Don’t guess my gender, gurl: The inadvertent impact of gender inferences

Eduard Fosch-Villaronga[1][0000-0002-8325-5871], Adam Poulsen [2][0000-0002-0001-3894], Roger A. Søraa[3][0000-0001-6800-0558], and Bart H.M. Custers[1][0000-0002-3355-8380]

[1] eLaw Center for Law and Digital Technologies, Leiden University, Leiden, The Netherlands
{e.fosch.villaronga,b.h.m.custers}@law.leidenuniv.nl
[2] School of Computing and Mathematics, Charles Sturt University, Australia,
apoulsen@csu.edu.au
[3] Department of Interdisciplinary Studies of Culture, Norwegian University of Science and Technology (NTNU), Norway, roger.soraa@ntnu.no

Abstract. Social media platforms employ inferential analytics methods to guess user preferences and sensitive attributes such as race, gender, sexual orientation, and opinions. These methods are often opaque, can predict behaviors for marketing purposes, influence behavior for profit, serve attention economics, and reinforce existing biases such as gender stereotyping. Although two international human rights treaties include express obligations relating to harmful and wrongful stereotyping, these stereotypes persist online and offline, as if platforms failed to understand that gender is not merely being a 'man' or a 'woman,' but a social construct. Our study investigates the impact of algorithmic bias on inadvertent privacy violations and the reinforcement of social prejudices of gender and sexuality through a multidisciplinary perspective, including legal, computer science, and queer media viewpoint. We conducted an online survey to understand whether Twitter inferred the gender of users and whether that was correct. Beyond Twitter's binary understanding of gender and the inevitability of the gender inference as part of Twitter's personalization trade-off, the results show that, in nearly 20% of the cases (N=109), it misgendered users. Although not apparently correlated, only 8% of the straight male respondents were misgendered, compared to 25% of gay males and 16% of straight females. Our contribution shows how the lack of attention to gender in gender classifiers exacerbates existing biases and affects marginalized communities. With our paper, we hope to promote the online account for privacy, diversity, inclusion, and advocate for the freedom of identity that everyone should have online and offline.

Keywords: Gender; Twitter; Inference; Gender Classifier; Privacy; Algorithmic Bias; Discrimination; LGBTQAI+; Gender stereotyping.

1 Introduction

Online and social media platform providers use users' traits, including name, age, and gender, to improve user experience and personalize online behavioral advertising. By knowing users' characteristics, corporations can target or exclude certain groups more efficiently, tailor their services to users, and increase the time they spend on the platform [1]. In such a way, profiling makes marketing more precise and effective. However, a growing concern is the increasing use of opaque inferential analytics that reveal sensitive user attributes that serve attention economics [2] and that may reinforce existing biases that, although not explicit, they can be very influential [3,4].

A recurrent bias is gender stereotyping. Gender stereotyping 'refers to the practice of ascribing to an individual woman or man specific attributes, characteristics, or roles by reason only of their membership in the social group of women or men' [5]. Two international human rights treaties include express obligations relating to harmful and wrongful stereotyping. Art. 5 of the Convention on the Elimination of All Forms of Discrimination against Women mandates States Parties to "take all appropriate measures to modify the social and cultural patterns of conduct of men and women, to achieve the elimination of prejudices and customary and all other practices which are based on the idea of the inferiority or the superiority of either of the sexes or stereotyped roles for men and women." Art. 8(1)(b) of the Convention on the Rights of Persons with Disabilities stresses that 'States Parties undertake to adopt immediate, effective and appropriate measures to combat stereotypes, prejudices and harmful practices relating to persons with disabilities, including those based on sex and age, in all areas of life.' However, these stereotypes persevere both online and offline [6,7], as if platforms failed to understand that gender is not merely being a 'man' or a 'woman,' but a social construct [8].

