Rethinking Personas for Fairness: Algorithmic Transparency and Accountability in Data-Driven Personas

Joni Salminen[1,2(✉)], Soon-gyo Jung[1], Shammur A. Chowdhury[1], and Bernard J. Jansen[1]

[1] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
jsalminen@hbku.edu.qa
[2] University of Turku, Turku, Finland

Abstract. Algorithmic fairness criteria for machine learning models are gathering widespread research interest. They are also relevant in the context of data-driven personas that rely on online user data and opaque algorithmic processes. Overall, while technology provides lucrative opportunities for the persona design practice, several ethical concerns need to be addressed to adhere to ethical standards and to achieve end user trust. In this research, we outline the key ethical concerns in data-driven persona generation and provide design implications to overcome these ethical concerns. Good practices of data-driven persona development include (a) creating personas also from outliers (not only majority groups), (b) using data to demonstrate diversity within a persona, (c) explaining the methods and their limitations as a form of transparency, and (d) triangulating the persona information to increase truthfulness.

Keywords: Personas · Data · Fairness · Algorithms · Transparency · Ethics

1 Introduction

Personas are a user-centric design technique made popular by Cooper [1] in software development and in Human-Computer Interaction (HCI). Personas are defined as profiles of people that fictive but realistic [2]. They embody central aspects of the user or customer segments they describe, giving "faces" to user data [3] and summarizing diverse and complex audiences into a few archetypes [4, 5].

A persona profile typically includes a name, a picture, and a description detailing the attitudes and behaviors of the user segment the persona portrays [6]. Personas have consistently been used in a variety of fields, including software development and design [7], marketing [8], and health informatics [9].

Data-driven personas (DDPs) are created using digital user data and quantitative methods. DDPs are usually evaluated for aspects of usability, user experience, and value [10], but these evaluations often tend to overlook ethical aspects, e.g., fairness, privacy, transparency, and trust [11]. These ethical considerations have been

H. Degen and L. Reinerman-Jones (Eds.): HCII 2020, LNCS 12217, pp. 82–100, 2020.
https://doi.org/10.1007/978-3-030-50334-5_6

acknowledged as essential for algorithmic decision-making systems [11–14] and have been explored in the broader persona literature [15–17] but not for DDPs specifically.

The consensus of research in tangential fields of computer science and HCI has shown that algorithmic systems, and machine learning (ML) in general, involve various ethical issues. As ML and automation are becoming more widely used for persona creation [3, 18, 19], these ethical considerations warrant an inquiry in the context of DDPs. Yet, thus far authors have evaluated personas as quantitative information [20], not as products embedded in ethical and political contexts (i.e., the real world). The tangential research in related fields on algorithmic fairness [21, 22] makes it strikingly obvious that researchers and practitioners should acknowledge ethical guidelines when creating personas using algorithms. The lack of such guidelines forms a critical research gap that we begin to address in this manuscript.

The HCI community is becoming increasingly aware of algorithmic biases. For DDPs, this means that data and algorithms may introduce undesired generalizations or preconceptions into the personas. Relying solely on quantitative data might, e.g., lead to ignoring minority groups and inclusivity, as statistical methods tend to "favor" majority groups and obscure the outliers and deviations within user groups. By implicitly assuming information of certain segments of a population, the use of DDPs may reinforce existing patterns of social advantage or produce new ones [23].

To this end, there is a need to investigate how ethical considerations appear in the design practice of DDPs (i.e., those using automatic data collection means and opaque algorithmic processes to output personas that are based on behavioral and demographic data about users). For this, we pose the following research questions (RQs):

RQ1: What does fairness mean in the context of data-driven persona development?
RQ2: What guidelines should researchers and practitioners follow to create fair data-driven personas?

Our conceptual analysis is inspired by the principles for *Algorithmic Transparency and Accountability* introduced by ACM in 2017[1]. These include (1) awareness, (2) access, (3) accountability, (4) explanation, (5) data provenance, (6) auditability, and (7) validation and testing [24]. While there are several fairness criteria, e.g., those by Green and Chen [11] (i.e., accuracy, reliability, fairness), the ACM criteria are comprehensive and capture the ethical dimensions appearing in other frameworks. Regarding the practical implications of our analysis, we propose *four ethical data-driven persona guidelines* that cover the ethical concerns throughout the lifecycle of data-driven personas, including data collection, persona creation, and the application of data-driven personas for decision making within organizations. We also pose *nine ethical questions (EQs) for persona creation*.

[1] https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf.

2 Data and Algorithms in Persona Generation

2.1 The Promise of Data

Persona creation has experienced dramatic changes in recent times. Quantitative methods have been leveraged to complement qualitative, interpretative methods in persona creation. Due to rapid development of data science algorithms, quantitative persona creation has been the topic in an increasing number of research articles [25–28]. Quantification of personas is contributing to the broader goal of creating more accurate and more compelling user archetypes from real data.

Moreover, it is seen that quantification can increase the scientific verifiability and credibility of personas, as quantitative methods have the clout of objectivity [29]. Statistical metrics can be used for interpreting how well a specific method creates personas [29, 30], which increases convenience for researchers.

Quantification of personas is also driven by the increasing availability of online user data [18]. When personas were first introduced in the 1990s, the Internet was still an emerging technology, and the tools to collect and process large amounts of user data were scarce. Since then, there has been tremendous progress in automatic collection of data via application programming interfaces (APIs) as well machine learning libraries (e.g., *scikit-learn*[2]) for automating the persona creation pipelines [31, 32], and automatic updating of the personas when the underlying data changes [33].

Simultaneously, data science techniques and algorithms have greatly evolved, including making a variety of statistical and computational approaches accessible for persona creation. For example, natural language processing (NLP) provides multiple methods for persona creation from textual data [34], and numerical data can be used for persona creation with the help of data dimensionality reduction algorithms such as factor analysis, clustering, and matrix factorization [18, 25, 35].

