ipm.2020.102234</ bib>
- < bib id="bib2">< number>[2]</ number> Jisun An, Haewoon Kwak, Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2018. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining* 8, 1 (2018), 54. DOI:https://doi.org/10.1007/s13278-018-0531-0</ bib>
- < bib id="bib3">< number>[3]</ number> Jisun An, Haewoon Kwak, Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2018. Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web (TWEB)* 12, 4 (2018), 27. DOI:https://doi.org/10.1145/3265986</ bib>
- < bib id="bib4">< number>[4]</ number> F. Anvari, D. Richards, M. Hitchens, and M. A. Babar. 2015. Effectiveness of Persona with Personality Traits on Conceptual Design. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, 263–272. DOI:https://doi.org/10.1109/ICSE.2015.155</ bib>
- < bib id="bib5">< number>[5]</ number> Farshid Anvari, Deborah Richards, Michael Hitchens, Muhammad Ali Babar, Hien Minh Thi Tran, and Peter Busch. 2017. An empirical investigation of the influence of persona with personality traits on conceptual design. *Journal of Systems and Software* 134, (December 2017), 324–339. DOI:https://doi.org/10.1016/j.jss.2017.09.020</ bib>
- < bib id="bib6">< number>[6]</ number> Farshid Anvari, Deborah Richards, Michael Hitchens, and Hien Minh Thi Tran. 2019. Teaching User Centered Conceptual Design Using Cross-Cultural Personas and Peer Reviews for a Large Cohort of Students. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*, 62–73. DOI:https://doi.org/10.1109/ICSE-SEET.2019.00015</ bib>
- < bib id="bib7">< number>[7]</ number> M. Aoyama. 2005. Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In *Proceedings of the 13th IEEE International Conference on Requirements Engineering (RE'05)*, Washington, DC, USA, 85–94. DOI:https://doi.org/10.1109/RE.2005.50</ bib>
- < bib id="bib8">< number>[8]</ number> M. Aoyama. 2007. Persona-Scenario-Goal Methodology for User-Centered Requirements Engineering. In *Proceedings of the 15th IEEE International Requirements Engineering Conference (RE 2007)*, Delhi, India, 185–194. DOI:https://doi.org/10.1109/RE.2007.50</ bib>
- < bib id="bib9">< number>[9]</ number> Ernesto Arroyo, Ted Selker, and Willy Wei. 2006. Usability tool for analysis of web designs using mouse tracks. In *CHI'06 extended abstracts on Human factors in computing systems*, 484–489.</ bib>
- < bib id="bib10">< number>[10]</ number> Homanga Bharadhwaj. 2019. Explainable recommender system that maximizes exploration. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion*, 1–2.</ bib>
- < bib id="bib11">< number>[11]</ number> Chris Chapman, Edwin Love, Russell P. Milham, Paul ElRif, and James L. Alford. 2008. Quantitative Evaluation of Personas as Information. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1107–1111. DOI:https://doi.org/10.1177/154193120805201602</ bib>
- < bib id="bib12">< number>[12]</ number> 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*, 634–636. DOI:https://doi.org/10.1177/154193120605000503</ bib>
- < bib id="bib13">< number>[13]</ number> Eric Chu, Prashanth Vijayaraghavan, and Deb Roy. 2018. Learning Personas from Dialogue with Attentive Memory Networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Brussels, Belgium, 2638–2646. Retrieved June 20, 2019 from https://www.aclweb.org/anthology/D18-1284</ bib>
- < bib id="bib14">< number>[14]</ number> Theresa Bilitski Clarke and Bernard J. Jansen. 2017. Conversion potential: a metric for evaluating search engine advertising performance. *Journal of Research in Interactive Marketing* 11, 2 (2017), 142–159.</ bib>
- < bib id="bib15">< number>[15]</ number> Argyris Constantinides, Marios Belk, Christos Fidas, and Andreas Pitsillides. 2020. An eye gaze-driven metric for estimating the strength of graphical passwords based on image hotspots. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 33–37.</ bib>
- < bib id="bib16">< number>[16]</ number> 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.</ bib>
- < bib id="bib17">< number>[17]</ number> Giuseppe Desolda, Rosa Lanzilotti, Danilo Caivano, and Maria F. Costabile. 2021. An experience on remote testing exploiting new web technology. In *INTERACT'21 Workshop on Remote user Testing - Experiences and Trends*, Bari, Italy.</ bib>
- < bib id="bib18">< number>[18]</ number> A. T. Duchowski. 2009. *Eye Tracking Methodology: Theory and Practice*. Springer, London.</ bib>
- < bib id="bib19">< number>[19]</ number> Hermann Ebbinghaus. 2013. Memory: A contribution to experimental psychology. *Annals of neurosciences* 20, 4 (2013), 155.</ bib>
- < bib id="bib20">< number>[20]</ number> Achim Ebert, Shah Rukh Humayoun, Norbert Seyff, Anna Perini, and Simone D.J. Barbosa (Eds.). 2016. *Usability- and Accessibility-Focused Requirements Engineering*. Springer International Publishing, Cham. DOI:https://doi.org/10.1007/978-3-319-45916-5</ bib>
- < bib id="bib21">< number>[21]</ number> Joy Goodman-Deane, Sam Waller, Dana Demin, Arantxa González-de-Heredia, Mike Bradley, and John P. Clarkson. 2018. Evaluating Inclusivity using Quantitative Personas. In *In the Proceedings of Design Research Society Conference 2018*, Limerick, Ireland. DOI:https://doi.org/10.21606/drs.2018.400</ bib>
- < bib id="bib22">< number>[22]</ number> Thomas Grindinger, Andrew T. Duchowski, and Michael Sawyer. 2010. Group-wise Similarity and Classification of Aggregate Scanpaths. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*, ACM, New York, NY, USA, 101–104. DOI:https://doi.org/10.1145/1743666.1743691</ bib>
- < bib id="bib23">< number>[23]</ number> Jonathan Grudin. 2006. Why Personas Work: The Psychological Evidence. In *The Persona Lifecycle*, John Pruitt and Tamara Adlin (eds.). Elsevier, 642–663. DOI:https://doi.org/10.1016/B978-012566251-2/50013-7</ bib>