Given that gender stereotypes persist online, and that the social media platform Twitter infers gender from a wide variety of sources,[1] we address the research question (RQ): *How accurate are Twitter's inferences of its users' gender identities and what implications does this social media practice have?* Addressing this RQ brings into view concerns of discrimination, misgendering, and exacerbation of existing biases that online platforms persist in replicating and that the literature starts highlighting [9-10]. Our goal is to investigate misgendering on Twitter and illustrate the impact of algorithmic bias on inadvertent privacy violations and reinforce social prejudices of gender and sexuality through a multidisciplinary perspective, including legal, computer science, and critical feminist media-studies viewpoint.

The reason behind our contribution lies in the idea that gender is a co-shaped, changing part of human identity tied into the socio-materiality of gendered relations often treated as a binary dichotomy. For instance, trans and non-binary users recently proclaimed being misgendered on Twitter because the categories 'female' and 'male' do not match who they are [11]. Second, platform providers no longer have to learn sensitive details about a particular user or correctly group users into categories for advertising to be effective, as advertising has a high tolerance for classification errors [12]. Nonetheless, not considering a broader understanding of gender in platforms can be socially harmful and costly, as technology usage and implementation may lead to further exacerbation of existing biases, including gender, race, and other minorities [13-15].

---

[1] *See* https://help.twitter.com/en/rules-and-policies/data-processing-legal-bases

In this article, we provide background information on inferential analytics to elucidate how companies infer specific user attributes, including gender, and how these techniques may harm users' rights in Section 2. In Section 3, we explain the methods for this study, and we introduce the results in Section 4. Our findings suggest that Twitter's binary understanding of gender excludes those not fitting the category 'male' and 'female.' The results also show that inferring gender is part of Twitter's personalization trade-off and misgenders users in nearly 20% of the cases. Out of these cases, the LGBTQAI+ and straight females were misgendered more frequently than straight males. We discuss in Section 5, the lack of diversity in social media platforms, and the role designers play in accounting for inclusivity and diversity. We conclude by presenting our future work, which includes a more extensive and refined survey to investigate this issue and the user's impressions further.

2 Gendering algorithms

2.1 Profiling, inference analytics, and discrimination

Profiling techniques like regression, classification, or clustering, mainly ascribe properties to people [16]. These methods infer distinct people's traits from different inputs of data, originating either from the very person (i.e., predicting recidivism based on someone's criminal record) or others (i.e., others who ordered these shoes, also like these shoes). Organizations use inferential analytics to induce user preferences using sensitive attributes such as race, gender, or sexual orientation, political interests, and opinions [17-19]. These techniques can predict behaviors for marketing purposes and influence behavior for profit [20]. A critical feature of inferential analytics is that companies infer information from data not directly or indirectly provided by data subjects [4]. Besides, these inferences may be precise (like inferring age from the date of birth) or estimates (like inferring emotional states such as happiness or even intelligence from Facebook likes) [21]. In this way, data analytics can predict qualities that a data subject may not want to disclose and attributes that a data subject does not even know and ascribe them to a person.

One of the parameters used to infer attributes from people is the 'like' button on many social media platforms [22]. In other words, what users like online tell something about who they are, including a person's income pretty accurately [23]. Gender can also be inferred from Facebook likes with very high accuracy [21]. With approximately 250 Facebook likes, gender could be predicted with accuracy rates of 93%. Although these seem many, gender could be predicted with accuracy rates of about 70% when using only five Facebook likes. Moreover, when using only one single Facebook like, the accuracy rates were approximately 60% for gender predictions. According to Kosinski et al., predictions for homosexuality were about 88% accurate for gays and 75% for lesbians, and predictions on being single versus in a relationship were about 67% accurate [21].

Inferential analytics may have some benefits. For instance, it can be a tool to fill gaps in fragmentary datasets or check the accuracy of available data by matching inferred data with the contested data. In this way, datasets enriched with many inferred attributes are likely to have higher levels of completeness and precision. In big data analytics, completeness and correctness of data is not a strict condition but can contribute to getting more well-defined and reliable results. Companies can identify that a particular customer prefers to consume video instead of text content, or is interested in learning about particular topics, like travel, fashion, or food. Companies use this information to personalize the user experience to fit the preferences of that particular individual.

