

ARE DATA-DRIVEN PERSONAS CONSIDERED HARMFUL? DIVERSIFYING USER UNDERSTANDINGS WITH MORE THAN ALGORITHMS

JONI SALMINEN HAMAD BIN KHALIFA UNIVERSITY & THE UNIVERSITY OF TURKU, SOON-GYO
JUNG HAMAD BIN KHALIFA UNIVERSITY AND BERNARD J. JANSEN HAMAD BIN KHALIFA
UNIVERSITY

ABSTRACT

In this work, we build on research on data-driven personas to present what might be “wrong with them”. From wrong assumptions by the client and wrong applications of methods to imbalanced, messy, or superficial data; a lack of communication regarding how these personas are created; and issues with usability, there are a plethora of issues that plague data-driven personas. We conclude by contemplating whether data-driven personas are even worthwhile and, if they are, then what are some of the immediate remedies required from the human-computer interaction community to make data-driven personas a viable tool for user understanding.

KEY WORDS

Personas, Data-Driven Personas, Harm

INTRODUCTION

Persona is a user-centred design (UCD) technique applied in HCI, user experience (UX), and other fields such as business and marketing. Personas, in these contexts, are defined as archetypes of user groups that share similar traits or behaviours (e.g., goals), and they are typically portrayed in the form of user profiles (Cooper 1999; Nielsen et al. 2015).

Data-driven personas, supposedly, elevate manually created personas from low-tech design artefacts to high-tech user representations (Jansen et al. 2020). While data-driven personas can be defined in several ways (Jansen et al. 2020; McGinn & Kotamraju 2008; Miaskiewicz & Luxmoore 2017; Mijač et al. 2018; Zhang et al. 2016), we crystallize the contemporary definition as follows: *a data-driven persona is a complete persona profile, created in a persona template using quantitative data about a given user population which is analysed using statistical techniques, including data science, and machine learning algorithms*. This definition, as can be seen, is heavily rooted in “quantitative data” and “algorithms,” to which we will return later.

It is standard that the research articles on data-driven personas (An, Kwak, Jung, et al. 2018; An, Kwak, Salminen, et al. 2018; Chapman et al. 2008; Goodman-Deane et al. 2018; Miaskiewicz & Luxmoore 2017; Mijač et al. 2018; Zhu et al. 2019) start out by criticizing manual

(“traditional”) persona creation methods, proposing data-driven personas as a remedy to shortcomings that include, *inter alia*, slowness, unreliability or risk of human analyst bias, small sample sizes and lack of representativeness, as well as un-reactiveness of the static personas created to the constant and dynamic changes in user behaviours, preferences, and characteristics (Chapman & Milham 2006; Howard 2015; Jansen et al. 2020; Salminen, Jansen, An, Kwak, et al. 2018).

Therefore, advocates of data-driven personas see these personas as superior to manually created personas, and as offering a *solution* to the problems of manual persona creation. Their argument is that personas are “fixed” when using “data” and “algorithms”. This kind of thinking poses hidden dangers. First, the goal of personas as “data-driven” (i.e., based on real user data, albeit often qualitative) was always present in the original conceptualization of Cooper (1999) and the further development of other HCI scholars (Anvari & Tran 2013; Nielsen 2019b; Pruitt & Grudin 2003; Turner & Turner 2011). Second, it is becoming increasingly clear from gained experience in the field as well as from the accumulated knowledge on over-reliance on “data” and “algorithms” that both the use and development of data driven personas are marred with challenges. These challenges logically evoke a variation of Dijkstra’s classic (Dijkstra 1968) question: Are data-driven personas considered harmful? In other words, should the research community pursue data-driven personas or are data-driven personas a dead end?

The purpose of this article is to highlight the variety and scope of challenges pertaining to data-driven personas. We not only consider the challenges with the development and evaluation of data-driven personas, as typically done in previous works (An, Kwak, Jung, et al. 2018; An, Kwak, Salminen, et al. 2018; Chapman et al. 2008) – where it is presumed that data-driven personas, as statistical creations, are primarily technical artefacts that should be judged by technical merits and metrics – but we also consider the larger schemes and ramifications of data-driven persona lifecycles, including how they fit into the HCI research community (which is predominantly qualitatively oriented when it comes to personas) and how they fit into organizations that, on one hand, tend to admire “data” and “algorithms” and, on the other hand, struggle to capture real value using data-driven personas.

We adopt the lifecycle view of personas (Adlin & Pruitt 2010), discussing challenges pertaining to different stages of the data-driven persona project, including its initiation, persona adoption, support, and general impediments that data-driven personas “inherit” from the essential nature of personas. Many observations in this manuscript are based on the authors’ encounters with HCI reviewers, especially those expressing scepticism towards data-driven personas. Due to anonymity, we cannot know the backgrounds of those reviewers, but based on their comments, it is often logical to assume that they come from a different tradition of creating and using personas and might, in some cases, be threatened by “data” and “algorithms” (we are basing this argument on the tone of some the reviews we have received over the years). In our opinion, it is important to make these points of criticism visible and put them forward to critically assess the current state and the future of the research on data-driven personas.

RELATED LITERATURE

Conceptually, persona criticism can be divided into three types. First, there is criticism towards personas as a design technique in general. This form of criticism applies to all types of personas, including manually created personas, data-driven personas, and mixed-method personas. Second, there is approach-specific criticism, e.g., that manually created personas are often based on low sample sizes (Chapman & Milham 2006). Third, there is method-specific criticism within

a specific approach, e.g., that K-means clustering would not be optimal for data-driven personas because it assigns each demographic group only to a single cluster (Kwak et al. 2017).