< bib id="bib24">< number>[24]</ number> Jonathan Grudin and John Pruitt. 2002. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. In *Proceedings of Participation and Design Conference (PDC2002)*, Sweden, 8.</ bib>

< bib id="bib25">< number>[25]</ number> Jacek Gwizdka, Rahilsadat Hosseini, Michael Cole, and Shouyi Wang. 2017. Temporal dynamics of eye-tracking and EEG during reading and relevance decisions. *Journal of the Association for Information Science and Technology* 68, 10 (October 2017), 2299–2312. DOI:https://doi.org/10.1002/asi.23904</ bib>

< bib id="bib26">< number>[26]</ number> Charles G. Hill, Maren Haag, Alannah Oleson, Chris Mendez, Nicola Marsden, Anita Sarma, and Margaret Burnett. 2017. Gender-Inclusiveness Personas vs. Stereotyping: Can We Have it Both Ways? In *Proceedings of the 2017 CHI Conference*, ACM Press, Denver, Colorado, USA, 6658–6671. DOI:https://doi.org/10.1145/3025453.3025609</ bib>

< bib id="bib27">< number>[27]</ number> 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*, ACM, Glasgow, United Kingdom, 1–6. DOI:https://doi.org/10.1145/3290607.3313006</ bib>

< bib id="bib28">< number>[28]</ number> Bernard J. Jansen, Soon-gyo Jung, and Joni Salminen. 2020. From flat file to interface: Synthesis of personas and analytics for enhanced user understanding. *Proceedings of the Association for Information Science and Technology* 57, 1 (October 2020). DOI:https://doi.org/10.1002/pr.a.2.215</ bib>