However, inferential analytics has some drawbacks. When people's attributes are predicted, privacy is at stake, especially if people did not want to disclose specific personal information. Furthermore, these inferences may contain errors, leading to biased and unfair decisions, and may lead to self-fulfilling prophecies [24]. These effects may amplify inequality, undermine democracy, and further push people into categories that are hard to break out [25].

Machine learning and data mining tools can be developed so that they do not grant discriminating patterns, such as gender stereotypes or profiles, something called discrimination-aware data mining [26]. The underlying idea is not to limit the data input (such as gender data), but to prevent the algorithms from yielding gender-based patterns, since not using gender data may still allow for predicting gender and indirect discrimination (discrimination by proxy). Focusing on the algorithms' design can prevent this when using gender in the development of data-driven decision models [27].

2.2 Gender inferences

Gender recognition can be useful to support applications, such as face recognition and smart human-computer interface aid in other domains [28]. Developers use algorithmic gender classification in human-computer interaction, security and surveillance industry, law enforcement, psychiatry, demographic research, education, commercial development, telecommunication, and mobile application and video games [28-30]. Depending on the application and dataset, developers may use vision-based and biological information-based methods to make inferences [29].

Gender classification systems (GCS) are trained using a training dataset (or corpus) of structured and labeled data. These labels categorize data, and the features within, as either masculine or feminine [28]. Training a GCS builds a classification algorithm (or classifier) that categorizes features, such as body movements, physiological and behavioral characteristics, and facial features [28], found in new data by comparing it to labeled features in the dataset. A GCS uses a feature extraction algorithm, classifier, and a dataset to make an inference [29].

Classifiers are trained in machine learning models. Exemplary models include neural networks [28], K-nearest neighbor [30], support vector machine [31], and Adaboost [32]. A classifier infers gender from video, images, or text, and the process is usually straightforward. First, data such as video or images are parsed into a GCS. Using a feature extraction algorithm, it then extracts features from the data, such as static body features, dynamic body features, apparel features, and biometrics [29,31,33]. Finally, it compares those features using a classifier to a feature dataset, which is categorized by gender and maps them to either category, inferring gender based on similarities in features [28,30].

Similarly, a text-based GCS infer gender using features such as language, vocabulary, and frequency of words [29]. Text-based GCS uses content found in forums, chat rooms, and social media, extracting features using text mining [29,34]. Beyond language, Corney et al. extended text-based feature extraction further into the typography field, training a classifier to make gender inferences based on style markers, structural characteristics, and gender-preferential language [35].

In the literature, developers have used classifiers to support text analysis techniques (e.g., sentiment and content analysis). Park et al. developed a GCS that supports sentiment analysis to identify the gender of persons making posts found on an online AIDS-related bulletin board [34]. The author's GCS used a feature dataset that paired gender with the frequency of sentiment-driven words. During training, the GCS learned that women tended to use the words 'thank', 'bless, 'scary,' and 'illness' about twice as often as men who used 'accurate,' 'important,' 'issue,' and 'aches' twice as often as women [34].

Several studies have made use of freely available Twitter user posts (or tweets) to train a GCS and infer the gender of other users [36-39]. Lopes Filho et al. utilized a dataset categorizing gender by 60 textual meta-attributes associated with characters, syntax, words, structure, and morphology for the extraction of gender expression linguistic cues in tweets after that gender inference [38]. The authors compared different classifiers, finding that each accurately determined the gender of Twitter users, 63.5%, 61.96%, and 68.08% of the time. Using word unigrams, hashtags, and psychometric properties as features, the GCS developed by Fink, Kopecky, & Morawski predicted the gender of Twitter users with 80% accuracy [39].