The literature does not always explicitly state the level of criticism it is engaged in. Hence, when reviewing the challenges related to data-driven personas, we need to consider if the specific points of criticism are valid for data-driven personas. In this brief literature review, we summarize well-cited articles presenting persona criticism, and relate that criticism to data-driven personas. The following section will dig deeper into this criticism. Additional resources for the reader include a literature review of quantitative persona creation (Salminen, Guan, Jung, et al. 2020) and a textbook focused on data-driven personas (Jansen et al. 2021).

One of the most cited persona critiques was put forth by Chapman and Milham (2006). While their criticism focused on several aspects that data-driven personas claim as benefits over manual personas (e.g., low sample size, staling), some of the concerns apply to data-driven personas as well. These include, at least, the *inconsistency problem* (one part of persona profile information can be from Source A and another from Source B and these may or may not refer to the same users) and the *granularity problem* (increasing the number of persona attributes requires more personas to be created in order to cover all possible segments).

Salminen, Jung, and Jansen (2020) mentioned ‘three Es’ as general challenges of personas: *Envision* (personas have no direct relationship to real user data), *Execution* (quality of the generated personas is low or unknown), and *Evaluation* (the success of personas is based on anecdotal feedback). The latter two can be considered as relevant concerns for data-driven and manual personas alike. In addition, Salminen, Guan, Jung, et al. (2020) mention the following challenges of quantitative persona creation: (1) lack of standards and best practices, (2) lack of ethical considerations, and (3) loss of immersion. These are critical issues that we expand on in the following section.

Multiple authors discuss the political challenges of persona use, particularly stereotyping (Marsden & Haag 2016; Rönkkö et al. 2004; Turner & Turner 2011). These concerns do not magically disappear with “data” and “algorithms” but can, in fact, become even more accentuated if the persona generation is done precariously or the results taken at face value. Therefore, these issues remain relevant for data-driven personas.

Howard (2015) posed the overarching question, “Are personas really usable?”. The crux of his criticism is that, although personas were originally introduced to facilitate communication among team members in UCD, in reality, personas do not solve the communication problems, but may even lead to further misunderstandings. A similar conclusion was made by Friess (2012) who, based on an ethnographic study, reported that designers rarely evoke or mention personas in their daily jobs, and by Matthews et al. (2012) whose participants found personas as abstract and misleading. Finally, De Voil (2010) raises several key issues regarding the concept of personas, proposing that personas are artificial thinking aids with severe limitations. These concerns remain topical for data-driven personas, and we will expand on them in the following section.

None of the criticism for data-driven personas in the past literature is comprehensive. Instead, the levels of criticism are often mixed, and most often the criticism focuses either on manually created personas or on personas as a user-centric design technique. While we do not claim to unveil all the challenges of data-driven personas in this article, we nonetheless make an analysis of how the central challenges regarding personas in general manifest in data-driven personas.

The challenges of data-driven personas are both theoretical and painstakingly practical, and the HCI community should be aware of them. It is our purpose to summarize the main ones and raise some novel ones that touch upon the transition of data-driven personas from “one-off exercises” into interactive persona systems (Jung et al. 2017, 2020; Jung, Salminen, Kwak, et al. 2018; Jung, Salminen, An, et al. 2018). This transition of personas moving from “static” layouts/templates/posters to interactive persona systems imposes some novel challenges regarding user experience (UX), user interfaces (UI), and interaction techniques) that the previous literature has not made explicit. This article opens discussion on those challenges, further expanding the list of “grand challenges” faced by data-driven personas.

APPROACH

We organize this work by themes, which represent central issues with data-driven personas. These are derived from authors’ prior experience with data-driven persona research (more than a dozen published papers, including several CHI publications) and with helping more than half a dozen client organizations in industries such as news and media, telecommunications, airline, e-commerce, and design/UX research, to apply data-driven personas. Drawing from this experience-based knowledge, we formulate central arguments as to why data-driven personas could be considered harmful. In deference to alternative views, we also solicited comments from two external researchers known to have extensive experience in persona research (one in qualitative personas and the other in quantitative personas). Their comments were incorporated in honing the arguments made in this manuscript. Finally, our work is based on perusing a large body of literature on data-driven personas (e.g., Adlin & Pruitt 2009; Goodman-Deane et al. 2018; Guo & Razikin 2015; McGinn & Kotamraju 2008; Miaskiewicz & Luxmoore 2017; Mijač et al. 2018; Watanabe et al. 2017; Zhang et al. 2016; Zhu et al. 2019 and many others) over a period of multiple years.

WHY ARE DATA-DRIVEN PERSONAS HARMFUL?

Starting from project definition and extending to persona data collection, creation, evaluation, and their eventual adoption in organizations, many things can go wrong with data-driven personas. We adopt a “lifecycle view” (cf. Adlin & Pruitt 2009, 2010) to inspect these challenges that have an adverse effect on the decision to undertake a data-driven persona project. In other words, the following challenges consider the data-driven persona project’s (a) *initiation* (expectations, objectivity, standards), (b) *adoption* (user perceptions and use), and (c) *support* (training, maintenance), as well as (d) *general impediments* (superficiality, aggregation, averages, and relevance).

Challenge 1: Inflated Expectations from Stakeholders

A common, yet often discarded aspect of applying algorithms for persona creation, is that the average stakeholder in a company attributes mythical properties and capabilities to these technological inventions, so that the mere mentioning of “data”, “algorithms”, or “non-negative matrix factorization” evokes positive qualities such as trustworthiness, efficiency, and transcendence of human capabilities. This effect, dubbed the “mystique of numbers” by Siegel (2010, p. 4721), refers to the phenomenon wherein stakeholders have unrealistic expectations from data-driven personas. As soon as stakeholders are informed that data-driven personas are based on “real data” and “millions of user interactions” which are analysed by (the implicitly objective) “algorithms”, they abandon their critical attitudes and become willing to take personas seriously. While this effect is beneficial for persona adoption, which is typically hindered by the lack of credibility, stakeholder commitment, and trust in the personas (Friess

2012; Jensen et al. 2017; Matthews et al. 2012; Nielsen 2019a; Rönkkö 2005; Rönkkö et al. 2004; Seidelin et al. 2014), it comes with the negative side-effect of hyperbolic expectations.