< bib id="bib29">< number>[29]</ number> Bernard J. Jansen, Joni 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. DOI:https://doi.org/10.2478/dim-2020-0005</ bib>

< bib id="bib30">< number>[30]</ number> Bernard Jansen, Joni Salminen, Soon-gyo Jung, and Kathleen Guan. 2021. *Data-Driven Personas* (1st ed.). Morgan & Claypool Publishers. Retrieved February 10, 2021 from https://www.morganclaypool.com/doi/abs/10.2200/S01072ED1V01Y202101HCI048</ bib>

< bib id="bib31">< number>[31]</ number> Joel Järvinen and Heikki Karjaluo. 2015. The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management* 50, Supplement C (October 2015), 117–127. DOI:https://doi.org/10.1016/j.indmarman.2015.04.009</ bib>

< bib id="bib32">< number>[32]</ number> Soon-gyo Jung, Jisun An, Haewoon Kwak, Moeed Ahmad, Lene Nielsen, and Bernard J. Jansen. 2017. Persona Generation from Aggregated Social Media Data. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*, ACM, Denver, Colorado, USA, 1748–1755.</ bib>

< bib id="bib33">< number>[33]</ number> Soon-Gyo Jung, Joni Salminen, Jisun An, Haewoon Kwak, and Bernard J Jansen. 2018. Automatically Conceptualizing Social Media Analytics Data via Personas. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2018)*, San Francisco, California, USA, 2.</ bib>

< bib id="bib34">< number>[34]</ number> Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2019. Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, ACM Press, Glasgow, Scotland Uk, 1–6. DOI:https://doi.org/10.1145/3290607.3312955</ bib>

< bib id="bib35">< number>[35]</ number> Soon-Gyo Jung, Joni Salminen, and Bernard J. Jansen. 2020. Giving Faces to Data: Creating Data-Driven Personas from Personified Big Data. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion (IUI '20)*, Association for Computing Machinery, Cagliari, Italy, 132–133. DOI:https://doi.org/10.1145/3379336.3381465</ bib>

< bib id="bib36">< number>[36]</ number> Soon-gyo Jung, Joni Salminen, and Bernard J Jansen. 2021. Implementing Eye-Tracking for Persona Analytics. In *ETRA '21 Adjunct: ACM Symposium on Eye Tracking Research and Applications*, ACM, Virtual conference, 1–4. DOI:https://doi.org/10.1145/3450341.3458765</ bib>

< bib id="bib37">< number>[37]</ number> Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2021. Persona Analytics: Implementing Mouse-tracking for an Interactive Persona System. In *Extended Abstracts of ACM Human Factors in Computing Systems - CHI EA '21*, ACM, Virtual conference. DOI:https://doi.org/10.1145/3411763.3451773</ bib>

< bib id="bib38">< number>[38]</ number> Soon-gyo Jung, Joni Salminen, Haewoon Kwak, Jisun An, and Bernard J. Jansen. 2018. Automatic Persona Generation (APG): A Rationale and Demonstration. In *CHIIR '18: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, ACM, New Jersey, USA, 321–324. DOI:https://doi.org/10.1145/3176349.3176893</ bib>

< bib id="bib39">< number>[39]</ number> Pascal J. Kieslich, Felix Henninger, Dirk U. Wulff, Jonas MB Haslbeck, and Michael Schulte-Mecklenbeck. 2019. Mouse-Tracking: A Practical Guide to Implementation and Analysis 1. In *A handbook of process tracing methods*. Routledge, 111–130.</ bib>

< bib id="bib40">< number>[40]</ number> Ari Kolbeinsson, Erik Brolin, and Jessica Lindblom. 2021. Data-Driven Personas: Expanding DHM for a Holistic Approach. In *International Conference on Applied Human Factors and Ergonomics*, Springer, 296–303.</ bib>