These developments have dramatically increased the feasibility of quantitative persona creation in online settings where personified big data about users or customers can be collected through social media and online analytics platforms (e.g., the APIs of Google, Facebook, and Twitter). Mijač et al. [36] argue that this constitutes a "*shift from using qualitative data towards using quantitative data for persona development*" [36] (p. 1427). In reality, mixed-method personas remain highly popular, with additional enrichment and validation afforded by qualitative analysis [37].

2.2 The Dangers of Data

Nonetheless, thus far, the ethical ramifications and impact of these profound changes in quantitative personas have been overlooked, with the majority of conceptual analyses focused on the role of personas amidst online analytics [8] or the research roadmap for fully automated persona creation [10, 38].

In contrast, only a few studies mention ethical considerations such as data privacy, algorithmic transparency, and risk of creating personas that represent averages or

[2] https://scikit-learn.org/stable/.

majority groups rather than diversity. Data privacy is mentioned by Wöckl [39] arguing that online datasets are typically aggregated, thus preserving the privacy of individual users. Even so, using social media data collected "in the wild" might involve issues of informed consent [40]. Moreover, using social media data presents confidentiality risks for participants, as users can be directly identified through profile characteristics or quotes.

We could locate only two previous studies specifically investigating *ethics and DDPs*. We note that the situation is different in the "mainstream" of persona research – this mainstream research, often qualitatively emphasized, has long recognized ethical issues such as inclusivity, stereotyping, and politics [15, 16, 41]. For example, Turner and Turner [41] provide an interesting conceptual analysis of stereotyping in personas, arguing that stereotyping might be inevitable, as personas always collate information from more detailed to less detailed – thus, naturally collapsing into large groups, common behaviors and tendencies, and the average. The rationale is similar to that of many algorithms of quantitative analysis – statistically speaking, concepts such as "regression to mean", "central limit theorem", "sampling", etc., all deal with representing and/or approaching the central tendency in the data [42]. Statistical methods, generally speaking, rely on means and modes that represent averages, not exceptions. A deviation from this rule can be found in methods of outlier detection [43] that specifically focus on discovering anomalies (i.e., deviations from the mean) in the data. However, outlier detection methods are not typically applied for persona creation, but the methods tend to rely on statistical generalization [37].

Of the studies discussing ethics of using algorithms and data for persona creation, the first one we located focuses only on demographic bias in DDPs [44], with the finding that DDPs inherit demographic imbalances from the source data, thus encouraging persona creators to consider the class balance in their datasets.

The second one analyzed inclusivity via quantitative personas [45], by combining exclusion assessment with DDPs. The findings of this proof-of-concept study show promise for analyzing disadvantaged groups via personas. These studies, although valuable, provide only a limited look into ethics in DDPs, which entails more issues of demographic bias and inclusivity. To this end, the study at hand is geared towards providing a more in-depth analysis of ethics in DDPs.

3 Ethics in Algorithmic Systems

There is a tremendous amount of research about the ethical aspects of computer science. Many aspects are, in fact, not novel but have been discussed over several decades. For example, the morality of computing systems was already discussed by Chorafas in 1966 [46] and Hamming in his 1969 essay in the Journal of the ACM [47]. Nonetheless, in the past few years, ethical themes have re-emerged as a trending topic in computer science and HCI, with an increasing research volume [23, 40, 48–51].

Ananny [48] places the ethics of algorithms within the broader framework of media ethics. He argues that algorithms should be understood beyond computer science's "purely mathematical, mechanistic focus". According to this view, algorithms exist within a complicated assemblage of "computational code, design assumptions, institutional contexts, folk theories [and] user models" [48] to form "Networked Information Algorithm" (NIA). As an NIA or unit of ethical analysis, DDPs describe not only an algorithmic system or merely the human interaction with that system, but "an intersection of technologies and people that makes some associations, similarities, and actions more likely than others" [48]. This embeddedness of algorithms within broader networks of human action forms the basis for understanding the normative and ethical dimensions of algorithmic systems. Gillespie proses that these ramifications may be traced along six ethical dimensions (EDs) [52]:

- *Patterns of inclusion*: the choices behind what makes it into an index in the first place, what is excluded, and how data is made algorithm ready.
- *Cycles of anticipation*: the implications of algorithm providers' attempts to thoroughly know and predict their users, and how the conclusions they draw can matter.
- *The evaluation of relevance*: the criteria by which algorithms determine what is relevant, how those criteria are obscured from us, and how they enact political choices about appropriate and legitimate knowledge.
- *The promise of algorithmic objectivity*: the way the technical character of the algorithm is positioned as an assurance of impartiality, and how that claim is maintained in the face of controversy.
- *Entanglement with practice*: how users reshape their practices to suit the algorithms they depend on, and how they can turn algorithms into terrains for political contest, sometimes even to interrogate the politics of the algorithm itself.
- *The production of calculated publics*: how the algorithmic presentation of publics back to themselves shape a public's sense of itself, and who is best positioned to benefit from that knowledge.

These EDs represent themes that reoccur across the various ethical frameworks in the field. Notions are also widely borrowed from ethical treatises in philosophy. For example, approaches that emphasize *procedural fairness* are based on the idea that algorithmic decision-making should be fair at every step and that algorithms should not be "forgiven" for making unfair decisions during the training process [6]. Adopting the idea of procedural justice, fairness in DDPs can best be assured when considering the normative dimensions of every step in the persona-creation process.

Procedural justice [10] requires that every step of the decision-making process is accurate, fair, consistent, correctable, and ethical [5, 6]. Each step in the creation *and* application of DDPs, therefore, has a normative dimension. An ethical inquiry into the DDPs will consider not only the fairness of the outcome of personas, but every step in the creation of DDPs, the human-algorithmic interface at each step (e.g., the algorithm's hyperparameters set by persona developers), the presentation of personas via the medium of choice, as well as the decision making processes relying on or influenced by DDPs. Each of these "links in chain" is a site for ethical inquiry.