3 Methods

Available scientific literature focuses on how gender can be inferred from user attributes [36-39]. However, there are not many studies that have compared the users' reported gender, the inferred gender from those attributes, and its correctness, although it is an increasing area of interest in social sciences [9]. How algorithms exacerbate existing biases and affect marginalized communities is also a nascent area of specialization [9,40,41]. Our work contributes to the literature on algorithmic bias and discrimination by exploring misgender on social media platforms like Twitter. Our work illustrates inferential analytics discrimination by means of a specific case study.

We conducted a short survey disseminated using Twitter. For four days, from 22 to 26 May 2020, N=109 Twitter users responded. The online survey was prepared in Qualtrics and included five specific questions revolving around whether Twitter algorithms inferred users' gender and whether it was correct. In particular, we asked what the user's sexual orientation (Q1), their gender identity (Q2), the pronouns they use (Q3), whether they provided Twitter with their gender information (Q4), and, if not, whether that was correctly assigned (Q5).[2] We gave the users instructions on how to find their assigned gender on Twitter,[3] and we processed anonymous data and surveyed the adult population only.

At the end of the survey, we exported, tabulated, and analyzed the data using Microsoft Excel Spreadsheet Software. The lead author analyzed the survey data, and the remaining authors examined the tabulated data and analysis to discuss discrepancies and ensure the reliability of the results. Understanding, empirically, if Twitter (mis)genders users lays the foundation for future work into the potential impacts in an extensive survey.

All of the respondents completed the survey entirely. However, the online survey has some limitations, including the small number of the respondents, which only amounts to N=109. This is due to the quick nature of our survey, within a limited, four-day timeframe. Another limitation may be the limited representativeness of the sample, which seems to over-represent the LGBTQAI+ community compared to the number of straight people in society in general. This potential bias may be due to one or more of the following reasons. First, the LGBTQAI+ community may be overrepresented among Twitter users (Twitter does not provide data on this) or be overrepresented in our Twitter networks. Moreover, people from the LGBTQAI+ community may be more inclined to complete the survey, perhaps because the survey topic appealed to them, as it may relate to past experiences of gender stereotyping or misrepresentation, on Twitter or elsewhere.

4 Results

Based on the conducted online survey, we identify the following data:

---

[2] The exact wording of the questions was: 1) What is your sexual orientation?; 2) What is your gender identity?; 3) What pronouns do you use?; 4) Did you at one point provide Twitter with your gender?; 5) If you did not include your gender in your profile, the gender that appears in your profile may have been assigned by Twitter, is the gender appearing here correct?

[3] To know the gender assigned by Twitter, go to picture > settings and privacy > account > your Twitter data > password > account > confirm password > gender.

Table 1. Twitter gender inference accuracy in a N=109 sample data.

	Self-reporting				Incorrect by Twitter				% Incorrect by Twitter			
	Female	Male	Non-binary	All	Female	Male	Non-binary	All	Female	Male	Non-binary	All
Straight	37	25		62	6	2		8	16%	8%		13%
Gay		24		24		6		6		25%		25%
Lesbian	2			2								
Bisexual	4	5	1	10	1	1	1	3	25%	20%	100%	30%
Asexual	3		2	5			2	2	0%		100%	40%
Questioning	2	1		3					0%	0%		
Other	2		1	3	2			2	100%			67%
Sum	**50**	**55**	**4**	**109**	**9**	**9**	**3**	**21**	**18%**	**16%**	**100%**	**19%**

Out of N=109 respondents, 19% had their gender wrongly assigned, whereas Twitter inferred users' gender correctly in 81% of the cases. Our central hypothesis revolved around differences between the self-reported gender identity (male), the sexual orientation of users (gay), and the correctness of the Twitter assigned gender (female). Comparing the variables' sexual orientation and the correctness of the inferred gender in our sample, there seems to be no correlation (R=0.009).