< bib id="bib41">< number>[41]</ number> Dannie Korsgaard, Thomas Bjørner, Pernille Krog Sørensen, and Paolo Burelli. 2020. Creating user stereotypes for persona development from qualitative data through semi-automatic subspace clustering. *User Model User-Adap Inter* 30, 1 (March 2020), 81–125. DOI:https://doi.org/10.1007/s11257-019-09252-5</ bib>

< bib id="bib42">< number>[42]</ number> Chris Lankford. 2000. Gazetracker: Software Designed to Facilitate Eye Movement Analysis. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA '00)*, ACM, New York, NY, USA, 51–55. DOI:https://doi.org/10.1145/355017.355025</ bib>

< bib id="bib43">< number>[43]</ number> Daniel D. Lee and Sebastian H. Seung. 1999. Learning the Parts of Objects by Non-negative Matrix Factorization. *Nature* 401, 6755 (1999), 788–791.</ bib>

< bib id="bib44">< number>[44]</ number> Michael Levandowsky and David Winter. 1971. Distance between Sets. *Nature* 234, 5323 (November 1971), 34–35. DOI:https://doi.org/10.1038/234034a0</ bib>

< bib id="bib45">< number>[45]</ number> Jingyi Li, Michelle X. Zhou, Huahai Yang, and Gloria Mark. 2017. Confiding in and listening to virtual agents: The effect of personality. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, ACM, 275–286.</ bib>

< bib id="bib46">< number>[46]</ number> Tianyi Li, Gregorio Convertino, Ranjeet Kumar Tayi, and Shima Kazerooni. 2019. What data should I protect? Recommender and planning support for data security analysts. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 286–297.</ bib>

< bib id="bib47">< number>[47]</ number> Yuting Liao and Jiagen He. 2020. Racial mirroring effects on human-agent interaction in psychotherapeutic conversations. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 430–442.</ bib>

< bib id="bib48">< number>[48]</ number> Nicola Marsden and Maren Haag. 2016. Stereotypes and politics: reflections on personas. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4017–4031.</ bib>

< bib id="bib49">< number>[49]</ number> Tara Matthews, Tejinder Judge, and Steve Whittaker. 2012. How do designers and user experience professionals actually perceive and use personas? In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, ACM Press, Austin, Texas, USA, 1219. DOI:https://doi.org/10.1145/2207676.2208573</ bib>

< bib id="bib50">< number>[50]</ number> Jennifer Jen McGinn and Nalini Kotamraju. 2008. Data-driven persona development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, Italy, 1521–1524. DOI:https://doi.org/10.1145/1357054.1357292</ bib>

< bib id="bib51">< number>[51]</ number> Tomasz Miaskiewicz and Coryndon Luxmoore. 2017. The Use of Data-Driven Personas to Facilitate Organizational Adoption—A Case Study. *The Design Journal* 20, 3 (2017), 357–374.</ bib>

< bib id="bib52">< number>[52]</ number> T. Mijač, M. Jadrič, 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)*, 1237–1242. DOI:https://doi.org/10.23919/MIPRO.2018.8400224</ bib>

< bib id="bib53">< number>[53]</ number> Thomas P. Moran. 1981. Guest Editor's Introduction: An Applied Psychology of the User. *ACM Comput Surv.* 13, 1 (March 1981), 1–11. DOI:https://doi.org/10.1145/356835.356836</ bib>

< bib id="bib54">< number>[54]</ number> Bennet B. Murdock Jr. 1962. The serial position effect of free recall. *Journal of Experimental Psychology* 64, 5 (1962), 482–488. DOI:https://doi.org/10.1037/h0045106</ bib>

< bib id="bib55">< number>[55]</ number> Vidhya Navalpakkam and Elizabeth Churchill. 2012. Mouse tracking: measuring and predicting users' experience of web-based content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2963–2972.</ bib>

< bib id="bib56">< number>[56]</ number> Lene Nielsen. 2010. Personas in Cross-Cultural Projects. In *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts*, Dinesh Katre, Rikke Orngreen, Pradeep Yammiyavar and Torkil Clemmensen (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 76–82. DOI:https://doi.org/10.1007/978-3-642-11762-6_7</ bib>