Another concept is *equal opportunity*, referring to the notion that all groups should be treated equally fair. Adopted to the design of DDPs, this could be considered via the concept of *demographic parity*, in that no age, race, or gender is put in a disadvantaged position when the algorithm creates the personas. Since this can be difficult within the constraints of statistical generalization, researchers have created algorithms that specifically encode protected attributes or classes [54]. However, these are yet to be incorporated in DDPs methodologies.

The above approaches focus on *distributive fairness* of ML with the goal of achieving parity of decision outcomes. Attempts have also been made to encode Rawlsian fairness – i.e., the principle of equal opportunity of *individuals* as opposed to *groups* – to combat algorithmic bias, while not mitigating the predictive accuracy [53].

4 Algorithmic Transparency and Accountability in DDPs

4.1 Conceptual Framework

Our conceptual analysis adopts the principles for *Algorithmic Transparency and Accountability* by ACM [24] that include (1) awareness, (2) access, (3) accountability, (4) explanation, (5) data provenance, (6) auditability, and (7) validation and testing. This framework is chosen for two reasons: first, it is comprehensive, covering the main aspects of ethics in algorithmic systems. Second, it is released by ACM and thus provides legitimacy when evaluation systems, especially those that are applied in nature. Personas, as a technique, are highly applied for which warranted to analyze them using a generalizable framework such as the ACM guidelines.

Overarchingly, DDPs may be thought of as an attempt to disaggregate the aggregated data. The most salient normative issues, therefore, concern the ethics of categorization, bias, and discrimination, and the question of algorithmic transparency. Thus, we expand the ACM framework to include these dimensions. The outcome is a conceptual analysis framework with ten “pillars” (see Fig. 1).

	Awareness	Access	Accountability	Explanation	Data provenance	Auditability	Validation and testing
Persona developers	??	??	??	??	??	??	??
Algorithms	??	??	??	??	??	??	??
Inspiration users	??	??	??	??	??	??	??
Decision makers	??	??	??	??	??	??	??
Decision targets	??	??	??	??	??	??	??

Fig. 1. Conceptual framework for analyzing ethical issues of DDPs

It is important to acknowledge that various roles are involved in ethical considerations revolving around DDPs (Fig. 1). These include at least:

- *Persona developers* – these are the creators of personas. The creators have major agency in the chain of ethical DDPs, as they make critical decisions about data collection, choice of algorithms, setting of hyperparameters (e.g., the number of generated personas), and so on.

- *Algorithms* – these are models, algorithms and computational techniques used when creating the personas and presenting them to decision makers.
- *Inspiration users* – these are the users whose data is used for persona creation (e.g., interviewees, respondents, social media users, website visitors, customers in the company's database…).
- *Decision makers* – these are the end-users that "use" personas, meaning they make decisions based on personas and/or refer to personas in their thinking about the users and in communication with other stakeholders in user-centric actions.
- *Decision targets* – these are the users that are facing the consequences of decisions made based on persona information.

Therefore, ethical issues are not isolated in only the confined chambers of researchers creating them, but rather cover the entire persona lifecycle [55], from their creation of exploitation. For example, decision makers can make "bad" or ethically unquestionable decisions (e.g., discriminate, enforce existing stereotypes) based on personas. If the decisions were based on inaccurate persona information, then we should emphasize the responsibility of persona developers. However, if the information was accurate and the decisions were still wrong, the emphasis of responsibility should be on the decision makers. This goes to show that persona creation and application are intertwined in a complex, and often intractable relationship, which requires any analysis on this topic to consider multiple stakeholder perspectives.

Moreover, algorithms, although not people or legal subjects, have agency (i.e., power) over the decisions. Algorithms are amoral (i.e., they do not recognize moral guidelines unless imposed as formal rules), but they can still behave immorally from a human perspective [56]. This can take place via statistical selection processes (i.e., "algorithmic decision making" [57]), but also in the choice of the medium for presenting the personas (e.g., certain persona information can be presented more saliently than other information in a persona UI). Even a seemingly simple issue such as selection of persona pictures can form a major ethical choice, as the picture can evoke gender [15] and racial [58] stereotypes. These issues cannot be neglected, as argued by Salminen et al. [58], as personas forcefully have a specific gender and race.

4.2 Analysis

In this section, we apply the ACM framework to discuss the ethical aspects of DDPs. In each subsection, we present important EQs for persona developers to address.

Awareness. Definition [24]: "*Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.*"

A literature review on quantitative persona generation [37] reveals there is, in general, little consideration to ethical matters in quantitative persona articles. Instead, the papers tend to focus on technical justification and evaluation of personas as information. The focus differs from the conceptually oriented persona research, with repeated studies on ethical matters, especially focused on stereotypes [16, 41], and inclusivity [15, 45]. Thus, the awareness aspect shows there is "work to do" to activate quantitatively oriented persona researchers to consider ethical aspects of DDPs.

To this end, the EQs include:

EQ1: How can creators of DDPs be made more aware of ethical concerns relating to DDPs?
EQ2: How can persona users be made aware of potential bias in DDPs?

Siegel [29] refers to the "mystique of numbers" that in their case manifested in company stakeholders not questioning the user segments because these were based on data and algorithms. Thus, there may be a fallacy of objectivity, in that decision makers in some cases are not questioning DDPs because they are seen as objective representations of real user data. Coincidentally, this line of thinking would actually be preferred by creators of DDPs in certain sense, as DDP methodologies are partly created to address the lack of credibility of personas, which has been found a real concern in empirical user studies [59, 60] as well as conceptual treatise of personas [30]. The empirical user studies of DDPs, in turn, provide evidence that decision makers remain critical to DDPs, questioning the data and the details of how the personas were created [58, 61]. Data resistance can be an issue, especially when personas are not believed because they contain information that contradicts the user's existing biases. Based on these slightly conflicting empirical findings, persona users can perhaps be divided into two main groups based on their trusting attitudes: those that accept the persona information as facts without questioning [29], and those that are skeptical and require further explanations to "believe" the persona is real [61].

Access and Redress. Definition [24]: "*Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions*."