Twitter infers their users' identity from a wide variety of sources, such as information from the account, interactions with links, and cookie data[4], but not from their sexual orientation. However, how apparently fair algorithmic designs and categorizations have ulterior and unintended consequences in specific communities is well-known in the literature [3,42,43]. For instance, our collected data shows that, out of the misgendered Twitter users that we analyzed, only 38% were straight. Only 8% of the straight male respondents were misgendered, compared to 25% of gay males and 16% straight females. Individuals that self-reported as bisexuals were misgendered in 25% of the cases for bisexual females and 20% for bisexual males. Respondents identified as non-binary were wrongly gendered in all cases. These results show that the LGBTQAI+ community and straight females were more often misgendered than straight males in our sample, although not only gay men compared to straight men, as we first hypothesized. Moreover, women and non-binary are usually more misgendered by Twitter than males.

The findings also seem to suggest that lesbian and questioning people are less likely to be misgendered, although the numbers are small in our sample (two lesbian and three questioning participants). One questioning and one lesbian participant answered that they had provided their gender to Twitter, meaning the gender of the remaining were inferred correctly by Twitter. The findings also show that non-binary participants (n=2) were misgendered, both of whom were also asexual participants. However, there were other asexual participants (n=3) whose gender (female in all

[4] *See* https://help.twitter.com/en/rules-and-policies/data-processing-legal-bases

cases) was correctly inferred by Twitter (each answered 'I do not know' when asked if they provided Twitter with their gender).

Of the 109 participants, only 15% provided Twitter with their gender, whereas 24% did not, and 61% responded do not remember doing so. 42% of those who did not provide Twitter with gender were from the LGBTQAI+ community. Of the 16 participants who provided their gender, all but one answered 'Yes' about whether or not the gender's appearing on their Twitter profile was correct. An outlier was an asexual, nonbinary person. This may indicate that either (1) some of those 16 participants were mistaken and had entered their gender into their Twitter profile previously or (2) Twitter may infer gender and change the one entered by the user.

Other findings resulted from discussions over Twitter, where we shared the online survey. Some respondents openly reported that Twitter used to misgender them, but that now Twitter gendered them correctly, probably due to their increasing interest in gender equality. Other respondents mentioned they had two profiles, but that Twitter misgendered the profile they used the most. A respondent suggested that, although gay, Twitter assigned his gender correctly, while another was surprised to be considered 'female' while being a 'male.'

5 Discussion

5.1 Misgendering in social media is discriminatory

Research affirms that gender identity is primarily subjective and internal, which juxtaposes with the idea that gender can be recognized automatically, at least with the state of art GCS [9]. Moreover, misgendering users via automated gender recognition systems have adverse implications, some of those being that they reinforce gender binarism, undermine autonomy, are a tool for surveillance, and threaten safety [9].

They also exacerbate existing stereotypes. Classifiers trained on real-world datasets are often biased because the data used to train them contains racial and gender stereotypes [44-47]. Female names are more associated with family than career words, with arts more than mathematics and science [48,49]. Datasets imSitu and MS-COCO are significantly gender-biased and "models trained to perform prediction on these datasets amplify the existing gender bias when evaluated on development data" [50]. For example, the verb 'cooking' is heavily biased towards females in a classifier trained using the imSitu dataset, amplifying existing gender stereotypes [50]. The same gender biases have been shown in natural language processing [51,52], another method used to support gender classifiers [53].

To be misgendered reinforces also the idea that society does not consider a person's gender real, causing rejection, impacting self-esteem and confidence, the felt authenticity, and increasing one's perception of being socially stigmatized [10]. These gender biases in the offline world may propagate to artificial intelligence if not addressed carefully [3], something worrying if we understand that available research suggests that many individuals perceive automatic misgendering as more harmful than human misgendering [9].