In the long run, the stakeholders' unrealistic expectations may result in various adverse outcomes. These include, e.g., disappointment in the fact that data-driven personas did not solve all analytics problems despite the superhuman capabilities of the algorithms. Similarly, there is a risk that hidden errors in the data, algorithms, or simply the misunderstandings of what certain data is and how it is created in the persona profile, skew the stakeholders' decision-making process thus defeating the original purpose of data-driven personas which is to provide valid, correct, and accurate information for stakeholders to consider real user needs, wants, interests, and goals.

Stakeholders may believe that statistical methods may simply be selected and applied to get "*the answer*," i.e., the immutable truth of their users (whereas, in reality, the truth is much more nuanced than what algorithms reveal). This is paired with a strong, but almost always unstated assumption that distinct types of people must exist. The contrary assumption, that people are approximately multivariate, normally distributed, and do not fall into neatly separable groups, is typically rejected from the outset. An analogy is slicing a pizza: there are infinite ways to do it, and none is correct or incorrect; it all depends on one's goal (Chapman & Feit 2019). As such, the data analysis efforts involved in data-driven persona creation can be more or less successful, but none of them is the only and perfect solution.

Furthermore, a predominant focus on statistical significance in data-driven persona creation may overlook the personas' practical significance, as these two concepts do not always equate in the real world. For example, there may be a statistically significant difference between two user groups with a low magnitude (Jansen et al. 2019), rendering this difference unimportant for decision making. Technically-oriented persona creators may want to optimize the accuracy or validity of the personas based on some metric to minimize or maximize, whereas stakeholders would want to optimize the usefulness of the personas, regardless of how they are created. A crucial question for the purpose of usefulness maximization is: are the similarities and differences among the personas truly so important that they matter for decision making? Such considerations are often omitted when reporting data-driven personas in academic literature. Consequently, data-driven personas may end up being abstract and esoteric — i.e., technically complex and difficult to communicate to stakeholders in ways that are both truthful and easy to understand (Salminen, Jung, & Jansen 2020).

In summary, the present ability to generate data-driven personas does not match the expected perfection, meaning that *there may be a gap of what the stakeholders think they get and what they actually get with data-driven personas*.

Challenge 2: Algorithms are Biased Too

Data-driven personas can be seen as design artefacts created by algorithms. As such, they are susceptible to what is known as algorithmic bias (Friedman & Nissenbaum 1996; Hajian et al. 2016), a tendency of algorithms to accentuate the properties of the data while ignoring fairness or legality of the outcomes. An example would be an algorithm repeatedly picking African-American names for criminal personas created from the data (Salminen, Froneman, Jung, et al. 2020).

In general, there are three sources of bias in algorithmic systems: (1) imbalanced/skewed datasets that "favour" one user group over another; (2) mathematics of the algorithm that accentuate the differences among the groups by "picking" certain groups over others; and (3) cultural assumptions that are encoded in datasets and systems, leading to

systematic discrimination by structural design (Hajian et al. 2016). Data-driven personas are not immune to these concerns. In fact, “data-driven,” when blindly applied, can unintentionally become “bias-driven”. Therefore, following the on-going research in the ethical analysis of algorithms (Eslami et al. 2018), an ethical review of data-driven persona development is necessary.

While research papers may claim that data-driven persona development increases objectivity (Jansen et al. 2020; Mijač et al. 2018), the deployment of algorithms for data analysis may present new sources of prejudice and lack of transparency (Salminen, Santos, Jung, et al. 2019). Additional ethical challenges include safeguarding the privacy of online users and giving stakeholders information and tools to assess how reliable and trustworthy the data-driven personas are (which is a non-trivial problem as the technical sophistication of end-users of personas greatly varies). Thus far, research on ethics in data-driven personas is scarce, with the exception of a couple of studies (Goodman-Deane et al. 2018; Salminen, Froneman, Jung, et al. 2020). It is uncertain if data-driven persona advocates recognise these ethical issues in their work, as most studies simply lack the discussion. For example, replacing the persona generation algorithm can have a drastic effect on the generated personas, even when the underlying data is the same (Brickey et al. 2012) and yet, there is virtually no work comparing what kind of personas different algorithms generate from the same user data. Thus, it is uncertain if data-driven personas can become biased and if they can, how can the issue be effectively addressed?

Challenge 3: Where are the Standards?

Data-driven personas paradoxically suffer from a lack of standards. The lack of standards is paradoxical because, being the result of quantitative data and objective/replicable processes, data-driven personas are, in theory, in a perfect position for standards to emerge. Yet, there are no standards or metrics even for measuring such a basic concept as persona quality, which would be fundamental for comparing and ranking different data-driven methods.

Unlike in computer science where researchers run experiments on baseline datasets that are the same for everyone, no baseline datasets exist for persona creation. Unlike in fields like psychology, where there are studies on norms of perception – e.g., how certain groups by age, gender, or culture view the world (Gosling et al. 2003) – data-driven persona studies propose no such norms or even discuss them. Hence, it is difficult to understand what features and expectations users have for data-driven persona systems. The lack of standardization also makes it difficult to obtain strong guidelines for persona creation that would be derived from empirical research. There are no empirically validated guidelines, for example, as to how many personas should be created, what metrics should be used to evaluate the personas, what ethical considerations should be made when collecting and processing data for persona generation, and so on.