< bib id="bib57">< number>[57]</ number> Lene Nielsen. 2019. *Personas - User Focused Design* (2nd ed. 2019 edition ed.). Springer, New York, NY, USA.</ bib>

< bib id="bib58">< number>[58]</ number> 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. *International Journal of Sociotechnology and Knowledge Development* 7, 1 (2015), 45–61. DOI:https://doi.org/10.4018/ijskd.2015010104</ bib>

< bib id="bib59">< number>[59]</ number> Lene Nielsen, Soon-gyo Jung, Jisun An, Joni Salminen, Haewoon Kwak, and Bernard J. Jansen. 2017. Who Are Your Users?: Comparing Media Professionals' Preconception of Users to Data-driven Personas. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction (OZCHI '17)*, ACM, Brisbane, Queensland, Australia, 602–606. DOI:https://doi.org/10.1145/3152771.3156178</ bib>

< bib id="bib60">< number>[60]</ number> Lene Nielsen and Kira Storgaard Hansen. 2014. Personas is applicable: a study on the use of personas in Denmark. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Toronto, Ontario, Canada, 1665–1674.</ bib>

< bib id="bib61">< number>[61]</ number> Stefan Palan and Christian Schitter. 2018. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17, (2018), 22–27.</ bib>

< bib id="bib62">< number>[62]</ number> Alexandra Papoutsaki, Aaron Gokaslan, James Tompkin, Yuze He, and Jeff Huang. 2018. The eye of the typer: a benchmark and analysis of gaze behavior during typing. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*, Association for Computing Machinery, New York, NY, USA, 1–9. DOI:https://doi.org/10.1145/3204493.3204552</ bib>

< bib id="bib63">< number>[63]</ number> Alexandra Papoutsaki, James Laskey, and Jeff Huang. 2017. SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval (CHIIR '17)*, Association for Computing Machinery, New York, NY, USA, 17–26. DOI:https://doi.org/10.1145/3020165.3020170</ bib>

< bib id="bib64">< number>[64]</ number> Alexandra Papoutsaki, Patson Sangkloy, James Laskey, Nediyan Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: Scalable Webcam Eye Tracking Using User Interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*, AAAI Press, New York, New York, USA, 3839–3845. Retrieved from <http://dl.acm.org/citation.cfm?id=3061053.3061156></ bib>

< bib id="bib65">< number>[65]</ number> Beatriz Plaza. 2011. Google Analytics for measuring website performance. *Tourism Management* 32, 3 (2011), 477–481.</ bib>

< bib id="bib66">< number>[66]</ number> John Pruitt and Jonathan Grudin. 2003. Personas: Practice and Theory. In *Proceedings of the 2003 Conference on Designing for User Experiences (DUX '03)*, ACM, San Francisco, California, USA, 1–15. DOI:https://doi.org/10.1145/997078.997089</ bib>

< bib id="bib67">< number>[67]</ number> Kun Qian, Marina Danilevsky, Yannis Katsis, Ban Kawas, Erick Oduor, Lucian Popa, and Yunyao Li. 2021. XNLP: A Living Survey for XAI Research in Natural Language Processing. In *26th International Conference on Intelligent User Interfaces*, 78–80.</ bib>

< bib id="bib68">< number>[68]</ number> Gertrude Raffle. 1936. Two Determinants of the Effect of Primacy. *The American Journal of Psychology* 48, 4 (1936), 654–657. DOI:https://doi.org/10.2307/1416518</ bib>

< bib id="bib69">< number>[69]</ number> Kari Rönkkö. 2005. An Empirical Study Demonstrating How Different Design Constraints, Project Organization and Contexts Limited the Utility of Personas. In *Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences - Volume 08 (HICSS '05)*, IEEE Computer Society, Washington, DC, USA. DOI:https://doi.org/10.1109/HICSS.2005.85</ bib>