When the tools used to extract patterns and profiles from data are not transparent, moreover, it may be hard for people to contend any decisions resulting from this, which may impede their freedom and autonomy and may inadvertently affect their privacy. In the EU, the collecting and processing of personal data are protected under the General Data Protection Regulation (GDPR), which also addresses discrimination issues in datasets. However, enforcing legislation in such cases is very challenging. For data protection, scholars note that information about a person's gender, age, financial situation, geolocation, and online profiles are not sensitive data according to Article 9 of the GDPR, despite often being grounds for discrimination [19]. Not being 'sensitive data' translates in not enjoying the extra protection (such as users' informed and explicit consent) sensitive information such as race, religion, or sexual orientation have. Discrimination in (patterns and profiles extracted from) large datasets can be hard to detect. Indirect discrimination takes place unintentionally when users are unaware of any harm profiles may be doing. However, it may also be the case that companies use profiles precisely to conceal discrimination, also named masking [24]. Because direct discrimination in data is hard to detect, and indirect discrimination is nearly impossible to detect, it can be challenging to enforce equal treatment acts and data protection legislation.

Many forms of discrimination are illegal in most Western jurisdictions. Not hiring someone based on gender, ethnicity, or sexual orientation, or because they have a criminal record, is prohibited. Not every decision based on the sensitive characteristics mentioned is forbidden, however. Legislation that forbids discrimination for specific characteristics concerns lists the characteristics that may not serve as a basis for making decisions (including gender, ethnicity, political preferences, trade union membership, or sexual orientation). Nonetheless, 'softer' forms of discrimination may occur in the formation of friendships, in the form of stigmatization of specific population groups. On a larger scale, this could lead to social polarization and segregation. For now, misgendering or addressing someone with the wrong pronoun is not sufficiently grave to be considered harassment under certain specific legal provisions, although several people advocate for remedies that make justice to these acts [54].

5.2 Organizations controversially infer gender for legitimate interests

Twitter makes inferences about users' accounts, including interests, age, and gender, to provide features such as account suggestions (e.g., suggested contacts, promoted accounts for the user to follow), advertising, recommendations, and timeline ranking.[5] Twitter uses users' content, activity, relationships, and interactions to genderize content production patterns [55], infer gender, and make these suggestions.[6] Twitter justifies making inferences about interests, age, and gender, stating that it helps tailor content to users, keeps the platform safe and enjoyable for all users and enables Twitter to provide compelling, targeted advertising. In other

[5] *See* https://help.twitter.com/en/rules-and-policies/data-processing-legal-bases; *see* also https://help.twitter.com/en/using-twitter/account-suggestions.
[6] *See* https://help.twitter.com/en/using-twitter/account-suggestions

words, users have to accept the trade-off if they want to have a personalized Twitter account.

The GDPR lists a limited number of legal grounds for data processing, including consent, the performance of a contract, or legitimate interests. Twitter states that it makes “inferences about your account - such as interests, age, and gender” for “legitimate purposes.” The appeal to legitimate interests as a legal basis for data processing is controversial, as legitimate purposes are only a solid legal basis if there is a necessity. It is questionable, however, whether gender inferences are necessary for Twitter. Although the legitimate interest seems less constraining than other grounds for data processing, it should not be considered a ‘last resort’ when all other grounds for lawful data processing fail [56].

Legitimate interest is the most appropriate legal ground for data processing if the data controller uses people's data in ways they would reasonably expect and have a minimal privacy impact, or where there is a compelling justification for the processing. If controllers choose this legal ground, they 'should take on extra responsibility for considering and protecting people's rights and interests' [57]. Thus, three elements configure the basis for legitimate interest: identify the legitimate interest, show that the processing is necessary to achieve it, and balance it against the individual's interests, rights, and freedoms.