This aspect concerns especially the individuals facing the consequences of decisions made based on DDPs. For example, organizations such as crime fighting agencies or insurance companies create "thug personas", "criminal personas", "diabetes personas", or "risk personas" that either over-generalize and thus provide basis for discriminatory decision making, or are accurate (e.g., capturing a person's higher risk of getting diabetes) and, therefore, make it possible to provide unfair terms for individuals at large. We can thus infer that providing free access to automatic persona generation systems can result in ethical vulnerabilities – the efficient use of personas can be unethical in nature.

The question hinges on an ethical understanding of algorithmic categories and probable similarity. As Gillespie [52] notes, categorization is a powerful political and semantic tool, particularly in the context of ML. Categories create order out of disparate information and present information in a fixed way that discourages alternatives. Minority groups can be especially vulnerable to being misinterpreted or having a lack of representation in DDPs, as data science algorithms tend to focus on averages and patterns and tendencies that reflect behaviors and traits of the majority subsets in the dataset. Thus, special consideration is needed to capture the diversity of user communities. While this can potentially be done within one generation of personas (depending on the correction methods available for a given algorithm), another option is to run algorithms several times: e.g., one set of personas from the full dataset and another set of "minority personas", identified by exploratory data analysis.

The matter of access is not only a question of obligation but can provide tangible benefits for decision makers. This is because useful insights for design of usability and user experience can often be found in outliers and minority segments (e.g., accessibility). Thus, DDPs can quantify issues of fairness and accessibility. For this, one of the central questions is:

EQ3: How can it be ensured that DDPs does not highlight marginalized, vulnerable, or otherwise disadvantaged populations in a harmful way?

Accountability. Definition [24] "*Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results*."

Algorithms and ML are increasingly understood as agential; as operating in terms that are becoming progressively unknowable and indecipherable to humans [62]. Holding algorithms accountable for potentially unethical DDPs is hampered by the opaqueness and lack of transparency of algorithms in general [50]. Thus, accountability of algorithms, therefore, relates closely to algorithmic transparency: the more is known about algorithmic decision-making, the better it can be evaluated for fairness. In most (but not all) cases, transparency of algorithmic decision-making leads to increased fairness [63, 64].

Clearly, the responsibility for the ethicality is shared by humans (persona creators and users) and algorithms. To be accountable, the humans involved in DDP projects need to understand the potential consequences of personas in the real world. This aspect of *actionability* is also noted by Gillespie. "What we need," notes Gillespie, "is an interrogation of algorithms as a key feature of our information ecosystem, and of the cultural forms emerging in their shadows with close attention to where and in what ways the introduction of algorithms into human knowledge practices may have political ramifications" [7].

Furthermore, it has been shown that automated systems may diminish people's sense of accountability and moral agency [23]. In other words, responsibility is shifted to the algorithm. This is a dangerous road, given the various types of biases involved with personas. Research has shown that personas are still generated mostly from survey data rather than behavioral data sources [37]. However, even when analyzed quantitatively, survey data may include several issues of validity (e.g., social desirability bias [65]). In a similar vein, setting the number of personas, applying hyperparameters for algorithms and other steps that involve manual tuning are subject to human bias. Therefore, "quantitative" does not automatically mean "objective" or "truthful", which is a critical consideration for accountability.

On the other hand, when personas are true to the data, this decreases – in theory [1, 66] – the decision makers' tendency to rely on user stereotypes that are compatible with their own biases. The implication is that adopting the principles of good ML (e.g., proper treatment of class imbalance [9] is ethical ML.

Some of the striking questions involve:

EQ4: What is the chain of responsibility in persona lifecycle, ranging from creation to application?
EQ5: Who is responsible for unethical choices based on personas; their creators or stakeholders applying them, and when?

Explanation. Definition [24]: "*Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts*."

Research suggests that people struggle to interpret and evaluate ML outcomes [67]. This also applies to personas that have been perceived as abstract [59], unrealistic [30], and confusing [68, 69]. Not only do people use algorithmic outcomes in unexpected and biased ways, but they are influenced by irrelevant information and display poor judgment in gauging the accuracy of algorithms. While it is not clear whether this phenomenon is fundamental to HCI or the result of factors like interface design or training [67], explainability inarguably poses a critical design challenge for DDPs.

Transparency has been suggested as a solution to trust concerns regarding data use and algorithmic decision-making [13, 70]. It is argued that by understanding how systems and algorithms work, decision makers using those systems or algorithms will feel more comfortable and trusting with the results [71]. On the other hand, previous findings on how to improve human-algorithmic interactions show that providing explanations or feedback does not necessarily improve human performance [67].

Particularly, users of DDPs may question the information in persona profiles because they are unsure of how it was created [61]. This problem is especially vexing for data-driven personas because their creation is an opaque algorithmic process. The more information and data the persona profile contains, the more complex its cognitive processing may become [61]. Thus, there is a trade-off of increasing informativeness ("roundedness" [7]) of personas and their understandability.

The challenge of explanations is further enhanced by the fact that the creation mechanisms of DDPs are complicated to understand by laymen and, at times, even other researchers. Moreover, if decision makers only see the DDPs without any explanations, they may still consider data-driven personas as untrustworthy because they may be unsure how the information in the persona profiles was inferred [30, 58].

Salminen et al. [72] investigated technically oriented explanations of information in DDP profiles and found that higher transparency through explanations increased the perceived completeness and clarity of the personas among end users. They encourage creators of DDPs to consider "persona transparency" by including clear statements of where the data originates, how it was collected, and what were the analysis steps that resulted in the personas shown to the decision makers.

Data Provenance. Definition [24]: "*A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of* *analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals*."

Using social media data may present confidentiality risks for participants, as participants can be directly identified through their profile characteristics or comments [40]. Therefore, privacy of individuals can be violated and/or their views misrepresented when automatically selecting social media quotes for persona profiles. In contrast, the aggregated and non-personally identifiable information regarding quantitative performance metrics such as click and view counts can be useful for safeguarding the privacy of individual users [3]. In this sense, the structured data afforded by many Web analytics and social media platforms can support the ethical creation of personas.