< bib id="bib70">< number>[70]</ number> Kari Rönkkö, Mats Hellman, Britta Kilander, and Yvonne Dittrich. 2004. Personas is Not Applicable: Local Remedies Interpreted in a Wider Context. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1 (PDC 04)*, ACM, Toronto, Ontario, Canada, 112–120. DOI:https://doi.org/10.1145/1011870.1011884</ bib>

< bib id="bib71">< number>[71]</ number> Darrell S. Rudmann, George W. McConkie, and Xianjun Sam Zheng. 2003. Eyetracking in cognitive state detection for HCI. In *Proceedings of the 5th international conference on Multimodal interfaces*, ACM, 159–163. Retrieved May 10, 2017 from <http://dl.acm.org/citation.cfm?id=958464></ bib>

< bib id="bib72">< number>[72]</ number> Joni Salminen, Kathleen Guan, Soon-gyo Jung, Shammur Absar Chowdhury, and Bernard J. Jansen. 2020. A Literature Review of Quantitative Persona Creation. In *CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu, Hawaii, USA, 1–14. DOI:https://doi.org/10.1145/3313831.3376502</ bib>

< bib id="bib73">< number>[73]</ number> Joni Salminen, Bernard J. Jansen, Jisun An, Soon-gyo Jung, Lene Nielsen, and Haewoon Kwak. 2018. Fixation and Confusion – Investigating Eye-tracking Participants' Exposure to Information in Personas. In *Proceedings of the ACM SIGIR Conference*

on *Human Information Interaction and Retrieval (CHIIR 2018)*, ACM, New Jersey, USA, 110–119.
DOI:<https://doi.org/10.1145/3176349.3176391>

< bib id="bib74">< number>[74]</ number> Joni Salminen, Bernard J. Jansen, Jisun An, Haewoon Kwak, and Soon-gyo Jung. 2018. Are personas done? Evaluating their usefulness in the age of digital analytics. *Persona Studies* 4, 2 (November 2018), 47–65.
DOI:<https://doi.org/10.21153/psj2018vol4no2art737>

< bib id="bib75">< number>[75]</ number> Joni Salminen, Bernard J. Jansen, Jisun An, Haewoon Kwak, and Soon-gyo Jung. 2019. Automatic Persona Generation for Online Content Creators: Conceptual Rationale and a Research Agenda. In *Personas - User Focused Design* (2nd ed.), Lene Nielsen (ed.), Springer London, London, 135–160. DOI:https://doi.org/10.1007/978-1-4471-7427-1_8

< bib id="bib76">< number>[76]</ number> Joni Salminen, Soon-gyo Jung, Jisun An, Haewoon Kwak, Lene Nielsen, and Bernard J. Jansen. 2019. Confusion and information triggered by photos in persona profiles. *International Journal of Human-Computer Studies* 129, (September 2019), 1–14. DOI:<https://doi.org/10.1016/j.ijhcs.2019.03.005>

< bib id="bib77">< number>[77]</ number> Joni Salminen, Soon-Gyo Jung, Kamal Chhirang, and Bernard Jansen. 2021. Instilling Knowledge Claims of Personas from 346 Research Articles. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–9. Retrieved August 19, 2021 from <https://doi.org/10.1145/3411763.3451619>

< bib id="bib78">< number>[78]</ number> Joni Salminen, Soon-gyo Jung, Shammur Absar Chowdhury, Sercan Sengün, and Bernard J Jansen. 2020. Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task. In *Proceedings of the ACM Conference of Human Factors in Computing Systems (CHI'20)*, ACM, Honolulu, Hawaii, USA. DOI:<https://doi.org/10.1145/3313831.3376770>

< bib id="bib79">< number>[79]</ number> Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2019. The future of data-driven personas: A marriage of online analytics numbers and human attributes. In *ICEIS 2019 - Proceedings of the 21st International Conference on Enterprise Information Systems*, SciTePress, Heraklion, Greece, 596–603.