Also, although data-driven personas could be generated from many alternative metrics to describe different behaviours (e.g., clicking behaviour, viewing behaviour, purchase behaviour), typically, studies use only one behavioural interaction metric at a time (An, Kwak, Salminen, et al. 2018). *Which metric(s) to choose, then?* This issue is akin to that in the field of analytics, where stakeholders need to define their questions well to avoid getting lost in the dozens of reports afforded by the analytics systems. For data-driven personas, there exists virtually no guidance for this metrics selection problem, but researchers and practitioners carry out the selection in an ad-hoc manner.

The lack of standards hinders data-driven persona creation (the choice of methods is unclear, as is the mutual comparison of methods), use (what are the standard use cases for data-

driven personas?), and understanding of the data-driven persona user behaviour (how many personas do user view? How long they spend, on average, on persona profiles? What information is the most crucial for decision making?). Apart from limited exploratory work on these matters (Salminen, Kathleen Guan, Jung, Chowdhury, et al., 2020; Salminen, Nielsen, Jung, An, et al., 2018; Salminen, Willemien Froneman, Jung, Chowdhury, et al., 2020; Salminen, Ying-Hsang Liu, Sengun, Santos, et al., 2020), no convincing standards for data-driven persona user behaviour have been developed to date.

Challenge 4: Mess, Confusing, and Difficult to Use

User studies report many issues with data-driven persona UX and UI (Salminen, Jung, An, et al. 2019; Salminen, Jansen, An, Jung, et al. 2018; Salminen, Jung, An, Kwak, et al. 2018; Salminen, Sengun, Jung, et al. 2019). These issues include, at least, confusion over what the information in persona profiles is and how it is generated (lack of transparency), how to get more information about a specific persona, questions about the reliability and trustworthiness of the information, and – the most vital question of all – *“Now what? How can I use this persona?”*.

According to our experience, stakeholders struggle to make use of persona systems, even when they are provided with multiple features, such as interest prediction, gap analysis, and search and navigation (Jansen et al. 2020). These features may appear unfamiliar to persona users, and it may be that it is more important for design outcomes that personas are inspirational and memorable rather than numerical and accurate. In this light, the definite proof of value for data-driven personas remains elusive. Moreover, at this stage, research on effective UIs for data-driven personas is still in its infancy (Salminen, Liu, Sengun, et al. 2020), and there is little empirical evidence about how stakeholders interact with these systems, what features are requested, and so on. It is fairly easy to generate proof-of-concepts (Mijač et al. 2018), but the leap from these prototypes into full-fledged production systems with an active user base is still in the horizon. Therefore, making data-driven personas user friendly and useful remains an obstacle for their wider application.

Challenge 5: Superficial and Unsurprising

It can be said there is a consensus among qualitative persona researchers that, even if not always obtained, the goal and purpose of personas is to provide in-depth understanding of different user types, that is to facilitate the sense of empathy (Blomquist & Arvola 2002; Haag & Marsden 2019; Nielsen et al. 2017; Nielsen & Storgaard Hansen 2014; Wright & McCarthy 2008). These insights are, on one hand, the result of the creation process itself; by immersing oneself into the user data, one achieves a thorough understanding of the user’s circumstances. On the other hand, gaining such insights relies on the innate ability of humans to understand other humans (Grudin 2006). Algorithms cannot think and, hence, they cannot compete with this ability.

Thus, a major concern with data driven personas is that the algorithms behind their generation often lack the ability to interpret, decipher, and encode common sense meanings. Cultural meanings and (tacit) distinctions are difficult even for untrained humans, and a data-driven persona algorithm is completely oblivious to them unless – with some method that has not been created yet – trained to classify information based on its cultural meaning. Cultural factors are lacking in the data-driven persona literature, despite an extensive body of literature on culture-cognizant application of manually created personas (Anvari et al. 2019; Jensen et al. 2017; Nielsen 2010). While data-driven persona profiles can include social media comments (Salminen, Şengün, Kwak, Jansen, et al. 2018), they cannot disambiguate their meanings or make any complex interpretations from these comments.

Finally, persona enrichment poses an issue since it tends to require using independent datasets (Mijač et al. 2018), evoking the consistency problem (see p. 1). Yet, without in-depth insights, data-driven personas risks remaining shallow alternative UIs for website analytics data, having little actionable information.

Challenge 6: Aggregation Makes Things Worse

Chapman and Milham (2006) were first to articulate the aggregation problem of personas. Later, other researchers have observed this issue (Bødker et al. 2012; Matthews et al. 2012; Salminen, Jansen, An, Kwak, et al. 2018); yet, no definitive solutions have been proposed. Personas are, by definition, aggregates: they group individual users to one user representation. Yet, each user is unique and different from others (sometimes referred to as segment of one [Lingel 2012]).

Chapman et al. (2008) analysed data-driven personas and found that the more granular representations of users we want, the more personas are needed. For example, if we want to represent users by gender, two personas (male and female) may be sufficient. However, if we want to represent both gender and age, assuming two age categories, we now need twice the number of personas: male-young, male-mature, female-young, and female-mature. The issue is that the selection of granularity of personas is arbitrary and there are no rules for deciding this granularity.

Another issue of data-driven personas is their potential weakness against the argument often mentioned by practitioners, “With individualized data, I can target the individual users, so why would I need personas/segments/clusters/etc.?”. This question is valid, as in use cases such as personalization and recommendation systems, the unit of analysis is the individual and the decision-maker is an algorithm (albeit, it is also true that many of such algorithms rely on dimensionality reduction, which is a form of grouping [Huang et al. 2019]). As the world moves towards automated decision-making, *is there room for data-driven personas?*

Challenge 7: The Average Persona Does Not Exist

The typical definition of personas is that they describe typical users (Marsden & Haag 2016; Sakata et al. 2014). Therefore, creating an average persona is the conceptual and practical default of many persona-creation projects. Its challenges relate, firstly, to stereotyping when focusing on the mean/average user (Marsden & Haag 2016; Marsden & Pröbster 2019; Turner & Turner 2011) and, secondly, to the focus of data-driven algorithms on the central tendency in the data.