In contrast, the extant trend [54, 73] to remove sensitive classes such as race and gender from the data can be problematic for ethical persona creation. This is because it reduces the ability of persona creators to, firstly, specifically portray marginalized group – when data is not available, these personas cannot easily be created and therefore understanding these user groups using DDPs becomes hard, if not possible. Secondly, the lack of protected class attributes can make it harder to fix the biases in quantitative algorithms – for example, an attribute such as race can be proxied by other variables in the dataset (e.g., income, location). This can especially take place with algorithms that learn latent patterns not directly visible in the dataset [21]. As a result, the decisions may involve a sort of a representation of the latent variable even when it is removed. For these purposes, masking data is challenging for ethical DDPs.

Auditability. Definition [24]: "*Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected*."

Although algorithms and ML influence human decision-making, how humans and algorithms interact to form decisions is not well understood [67]. This may hinder the scrutiny of DDPs, as it might not be clear for researchers or practitioners how to audit personas. One promising alternative is to provide so-called "full stack personas" (*forthcoming*), using a persona system through which decision makers can download the raw data of their personas (called "interaction matrix").

Another challenge is the trade-off of private vs. publicly available data. Naturally, for replication and scrutiny, data used for DDP creation would need to be available for other researchers. However, making the data available can, on one hand, violate the terms of service (TOS) in online analytics platforms – for example, Twitter disallows direct sharing of tweets (they can be shared using Tweet IDs). On the other hand, if persona creation data is made available, this can violate the privacy of individuals based on whose information the personas are created – thus, researchers should consider getting the users' consent while adhering to TOS' of online platforms. Because this adds the complexity and required effort, most DDP studies fail to share their data [37].

Auditability can also involve aspects of users' choice – relevant questions here include, for example:

EQ6: Can users see their corresponding personas?
EQ7: Can user correct misinformation/mismatches of personas?
EQ8: Can online users "opt out" of their data being used for persona creation?

Validation and Testing. Definition [24]: "*Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public*." Validation of persona ethics suffers from the lack of standards and metrics. What is the metric for an ethical persona? The issue can be demonstrated via the example of representativeness, which is understood very differently whether one comes from a statistical background or from an ethics background.

In many of studies developing DDPs, representativeness (or inclusivity) tends to be considered from the perspective of statistics [37], not from the perspective of fairness. The difference is such a representative persona set describes the main tendencies of the data via personas, whereas an inclusive persona set would include personas evenly for each defined class. The objective of the former is efficient representation of central data, while the objective of the latter is the maximization of diversity [45].

As these two approaches appear incommensurable, the outcome is real challenge for validation of DDPs – or, as Hill et al. [15] put it, "can we have it both ways?".

Testing and validating DDPs can, nonetheless, provide answers to ethical questions. For example, consider the trade-off regarding complexity vs. comprehension. In other words, when designing explanations for explainable DDPs, the outputs can easily become too complex [72], which defeats the purpose. This trade-off prompts persona developers to carry out empirical testing and validation towards the goal of finding the optimal "simplicity-informativeness" ratio.

EQ9: How can more detailed reporting of ethical aspects be promoted within DDP creation?

5 Discussion

5.1 The Good, the Bad, and the Ugly of Personas

The findings of the conceptual analysis indicate that personas cannot be created in "blind faith" with the assumption that the underlying data and applied algorithms would automatically yield "objective" outcomes; rather, the risks and biases need to be properly scrutinized for each DDP project. This is important to avoid biased decisions based on the personas by stakeholders that are using them. In other words, issues relating to the nature of data (i.e., measurement errors, imbalance, protected classes), as well as the statistical nature of algorithms (overgeneralization) need to be considered when applying automatic quantitative methods for persona creation.

Somewhat ironically, DDPs were originally introduced to address the issue of human bias and limited data when using qualitative persona creation [10, 19, 27, 36]. However, new sources of bias emerged, forcing the creators of DDPs to exit the curtain of (alleged) objectivity. These new sources involve both human and machine bias. The former is exemplified by selection of data and algorithms, as well as setting the hyperparameter values such as the number of personas. The latter is exemplified by tendency towards means, modes, and averages. The former should be addressed by explicating and justifying the manual choices in DDP creation process. The latter can be addressed using statistical methods such as dividing the data into even subsets and/or applying outlier detection. In ML studies, there are several approaches to class balancing [74, 75] that can help process the data for fairer personas.

Overall, the ethical concerns in DDPs stem from (a) the increase in the use of online user data (e.g., social media profiles), as well as from (b) the use of opaque algorithmic processes to generate the personas. The automated processing of user data to create artificial user profiles (i.e., personas) thus transcends ethical questions about the data itself (privacy, ownership) as well as the algorithms involved in manipulating the data (transparency, fairness). All these factors must be considered and be subject to further research and development of ethical DDP methodologies.

We would like to point out that not all matters in this space are gloomy "risks" or "threats". There are positive opportunities as well. One of these is correcting biases and stereotypes decision makers have about users – because DDPs are based on quantitative evidence, they might be more believable for (at least skeptical) stakeholders. Thus, personas could be actually used to correct biases and stereotypes.

Another aspect is using DDPs as tools to pinpoint underprivileged groups. For example, if certain demographics are missing from the data, then communicating the absence of these groups via personas could be a compelling method to show "who is missing". It is then not necessarily the personas that are biased but the social structures that yielded the data, and personas simply reflected these structures.

In many respects, DDPs may be seen as offering a solution to the inscrutability of big data analytics. By presenting big data in a persona format, DDPs aim to humanize algorithmic machine learning and to package this information into a representation that is understandable to human reasoning. However, because of the unpredictable nature of human decision-making, the normative dimension of the interface between DDPs and end-users should be of concern to persona designers.