< bib id="bib80">< number>[80]</ number> Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2020. Explaining Data-Driven Personas. In *Proceedings of the Workshop on Explainable Smart Systems for Algorithmic Transparency in Emerging Technologies co-located with 25th International Conference on Intelligent User Interfaces (IUI 2020)* (Vol-2582), CEUR Workshop Proceedings, Cagliari, Italy, 7.
DOI:<https://doi.org/urn:nbn:de:0074-2582-4>

< bib id="bib81">< number>[81]</ number> Joni Salminen, Soon-gyo Jung, João M. Santos, Shammur Chowdhury, and Bernard J. Jansen. 2020. The Effect of Experience on Persona Perceptions. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '20)*, Association for Computing Machinery, Honolulu, HI, USA, 1–9. DOI:<https://doi.org/10.1145/3334480.3382786>

< bib id="bib82">< number>[82]</ number> Joni Salminen, Soon-gyo Jung, João M. Santos, and Bernard J. Jansen. 2019. Does a Smile Matter if the Person Is Not Real?: The Effect of a Smile and Stock Photos on Persona Perceptions. *International Journal of Human-Computer Interaction* 0, 0 (September 2019), 1–23. DOI:<https://doi.org/10.1080/10447318.2019.1664068>

< bib id="bib83">< number>[83]</ number> Joni Salminen, Soon-gyo Jung, João M. Santos, Ahmed Mohamed Kamel, and Bernard J. Jansen. 2021. Picturing It!: The Effect of Image Styles on User Perceptions of Personas. In *Proceedings of ACM Human Factors in Computing Systems (CHI'21)*, ACM, Virtual conference, 1–6. DOI:<https://doi.org/10.1145/3411764.3445360>

< bib id="bib84">< number>[84]</ number> Joni Salminen, Ilkka Kaate, Ahmed Mohamed Sayed Kamel, Soon-gyo Jung, and Bernard J. Jansen. 2020. How Does Persona Interaction Impact Ad Performance and Empathy? An Experiment with Online Advertising. *International Journal of Human-Computer Interaction* 0, 0 (August 2020), 1–15. DOI:<https://doi.org/10.1080/10447318.2020.1809246>

< bib id="bib85">< number>[85]</ number> Joni Salminen, Ying-Hsang Liu, Sercan Sengun, João M. Santos, Soon-gyo Jung, and Bernard J. Jansen. 2020. The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces. In *IUI '20: Proceedings of the 25th International Conference on Intelligent User Interfaces*, ACM, Cagliari, Italy, 357–368.
DOI:<https://doi.org/10.1145/3377325.3377492>

< bib id="bib86">< number>[86]</ number> Joni Salminen, Lene Nielsen, Soon-gyo Jung, Jisun An, Haewoon Kwak, and Bernard J Jansen. 2018. "Is More Better?": Impact of Multiple Photos on Perception of Persona Profiles. In *Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI2018)*, ACM, Montréal, Canada. DOI:<https://doi.org/10.1145/3173574.3173891>

< bib id="bib87">< number>[87]</ number> Joni Salminen, Rohan Gurunandan Rao, Soon-gyo Jung, Shammur A. Chowdhury, and Bernard J. Jansen. 2020. Enriching Social Media Personas with Personality Traits: A Deep Learning Approach Using the Big Five Classes. In *International Conference on Human-Computer Interaction*, Springer, 101–120.

< bib id="bib88">< number>[88]</ number> Joni Salminen, Joao M. Santos, Soon-gyo Jung, Motahhare Eslami, and Bernard J. Jansen. 2019. Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. *International Journal of Human-Computer Interaction* 0, 0 (November 2019), 1–13. DOI:<https://doi.org/10.1080/10447318.2019.1688946>

< bib id="bib89">< number>[89]</ number> Joni Salminen, Sercan Şengün, Haewoon Kwak, Bernard J. Jansen, Jisun An, Soon-gyo Jung, Sarah Vieweg, and D. Fox Harrell. 2018. From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. *FM* 23, 6 (June 2018). DOI:<https://doi.org/10.5210/fm.v23i6.8415>