What we mean by this can be illustrated with a simple example. Assume two datapoints about users, with numerical values of “1” and “5”. Their average is “3” which is equally far from both observations and thus does not well represent either datapoint. This “flaw of averages” is well documented in a classic study conducted by the United States Air Force in 1950, finding that, among 4,000 measured pilots, no pilots matched all the average attributes of height, weight, etc. (Hertzberg et al. 1954). The problem with mean-centred personas (i.e., those that describe average, typical users) is the general problem with the mean: if half of your users are right-handed and half are left-handed, should your persona be middle-handed? Obviously not. Instead, you need personas for both left- and right-handed users. This is what we mean we talk about diversity of personas – a good persona set is one that covers various user types, not only their hypothetical amalgamation.

Yet, by picking “representative” behaviours and characteristics for the data-driven personas, we tend to overlook the extremes. These extremes, anomalies, deviations, minorities,

and fringe groups are, therefore, not considered by the stakeholders using the data-driven personas, as these segments are hidden; they do not exist, as far as the stakeholder is considered (Salminen, Froneman, Jung, et al. 2020). This rounding up of characteristics may end up with eliminating everything that makes a user unique, resulting in bland and unimaginative user profiles that feed rather than curb stereotypes. Therefore, representativeness comes at the cost of diversity.

Challenge 8: Maintenance Cost

Unlike traditional personas that are created once and then used for some time, data-driven personas require constant nurturing, care, and maintenance. This maintenance is costly and time-consuming. The reason for maintenance stems from the reliance on live datasets (Jung, Salminen, An, et al. 2018). As platforms such as YouTube and Facebook repeatedly change their terms of service and APIs, often without proper documentation or notifications for developers, persona systems reliant on these data sources are “broken” until the necessary updates are made.

Similarly, software packages and algorithms are frequently updated, requiring the data-driven persona developers to monitor and implement these updates to ensure the continued functionality of the system. Thus, unlike traditional personas that are independent and contained, data-driven personas tend to have complex linkages to sub-systems, data science libraries, and Web technologies that come with a built-in technical debt (Thomas et al. 2018).

Related problems are missing data, unknown measurement errors in data exports, sampling/thresholds that limit the data collection speed and may skew the data distributions, and the adding/removing of data variables and classes by the online platforms without providing any say to researchers on these decisions. When personas are built around data sources owned by multi-national corporations such as Facebook and Google, the dependence on the goodwill of these organizations to continue sharing user data is high. If these platforms were to consider it strategically unwise to continue sharing data via their APIs, data-driven personas would be quickly broken.

Finally, privacy issues such as the General Data Protection Regulation (GDPR) in the European Union and similar legislative initiatives in other economic areas may further limit the availability of user data for applications such as data-driven persona generation in the future. While the benefits of data-driven personas can be seen in the abundance of online user data and the effect of democratizing personas for even smaller organizations that can have access to this data, a future scenario where the data becomes less accessible and perhaps only accessible for large corporations against payment can be envisioned. Future developments bring forth this cloud of uncertainty for data-driven applications such as online user personas.

Challenge 9: Personas are Passé!

A compelling argument against data-driven personas that one often reads in some reviews and online discussions among UX professionals is that personas are not relevant anymore and organisations are using other methods. Blažica (2014) surveyed start-up companies about their use of UX techniques and observed that personas ranked the fourth last (out of eight techniques) in terms of stakeholder familiarity and also fourth last in terms of regular use. While ten respondents indicated that they had used personas “a few times”, only one respondent reported regular use. This, indeed, warrants concern as personas did relatively poorly compared to other methods.

If there is general redundancy for personas among HCI professionals, then this sentiment is inherited to data-driven personas that, after all, are personas. Indeed, multiple alternative techniques for UCD exist — e.g., interviews, focus groups, surveys, participant observation, user narratives, jobs-to-be-done, scenarios, customer journeys, and so on (Blažica 2014; Carroll 1997; Goodman et al. 2013; Kliman-Silver et al. 2020). Similarly, there are a plethora of analytics tools that provide numerical information about users in the form of charts, numbers, and tables (e.g., YouTube Analytics, Google Analytics, Facebook Insights, IBM Analytics, etc.). So, *why are data-driven personas needed?*

| Challenge | Applies to... | |
|--|----------------------|---|
| | Data-driven personas | All personas (also ones created manually) |
| C1: Inflated expectations | x | |
| C2: Algorithmic bias | x | |
| C3: Lack of standards | x | x |
| C4: User perceptions and difficulty of use | x | x |
| C5: Superficiality | x | |
| C6: Problem of aggregation | x | x |
| C7: Problem of averages | x | x |
| C8: Maintenance cost | x | |
| C9: Irrelevance | x | x |

Table 1: Data-driven personas inherit challenges from the concept of personas but there are also challenges unique to them.

DISCUSSION

We presented nine challenges (see Table 1) that might imply that data-driven personas are harmful. These challenges are far more serious than generally believed. They start with wrong assumptions from the stakeholders, and extend to precarious application of methods, imbalanced or messy data, access to superficial data only, lack of communication how they were created, too complex UIs, unclear or lacking definitions of persona content, and omission of ethical considerations. The crucial message of our work is that the state-of-the-art of data-driven personas does not create perfect personas, despite the somewhat illusionary and impressive use of technical jargon such as “data”, “algorithms”, and so on.

Are data-driven personas worthwhile, then? Indeed, at first glance, the challenges may seem overwhelming. It is, in any case, certain that no single paper or research project can solve them. For the discovery of solutions, researchers within the persona domain need to work in unison. Probabilistic methods can assist with the aggregation problem (Chapman et al. 2015). Other potential solutions involve developing standards for the choice process for hyperparameters (most importantly, the number of personas) and for evaluation metrics that need to be consistent among the numerous data-driven persona methodologies.