5.2 Ethical Data-Driven Persona Guidelines

Persona creators should be aware that harm from online research can occur for classes of people and communities [49]. The ethically questionable practices to avoid for persona creation (and conversely to strive for) are proposed in Table 1.

Table 1. What to avoid and what to strive for when creating ethical data-driven personas

Bad way	Good way
Generating personas based on averages and majorities while overlooking deviant or minority personas (a concern for inclusivity [45])	Creating personas also from outliers and deviating behaviors. Using subsets of data that describe marginalized groups, or specifically acquiring such data
Reinforcing stereotypes as a consequence of the previous point (a concern for application) [41]	Increasing the number of personas created to cover more subsegments in data [33], and using data to demonstrate diversity within a persona, such as showing multiple pictures for gender diversity [15]
Creating personas that appear "objective" and "perfect" because they are created using numerical data and algorithms (the "mystique of numbers" [29]), without communicating the limitations that each method inevitably has to the end users of personas (a concern of transparency)	Being frank about the applied methods and their limitations, adding explanations and other forms of transparency in persona systems [72]
Not corroborating the DDPs using triangulation or qualitative insights [2], while relying on the black-box data from online platforms whose sources of error and bias remain unknown (a concern of truthfulness)	Creating "hybrid personas" that are based on quantitative and qualitative insights [5] as a form of triangulation, and using both text and numbers to describe the personas [76]

In the *worst ethical scenario*, stakeholders are presented with personas that represent only majority user segments, without explaining how the personas were created (persona transparency [72]) and what are the drawbacks of the methods applied, and without ensuring that the data upon which the personas are built is actually valid. In the *best ethical scenario*, the ethically ideal personas (a) capture the diversity of the user segments, are (b) transparent in the sense that their generation and information is well explained and understood by decision makers and replicable if needed, and are (c) corroborated by using methods of triangulation.

Persona developers also have a responsibility in verifying that the users really understand the limitations of each method. This is more complex than it seems, as users can easily argue they understand (e.g., non-verbally nodding), whereas in reality, they do not understand. Asking the users to explain the DDPs in their own words is one tactic for ensuring proper understanding.

Decision makers should not blindly believe the outputs of DDP creation algorithms. Additional steps, such as ensuring data quality and triangulating the results with other methods, such as traditional qualitative interviews, are necessary. This is not a novel recommendation, as persona scholars have consistently advocated mixed-method personas [1, 2, 5] – however, in the "hype of data science", this old wisdom can easily be forgotten. In practice, practitioners with limited knowledge about quantitative methods should "ask stupid questions" to avoid the "mystique of numbers" [29], including asking clarification about how the personas were created, what manual choices the creation process involved, and how the results were validated.

6 Conclusion

Our goal was to tie DDPs into the algorithmic fairness, accountability, and transparency discussion. Through this linkage, we provide guidelines for data-driven persona creation that include (a) creating personas also from outliers (not only majority groups), (b) using data to demonstrate diversity within a persona, (c) explaining the methods and their limitations as a form of transparency, and (d) triangulating the persona information to increase truthfulness. These recommendations provide a starting point for developing standards for ethical data-driven persona creation.

References

1. Cooper, A.: The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams - Pearson Education, Indianapolis (1999)
2. Pruitt, J., Grudin, J.: Personas: practice and theory. In: Proceedings of the 2003 Conference on Designing for User Experiences, San Francisco, California, USA, pp. 1–15. ACM (2003). https://doi.org/10.1145/997078.997089
3. An, J., Kwak, H., Salminen, J., Jung, S., Jansen, B.J.: Imaginary people representing real numbers: generating personas from online social media data. ACM Trans. Web (TWEB) **12** (4), 1–26 (2018)
4. Salminen, J., et al.: Generating cultural personas from social data: a perspective of middle eastern users. In: Proceedings of the Fourth International Symposium on Social Networks Analysis, Management and Security (SNAMS-2017), Prague, Czech Republic. IEEE (2017). https://doi.org/10.1109/FiCloudW.2017.97
5. Salminen, J., et al.: From 2,772 segments to five personas: summarizing a diverse online audience by generating culturally adapted personas. First Monday **23**, 8415 (2018). https://doi.org/10.5210/fm.v23i6.8415
6. Nielsen, L., Hansen, K.S., Stage, J., Billestrup, J.: A template for design personas: analysis of 47 persona descriptions from Danish industries and organizations. Int. J. Sociotechnol. Knowl. Dev. **7**, 45–61 (2015). https://doi.org/10.4018/ijskd.2015010104
7. Nielsen, L.: Personas - User Focused Design. Springer, New York (2019). https://doi.org/10.1007/978-1-4471-7427-1
8. Salminen, J., Jansen, B.J., An, J., Kwak, H., Jung, S.: Are personas done? Evaluating their usefulness in the age of digital analytics. Pers. Stud. **4**, 47–65 (2018). https://doi.org/10.21153/psj2018vol4no2art737
9. Zhu, H., Wang, H., Carroll, J.M.: 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 2019, San Diego, CA, USA, pp. 61–70. ACM (2019). https://doi.org/10.1145/3322276.3322285
10. Salminen, J., Jung, S., Jansen, B.J.: 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, Heraklion, Greece, pp. 596–603. SciTePress (2019)

11. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. In: Proceedings of the ACM on Human-Computer Interaction (CSCW), pp. 1–24 (2019). https://doi.org/10.1145/3359152
12. Chander, A.: The racist algorithm. Mich. L. Rev. **115**, 1023 (2016)
13. Diakopoulos, N., Koliska, M.: Algorithmic transparency in the news media. Digit. J. **5**, 809–828 (2017)
14. Eslami, M., Vaccaro, K., Karahalios, K., Hamilton, K.: "Be careful; things can be worse than they appear": understanding biased algorithms and users' behavior around them in rating platforms. In: Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM), Montréal, Canada, pp. 62–71 (2017)
15. Hill, C.G., et al.: Gender-inclusiveness personas vs. stereotyping: can we have it both ways? In: Proceedings of the 2017 CHI Conference, pp. 6658–6671. ACM (2017). https://doi.org/10.1145/3025453.3025609
16. Marsden, N., Haag, M.: Stereotypes and politics: reflections on personas. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4017–4031. ACM, New York (2016). https://doi.org/10.1145/2858036.2858151
17. Pröbster, M., Haque, M.E., Marsden, N.: Perceptions of personas: the role of instructions. In: 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1–8 (2018). https://doi.org/10.1109/ICE.2018.8436339
18. An, J., Kwak, H., Jung, S., Salminen, J., Jansen, B.J.: Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Soc. Netw. Anal. Min. **8**, 54 (2018). https://doi.org/10.1007/s13278-018-0531-0
19. Zhang, X., Brown, H.-F., Shankar, A.: Data-driven personas: constructing archetypal users with clickstreams and user telemetry. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5350–5359. ACM, New York (2016)
20. Chapman, C.N., Love, E., Milham, R.P., ElRif, P., Alford, J.L.: Quantitative evaluation of personas as information. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1107–1111 (2008). https://doi.org/10.1177/154193120805201602
21. Hajian, S., Bonchi, F., Castillo, C.: 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, pp. 2125–2126. ACM (2016)
22. Stoyanovich, J., Abiteboul, S., Miklau, G.: Data, responsibly: fairness, neutrality and transparency in data analysis. In: International Conference on Extending Database Technology, Bordeaux, France (2016)
23. Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., Floridi, L.: The ethics of algorithms: mapping the debate. Big Data Soc. **3** (2016). https://doi.org/10.1177/2053951716679679
24. Garfinkel, S., Matthews, J., Shapiro, S.S., Smith, J.M.: Toward algorithmic transparency and accountability. Commun. ACM **60**, 5 (2017). https://doi.org/10.1145/3125780
25. Brickey, J., Walczak, S., Burgess, T.: A comparative analysis of persona clustering methods. In: AMCIS 2010 Proceedings (2010)
26. Laporte, L., Slegers, K., De Grooff, D.: Using correspondence analysis to monitor the persona segmentation process. In: Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, pp. 265–274. ACM, New York (2012). https://doi.org/10.1145/2399016.2399058
27. McGinn, J.J., Kotamraju, N.: Data-driven persona development. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, pp. 1521–1524. ACM (2008). https://doi.org/10.1145/1357054.1357292

28. Miaskiewicz, T., Sumner, T., Kozar, K.A.: A latent semantic analysis methodology for the identification and creation of personas. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1501–1510. ACM (2008)
29. Siegel, D.A.: The mystique of numbers: belief in quantitative approaches to segmentation and persona development. In: CHI 2010 Extended Abstracts on Human Factors in Computing Systems, pp. 4721–4732. ACM, New York (2010). https://doi.org/10.1145/1753846.1754221
30. Chapman, C.N., Milham, R.P.: The personas' new clothes: methodological and practical arguments against a popular method. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 634–636 (2006). https://doi.org/10.1177/154193120605000503
31. Jung, S., Salminen, J., An, J., Kwak, H., Jansen, B.J.: Automatically conceptualizing social media analytics data via personas. In: Presented at the 12th International AAAI Conference on Web and Social Media (ICWSM 2018), San Francisco, California, USA, 25 June 2018 (2018)
32. Jung, S., Salminen, J., Kwak, H., An, J., Jansen, B.J.: Automatic Persona Generation (APG): a rationale and demonstration. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA, pp. 321–324. ACM (2018). https://doi.org/10.1145/3176349.3176893
33. Jung, S., Salminen, J., Jansen, B.J.: 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, Glasgow, UK, pp. LBW2714:1–LBW2714:6. ACM (2019). https://doi.org/10.1145/3290607.3312955
34. Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., Dolan, B.: A persona-based neural conversation model. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 994–1003. Association for Computational Linguistics, Berlin (2016). https://doi.org/10.18653/v1/P16-1094
35. Kim, H.M., Wiggins, J.: A factor analysis approach to persona development using survey data. In: Proceedings of the 2016 Library Assessment Conference, p. 11 (2016)
36. Mijač, T., Jadrić, M., Ćukušić, M.: 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 (2018). https://doi.org/10.23919/MIPRO.2018.8400224
37. Salminen, J., Guan, K., Jung, S.-G., Chowdhury, S.A., Jansen, B.J.: A literature review of quantitative persona creation. In: Proceedings of the ACM Conference of Human Factors in Computing Systems (CHI 2020), Honolulu, Hawaii, USA. ACM (2020)
38. Salminen, J., Jansen, B.J., An, J., Kwak, H., Jung, S.: Automatic persona generation for online content creators: conceptual rationale and a research agenda. In: Nielsen, L. (ed.) Personas - User Focused Design. HIS, pp. 135–160. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7427-1_8
39. Wöckl, B., Yildizoglu, U., Buber, I., Aparicio Diaz, B., Kruijff, E., Tscheligi, M.: Basic senior personas: a representative design tool covering the spectrum of european older adults. In: Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 25–32. ACM, New York (2012). https://doi.org/10.1145/2384916.2384922
40. Fiesler, C., Proferes, N.: "Participant" perceptions of twitter research ethics. Soc. Media +Soc. **4**, 1–14 (2018). https://doi.org/10.1177/2056305118763366
41. Turner, P., Turner, S.: Is stereotyping inevitable when designing with personas? Des. Stud. **32**, 30–44 (2011)
42. Spiegel, M.R., Constable, R.L.: Theory and Problems of Statistics. Schaum, New York (1961)