< bib id="bib90">< number>[90]</ number> Joni Salminen, Jukka Vahlo, Aki Koponen, Soon-gyo Jung, Shammur A. Chowdhury, and Bernard J. Jansen. 2020. Designing Prototype Player Personas from a Game Preference Survey. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '20)*, Association for Computing Machinery, Honolulu, HI, USA, 1–8.
DOI:<https://doi.org/10.1145/3334480.3382785>

< bib id="bib91">< number>[91]</ number> Jayasankar Santhosh, Shoya Ishimaru, and Andreas Dengel. 2020. Generating Heatmap for Unknown Documents towards Readability Measurement. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, 47–48.

< bib id="bib92">< number>[92]</ number> Claude E Shannon, Warren Weaver, and Shannon. 1998. *The Mathematical Theory of Communication* (1ST ed.). University of Illinois Press.

< bib id="bib93">< number>[93]</ number> H. A. Simon. 1956. Rational choice and the structure of the environment. *Psychological Review* 63, 2 (1956), 129–138. DOI:<https://doi.org/10.1037/h0042769>

< bib id="bib94">< number>[94]</ number> Harold Soh, Scott Sanner, Madeleine White, and Greg Jamieson. 2017. Deep Sequential Recommendation for Personalized Adaptive User Interfaces. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, ACM, Limassol Cyprus, 589–593. DOI:<https://doi.org/10.1145/3025171.3025207>

< bib id="bib95">< number>[95]</ number> Dimitris Spiliotopoulos, Dionisis Margaritis, and Costas Vassilakis. 2020. Data-Assisted Persona Construction Using Social Media Data. *Big Data and Cognitive Computing* 4, 3 (September 2020), 21. DOI:https://doi.org/10.3390/bdcc4030021</ bib>

< bib id="bib96">< number>[96]</ number> Phillip Douglas Stevenson and Christopher Andrew Mattson. 2019. The Personification of Big Data. *Proceedings of the Design Society: International Conference on Engineering Design* 1, 1 (July 2019), 4019–4028. DOI:https://doi.org/10.1017/dsi.2019.409</ bib>

< bib id="bib97">< number>[97]</ number> Yasuhiro Watanabe, Hironori Washizaki, Kiyoshi Honda, Yuki Noyori, Yoshiaki Fukazawa, Aoi Morizuki, Hiroyuki Shibata, Kentaro Ogawa, Mikako Ishigaki, Satiyo Shiizaki, Teppei Yamaguchi, and Tomoaki Yagi. 2017. ID3P: Iterative Data-driven Development of Persona Based on Quantitative Evaluation and Revision. In *Proceedings of the 10th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE '17)*, IEEE Press, Piscataway, NJ, USA, 49–55. DOI:https://doi.org/10.1109/CHASE.2017.9</ bib>

< bib id="bib98">< number>[98]</ number> Xiaozhi Yang and Ian Krajbich. 2020. Webcam-based online eye-tracking for behavioral research. (2020).</ bib>

< bib id="bib99">< number>[99]</ number> Robert K. Yin. 2003. *Case Study Research: Design and Methods*. SAGE Publications.</ bib>

< bib id="bib100">< number>[100]</ number> Melita Zajc. 2015. The social media dispositive and monetization of user-generated content. *The Information Society* 31, 1 (2015), 61–67.</ bib>

< bib id="bib101">< number>[101]</ number> Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. 2016. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*, ACM, San Jose, California, USA, 5350–5359.</ bib>

< bib id="bib102">< number>[102]</ number> Haining Zhu, Hongjian Wang, and John M. Carroll. 2019. Creating Persona Skeletons from Imbalanced Datasets - A Case Study using U.S. Older Adults' Health Data. In *Proceedings of the 2019 on Designing Interactive Systems Conference - DIS '19*, ACM Press, San Diego, CA, USA, 61–70. DOI:https://doi.org/10.1145/3322276.3322285</ bib>