It can also be said that ‘no persona is an island’, meaning that data-driven personas co-exist and co-evolve together with computer science, HCI, and other related fields. For example, studies in algorithmic bias (Diaz et al. 2018; Hajian et al. 2016; Salminen, Jung, & Jansen 2019) apply to data-driven personas, for example, through the process of selecting persona name, persona’s demographic traits, and interpreting persona’s sentiment by using tools of Natural Language Processing (NLP). Data-driven personas lay on a foundation of algorithmic and technological work, implying that their future is intertwined with the progress in these fields that support and enable the technical back-end of data-driven personas.

CONCLUSION

Data-driven personas may provide many benefits relative to manually created personas. However, the implicit assumption that data-driven personas would be “perfect” or “easy” is not correct. On the other hand, organizations show a continuous interest in data-driven personas, albeit often with unrealistic expectations. Hence, the research efforts in this space are valuable and worthwhile. The future will reveal if data-driven personas remain at the level of perpetual promise or, if at some point, they redeem the high expectations that their (small but persistent) group of advocates claim.

WORKS CITED

- Adlin T & Pruitt J 2009, ‘Putting personas to work: Using data-driven personas to focus product planning, design, and development; in Sears A and Jacko JA (eds) *Human-Computer Interaction: Development Process*, CRC Press, New York, pp. 95–120.
- 2010, *The Essential Persona Lifecycle: Your Guide to Building and Using Personas*. 1st ed. Morgan Kaufmann Publishers Inc. San Francisco.
- An J, Kwak H, Jung S-G, Salminen J & Jansen BJ 2018, ‘Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data’, *Social Network Analysis and Mining*, vol. 8, no. 1: 54. DOI: 10.1007/s13278-018-0531-0.
- An J, Kwak H, Salminen J, Jung S-G & Jansen BJ 2018, ‘Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data’, *ACM Transactions on the Web (TWEB)*, vol. 12, no. 4: 27. DOI: 10.1145/3265986.
- Anvari F & Tran HMT 2013, ‘Persona ontology for user centred design professionals’, in *The ICIME 4th International Conference on Information Management and Evaluation*, Ho Chi Minh City, Vietnam, 2013, pp. 35–44.
- Anvari F, Richards D, Hitchens M & Tran H 2019, ‘Teaching user centered conceptual design using cross-cultural personas and peer reviews for a large cohort of students’, in *Proceedings of the 41st International Conference on Software Engineering: Software Engineering Education and Training*, IEEE Press, Piscataway, NJ, 2019, pp. 62–73.
- Blažica B 2014, *Use of UX and HCI tools among start-ups*. Working paper. Ljubljana, Slovenia: XLAB Research.
- Blomquist A & Arvola M 2002, ‘Personas in action: Ethnography in an interaction design team’, in *Proceedings of the second Nordic conference on Human-computer interaction*, ACM Press, New York, pp. 197–200.
- Bødker S, Christiansen E, Nyvang T & Zander P 2012, ‘Personas, people and participation: challenges from the trenches of local government’, in *Proceedings of the 12th Participatory Design Conference on Research Papers: Volume 1 - PDC '12*, ACM Press, Roskilde, p. 91. ACM Press. DOI: 10.1145/2347635.2347649.

- Brickey J, Walczak S & Burgess T 2012, 'Comparing Semi-Automated Clustering Methods for Persona Development', *IEEE Transactions on Software Engineering*, vol. 38, no. 3, pp. 537–546. DOI: 10.1109/TSE.2011.60.
- Carroll JM 1997, 'Chapter 17 - Scenario-Based Design', in Helander MG, Landauer TK, and Prabhu PV (eds) *Handbook of Human-Computer Interaction (Second Edition)*. Amsterdam: North-Holland, pp. 383–406. DOI: 10.1016/B978-044481862-1.50083-2.
- Chapman C & Feit EM 2019, *R For Marketing Research and Analytics*. Springer.
- Chapman C & Milham RP 2006, 'The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method', in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1 October 2006, pp. 634–636. DOI: 10.1177/154193120605000503.
- Chapman C, Love E, Milham RP, ElRif P & Alford J 2008, 'Quantitative Evaluation of Personas as Information', in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1 September 2008, pp. 1107–1111. DOI: 10.1177/154193120805201602.
- Chapman C, Krontiris K & Webb J 2015, 'Profile CBC: Using Conjoint Analysis for Consumer Profiles', in *Sawtooth Software Conference Proceedings*, Google Research. Available at: <https://research.google.com/pubs/archive/44167.pdf>.
- Cooper A 1999, *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity*. 1st ed. Sams-Pearson Education, Indianapolis.
- De Voil N 2010, *Personas considered harmful*. Industry report. London, UK: De Voil Consulting. Available at: <http://www.devoil.com/papers/PersonasConsideredHarmful.pdf>. Retrieved November 2019.
- Diaz M, Johnson I, Lazar A, Piper A & Gergle D 2018, 'Addressing Age-Related Bias in Sentiment Analysis', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, pp. 1–14. CHI '18. Association for Computing Machinery. DOI: 10.1145/3173574.3173986.
- Dijkstra EW 1968, 'Letters to the editor: go to statement considered harmful', *Communications of the ACM*, vol.11, no. 3, pp.147–148.
- Eslami M, Krishna Kumaran SR, Sandvig C & Karahalios K 2018, 'Communicating Algorithmic Process in Online Behavioral Advertising', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, Montréal, p. 432.
- Friedman B & Nissenbaum H 1996, 'Bias in computer systems', *ACM Transactions on Information Systems (TOIS)*, vol.14, no. 3, pp. 330–347.
- Friess E 2012, 'Personas and decision making in the design process: an ethnographic case study', in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1209–1218. DOI: <https://doi.org/10.1145/2207676.2208572>.
- Goodman E, Kuniavsky M & Moed A 2013, *Observing the User Experience: A Practitioner's Guide to User Research*. Morgan Kaufmann.
- Goodman-Deane J, Waller S, Demin D, González de Heredia A, Bradley M & Clarkson J 2018, 'Evaluating Inclusivity using Quantitative Personas', in *In the Proceedings of Design Research Society Conference 2018*, Limerick. DOI: 10.21606/drs.2018.400.
- Gosling SD, Rentfrow PJ & Swann WB 2003, 'A very brief measure of the Big-Five personality domains', *Journal of Research in personality*, vol. 37, no.6, pp: 504–528.
- Grudin J 2006, 'Why Personas Work: The Psychological Evidence', in Pruitt J and Adlin T (eds) *The Persona Lifecycle*. Elsevier, pp. 642–663. DOI: 10.1016/B978-012566251-2/50013-7.
- Guo H & Razikin KB 2015, 'Anthropological User Research: A Data-Driven Approach to Personas Development', in *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, OzCHI '15*. ACM, New York, pp. 417–421. DOI: 10.1145/2838739.2838816.
- Haag M & Marsden N 2019, 'Exploring personas as a method to foster empathy in student IT design teams', *International Journal of Technology and Design Education*, vol. 29, no. 3, pp.565–582. DOI: 10.1007/s10798-018-9452-5.