43. Ben-Gal, I.: Outlier detection. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 131–146. Springer, Boston (2005). https://doi.org/10.1007/0-387-25465-X_7
44. Salminen, J., Jung, S., Jansen, B.J.: Detecting demographic bias in automatically generated personas. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. LBW0122:1–LBW0122:6. ACM, New York (2019). https://doi.org/10.1145/3290607.3313034
45. Goodman-Deane, J., Waller, S., Demin, D., González-de-Heredia, A., Bradley, M., Clarkson, J.P.: Evaluating inclusivity using quantitative Personas. In: Presented at the Design Research Society Conference 2018, 28 June (2018). https://doi.org/10.21606/drs.2018.400
46. Chorafas, D.N.: Control Systems Functions and Programming Approaches by Dimitris N Chorafas. Academic Press, New York (1966)
47. Hamming, R.W.: One man's view of computer science. J. ACM (JACM) **16**, 3–12 (1969)
48. Ananny, M.: Toward an ethics of algorithms: convening, observation, probability, and timeliness. Sci. Technol. Hum. Values **41**, 93–117 (2016)
49. Hoffmann, A.L., Jonas, A.: Recasting justice for internet and online industry research ethics. In: Zimmer, M., Kinder-Kuranda, K. (eds.) Internet Research Ethics for the Social Age: New Cases and Challenges. Peter Lang, Bern (2016)
50. Neyland, D.: Bearing accountable witness to the ethical algorithmic system. Sci. Technol. Hum. Values **41**, 50–76 (2016)
51. Ullmann, S., Tomalin, M.: Quarantining online hate speech: technical and ethical perspectives. Ethics Inf. Technol. (2019). https://doi.org/10.1007/s10676-019-09516-z
52. Gillespie, T.: The relevance of algorithms. In: Gillespie, T., Boczkowski, P., Foot, K. (eds.) Media Technologies: Essays on Communication, Materiality, and Society, pp. 167–194. MIT Press, Cambridge (2014)
53. Joseph, M., Kearns, M., Morgenstern, J., Neel, S., Roth, A.: Rawlsian fairness for machine learning. arXiv preprint arXiv:1610.09559 (2016)
54. Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., Baeza-Yates, R.: FA*IR: a fair top-k ranking algorithm. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 1569–1578. ACM (2017)
55. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Morgan Kaufmann, Boston (2006)
56. Arsiwalla, X.D., Freire, I.T., Vouloutsi, V., Verschure, P.: Latent morality in algorithms and machines. In: Martinez-Hernandez, U., et al. (eds.) Living Machines 2019. LNCS (LNAI), vol. 11556, pp. 309–315. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24741-6_27
57. Rouse, W.B., Sheridan, T.B.: Computer-aided group decision making: theory and practice. Technol. Forecast. Soc. Change **7**, 113–126 (1975)
58. Salminen, J., Nielsen, L., Jung, S., An, J., Kwak, H., Jansen, B.J.: "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), Montréal, Canada (2018)
59. Matthews, T., Judge, T., Whittaker, S.: How do designers and user experience professionals actually perceive and use personas? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA, pp. 1219–1228. ACM (2012). https://doi.org/10.1145/2207676.2208573
60. Rönkkö, K., Hellman, M., Kilander, B., Dittrich, Y.: 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, vol. 1, Toronto, Ontario, Canada, pp. 112–120. ACM (2004). https://doi.org/10.1145/1011870.1011884

61. Salminen, J., Jung, S., An, J., Kwak, H., Nielsen, L., Jansen, B.J.: Confusion and information triggered by photos in persona profiles. Int. J. Hum. Comput. Stud. **129**, 1–14 (2019). https://doi.org/10.1016/j.ijhcs.2019.03.005
62. Ziewitz, M.: Governing algorithms: myth, mess, and methods. Sci. Technol. Hum. Values **41**, 3–16 (2016)
63. Lee, M.K., Jain, A., Cha, H., Ojha, S.: Procedural justice in algorithmic fairness: leveraging transparency and outcome control for fair algorithmic mediation. Psychology **3**, 14 (2019)
64. Zarsky, T.: The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. Sci. Technol. Hum. Values **41**, 118–132 (2016)
65. Fisher, R.J.: Social desirability bias and the validity of indirect questioning. J. Consum. Res. **20**, 303–315 (1993)
66. Nielsen, L., Storgaard Hansen, K.: Personas is applicable: a study on the use of personas in Denmark. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Ontario, Canada, pp. 1665–1674. ACM (2014)
67. Green, B., Chen, Y.: The principles and limits of algorithm-in-the-loop decision making. Proc. ACM Hum. Comput. Interact. **3**, 50–74 (2019)
68. Salminen, J., Jung, S., An, J., Kwak, H., Jansen, B.J.: Findings of a user study of automatically generated personas. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montréal, Canada, pp. LBW097:1–LBW097:6. ACM (2018). https://doi.org/10.1145/3170427.3188470
69. Salminen, J., Sengun, S., Jung, S., Jansen, B.J.: Design issues in automatically generated persona profiles: a qualitative analysis from 38 think-aloud transcripts. In: Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), Glasgow, UK, pp. 225–229. ACM (2019). https://doi.org/10.1145/3295750.3298942
70. Ananny, M., Crawford, K.: Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc. **20**, 973–989 (2018)
71. Kizilcec, R.F.: How much information?: effects of transparency on trust in an algorithmic interface. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, USA, pp. 2390–2395. ACM (2016)
72. Salminen, J., Santos, J.M., Jung, S., Eslami, M., Jansen, B.J.: Persona transparency: analyzing the impact of explanations on perceptions of data-driven personas. Int. J. Hum. Comput. Interact. **36**, 1–13 (2019). https://doi.org/10.1080/10447318.2019.1688946
73. Ajunwa, I., Friedler, S., Scheidegger, C.E., Venkatasubramanian, S.: Hiring by algorithm: predicting and preventing disparate impact. Available at SSRN (2016)
74. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
75. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: Presented at the 2018 IEEE International Joint Conference on Neural Networks, June 2008. https://doi.org/10.1109/IJCNN.2008.4633969
76. Salminen, J., Liu, Y.-H., Sengun, S., Santos, J.M., Jung, S., Jansen, B.J.: The Effect of numerical and textual information on visual engagement and perceptions of AI-Driven persona interfaces. In: Proceedings of the ACM Intelligent User Interfaces (IUI 2020), Cagliary, Italy. ACM (2020)