- Hajian S, Bonchi F & Castillo C 2016, 'Algorithmic Bias: From Discrimination Discovery to Fairness-aware Data Mining', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. Association for Computing Machinery San Francisco, pp. 2125–2126. DOI: 10.1145/2939672.2945386.
- Hertzberg HT, Daniels GS & Churchill E 1954, *Anthropometry of Flying Personnel-1950*. Antioch, Yellow Springs.
- Howard TW 2015, 'Are Personas Really Usable?', *Communication Design Quarterly Review*, vol. 3, no. 2, pp. 20–26. DOI: <https://doi.org/10.1145/2752853.2752856>.
- Huang X, Wu L & Ye Y 2019, 'A Review on Dimensionality Reduction Techniques', *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 10. DOI: 10.1142/S0218001419500174.
- Jansen BJ, Jung S-G & Salminen J 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, pp. 1–6. DOI: 10.1145/3290607.3313006.
- Jansen BJ, Salminen J, Jung S-G, & Guan K 2021, *Data-Driven Personas*. 1st ed. Synthesis Lectures on Human-Centered Informatics. Morgan & Claypool Publishers. Available at: <https://www.morganclaypool.com/doi/abs/10.2200/S01072ED1V01Y202101HCI048> (accessed 10 February 2021).
- Jansen BJ, Salminen J & Jung S-G 2020, 'Data-Driven Personas for Enhanced User Understanding: Combining Empathy with Rationality for Better Insights to Analytics', *Data and Information Management*, vol. 4, no. 1, pp. 1–17. DOI: <https://doi.org/10.2478/dim-2020-0005>.
- Jensen I, Hautopp H, Nielsen L & Madsen S 2017, 'Developing international personas: A new intercultural communication practice in globalized societies', *Journal of Intercultural Communication*, vol. 43.
- Jung S-G, An J, Kwak H, Ahmad M, Nielsen L & Jansen BJ 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, pp. 1748–1755.
- Jung S-G, Salminen J, Kwak H & Jansen BJ 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, pp. 321–324. ACM. DOI: <https://doi.org/10.1145/3176349.3176893>.
- Jung S-G, Salminen J, An J & Jansen BJ 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, p. 2.
- Jung S-G, Salminen J & Jansen BJ 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, pp. 132–133. DOI: 10.1145/3379336.3381465.
- Kliman-Silver C, Siy O, Awadalla K, Lentz A, Convertino G & Churchill E 2020, 'Adapting User Experience Research Methods for AI-Driven Experiences', in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–8.
- Kwak H, An J & Jansen BJ 2017, 'Automatic Generation of Personas Using YouTube Social Media Data', in *Proceedings of the Hawaii International Conference on System Sciences (HICSS-50)*, Waikoloa, pp. 833–842.
- Lingel J 2012, 'Ethics and dilemmas of online ethnography', in *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 41–50.
- Marsden N & Haag M 2016, 'Stereotypes and politics: reflections on personas', in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 4017–4031.
- Marsden N & Pröbster M 2019, 'Personas and Identity: Looking at Multiple Identities to Inform the Construction of Personas', in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*, ACM Press, Glasgow, pp. 1–14. DOI: 10.1145/3290605.3300565.

- Matthews T, Judge T & Whittaker S 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, p. 1219. DOI: 10.1145/2207676.2208573.
- McGinn JJ & Kotamraju N 2008, 'Data-driven persona development', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Florence, pp. 1521–1524. DOI: 10.1145/1357054.1357292.
- Miaskiewicz T & Luxmoore C 2017, 'The Use of Data-Driven Personas to Facilitate Organizational Adoption—A Case Study', *The Design Journal*, vol. 20, no. 3, pp. 357–374.
- Mijač T, Jadrić M & Ćukušić M 2018, 'The potential and issues in data-driven development of web personas', in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1237–1242. DOI: 10.23919/MIPRO.2018.8400224.
- Nielsen L 2010, 'Personas in Cross-Cultural Projects', in Katre D, Orngreen R, Yammiyavar P, et al. (eds) *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts*. IFIP Advances in Information and Communication Technology. Springer, Berlin, pp. 76–82. DOI: 10.1007/978-3-642-11762-6_7.
- 2019a, 'Going Global—International Personas', in Nielsen L (ed.) *Personas - User Focused Design*. Human-Computer Interaction Series. Springer, London, pp. 123–133. DOI: 10.1007/978-1-4471-7427-1_7.
- 2019b, *Personas - User Focused Design*. 2nd ed. 2019 edition. Springer, New York.
- Nielsen L & Storgaard Hansen K 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, pp. 1665–1674.
- Nielsen L, Hansen KS, Stage J & Billestrup J 2015, 'A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations', *International Journal of Sociotechnology and Knowledge Development*, vol. 7, no. 1, pp. 45–61. DOI: 10.4018/ijskd.2015010104.
- Nielsen L, Jung S-G, An J, Salminen J, Kwak H & Jansen BJ 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*, Brisbane, ACM, Queensland, pp. 602–606. OZCHI '17. ACM. DOI: 10.1145/3152771.3156178.
- Pruitt J & Grudin J 2003, 'Personas: Practice and Theory', in *Proceedings of the 2003 Conference on Designing for User Experiences, DUX '03*. ACM, San Francisco, pp. 1–15. DOI: 10.1145/997078.997089.
- Rönkkö K 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. DOI: 10.1109/HICSS.2005.85.
- Rönkkö K, Hellman M, Kilander B & Dittrich Y 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, pp. 112–120. DOI: 10.1145/1011870.1011884.
- Sakata J, Zhang M, Pu S, Xing J & Versha K 2014, 'Beam: a mobile application to improve happiness and mental health', in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pp. 221–226.
- Salminen J, Jansen BJ, An J, Kwak H & Jung S-G 2018, 'Are personas done? Evaluating their usefulness in the age of digital analytics', *Persona Studies*, vol. 4, no. 2, pp. 47–65. DOI: 10.21153/psj2018vol4no2art737.
- Salminen J, Jansen BJ, An J, Jung S-G, Nielsen L & Kwak H 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, pp. 110–119. DOI: 10.1145/3176349.3176391.

- Salminen J, Jung S-G, An J, Kwak H & Jansen BJ 2018, 'Findings of a User Study of Automatically Generated Personas', in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*, ACM Press, Montreal, pp. 1–6. DOI: 10.1145/3170427.3188470.
- Salminen J, Jung S-G, An J, Kwak H, Nielsen L & Jansen BJ 2019, 'Confusion and information triggered by photos in persona profiles', *International Journal of Human-Computer Studies*, vol. 129, pp. 1–14. DOI: 10.1016/j.ijhcs.2019.03.005.
- Salminen J, Jung S-G & Jansen BJ 2019, 'Detecting Demographic Bias in Automatically Generated Personas', in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19. ACM, New York, pp. LBW0122:1-LBW0122:6. DOI: 10.1145/3290607.3313034.
- Salminen J, Nielsen L, Jung S-G, An J, Kwak H & Jansen BJ 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, DOI: 10.1145/3173574.3173891.
- Salminen J, Sengün S, Jung S-G & Jansen BJ 2019, 'Design Issues in Automatically Generated Persona Profiles: A Qualitative Analysis from 38 Think-Aloud Transcripts', in *Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, ACM, Glasgow, pp. 225–229. DOI: 10.1145/3295750.3298942.
- Salminen J, Şengün S, Kwak H, Jansen BJ, An J, Jung S-G, Vieweg S & Harrell DF 2018, 'From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas', *First Monday*, vol. 23, no. 6. DOI: 10.5210/fm.v23i6.8415.
- Salminen J, Santos Joao M., Jung S-G, Eslami M Jansen BJ 2019, 'Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas', *International Journal of Human-Computer Interaction*, vol. 0, no. 0, pp.1–13. DOI: 10.1080/10447318.2019.1688946.
- Salminen J, Guan Kathleen, Jung S-G, Chowdhury SA & Jansen BJ 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, pp. 1–14. DOI: <https://doi.org/10.1145/3313831.3376502>.
- Salminen J, Jung S-G & Jansen BJ 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)*, CEUR Workshop Proceedings, Cagliari, p. 7. DOI: urn:nbn:de:0074-2582-4.
- Salminen J, Jung S-G, Chowdhury SA, Sengün S & Jansen BJ 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, DOI: <https://doi.org/10.1145/3313831.3376770>.
- Salminen J, Liu Ying-Hsang, Sengün S, Santos JM, Jung S-G & Jansen BJ 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, pp. 357–368. DOI: <https://doi.org/10.1145/3377325.3377492>.
- Salminen J, Froneman Willemien, Jung S, Chowdhury S & Jansen BJ 2020, 'The Ethics of Data-Driven Personas', in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI '20. Association for Computing Machinery, Honolulu, pp. 1–9. DOI: 10.1145/3334480.3382790.
- Seidelin C, Jonsson A, Høgild M, Rømer J & Diekmann P 2014, 'Implementing personas for international markets: a question of UX maturity', in *Proceedings at SIDER*.
- Siegel DA 2010, 'The Mystique of Numbers: Belief in Quantitative Approaches to Segmentation and Persona Development', in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10. ACM, New York, pp. 4721–4732. DOI: 10.1145/1753846.1754221.

- Thomas TW, Tabassum M, Chu B & Lipford H 2018, 'Security during application development: An application security expert perspective', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12.
- Turner P & Turner S 2011, 'Is stereotyping inevitable when designing with personas?' *Design studies*, vol. 32, no. 1, pp. 30–44.
- Watanabe Y, Washizaki H, Honda K, Noyori Y, Fukazawa Y, Morizuki A, Shibata H, Ogawa K, Ishigaki M, Shiizaki S, Yamaguchi T & Yagi T 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, pp. 49–55. DOI: 10.1109/CHASE.2017.9.
- Wright P & McCarthy J 2008, 'Empathy and Experience in HCI', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08. ACM, Florence, pp. 637–646. DOI: 10.1145/1357054.1357156.
- Zhang X, Brown H-F & Shankar A 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, pp. 5350–5359.
- Zhu H, Wang H & Carroll JM 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, pp. 61–70. DOI: 10.1145/3322276.3322285.