Our survey findings highlight a significant number of misgendered users and question whether Twitter did balance their interests against individuals’ interests. First, out of the 109 participants, only 15% provided Twitter with their gender, while Twitter inferred their gender. Second, our results show that the LGBTQAI+ community and straight females are more often misgendered than straight males. Third, remedies for opposing the processing seem not to correspond in magnitude to the subsequent impact of being misgendered. A user can modify or rectify the inferred gender, but cannot escape that inference unless she unticks the box of personalization. Making users choose between these two is as if in times of COVID-19, developers made users choose between health or privacy [58]. Moreover, it results in a privacy paradox: the gender inference causes a privacy issue (i.e., disclosing information people may want to keep to themselves), but to address this, users have to provide additional information, disclosing even more (or more detailed) information about themselves [24]. This is particularly problematic for communities that society has been historically discriminated against and in which gender is a sensitive part of their identity [11,59].

5.3 Accounting for diversity in social media

Platforms exclude and misrepresent a large number of potential users if they are not respectful and inclusive towards their gender identity or sexual orientation. The assumption that gender is physiologically-rooted harms trans people overall by essentializing the body as the source of gender, and also harms non-binary people, who cannot be accurately classified [10]. As Fergus announced, transexual and non-binary users reported being misgendered by Twitter, which we found to be the case in our survey (100% of the non-binary participants reported being misgendered) [11]. These findings may result from the fact that twitter gender classifiers do not account for diversity and work on a male/female binary categorization that does represent other people's gender expression, not doing justice to the freedom of identity that everyone should have.

Our study shows that when it comes to diversity and more inclusive engagement, social media platforms like Twitter still have a long way to go to become a more open and welcoming platform for a wide variety of users. Misgendering users in the background is not good practice, and can lead to privacy and discrimination issues, beyond echoing deeply rooted stereotypes. The lack of diversity in marketing strategies is apparent when users can be gendered as male or female only. However, making strategies for a diverse engagement with the "queer rainbow economy" can make for more affluent and more diverse revenue streams [60-62].

From all this, it is clear that digital identity and participatory culture play a massive role in the sense of self in the modern world and that there should be more effort to realize diversity and inclusion in the online world [63] to not perpetuate the normative view that certain collectives such as trans or non-binary do not exist [10].

6 Summary

An online survey showed that, out of N=109 respondents, Twitter correctly inferred users' gender in 81% of the cases, and 19% were misgendered. Although sexual orientation and misgender were not correlated, a closer look at the results shows that only 8% of the straight male respondents were misgendered, compared to 25% of gay males and 16% of straight females, while non-binary users were misgendered in all the cases.

Social media platforms like Twitter have economic incentives to know users' genders for commercialization and targeted advertisements. However, our investigation shows that inferring a user's gender with automated means clashes with the understanding that gender is subjective and internal. Misgendering has also alerting and broader consequences, leading to serious privacy, discrimination, autonomy, and self-identity issues. Misgendering reinforces gender stereotypes, accentuates gender binarism, undermines autonomy, and leads to toxic cultures and algorithmic bias [9,64]. Moreover, misgendering causes a feeling of rejection, impacting one's self-esteem, confidence, and authenticity, increasing social stigmatization [10].

If users do not provide a gender parameter choice themselves, platforms may infer the user's gender from a wide variety of data sources, including personal data. Therefore, gender classifiers should account for diversity and inclusion, using a more accurate understanding of gender to represent contemporary society fully. Otherwise, inferential analytics may reinforce existing biases about gender stereotyping. By including diverse users early on, during the design, and with the possibility to provide feedback afterward, the technology can be experienced as more just and fair. Inclusive engagement that reflects on the users as not homogeneous can have a positive impact on technology.

By identifying how inferential analytics may reinforce gender stereotyping and affect marginalized communities, we hope to continuously contribute to promoting the online account for privacy, diversity, inclusion, and advocate for the freedom of identity that everyone should have online and offline. Looking forward, a more robust survey ought to be undertaken to further explore the social implications of gender inference on Twitter, such as discrimination and diversity in social media.

7 References

1. Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y.:. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In: Proceedings of the ACM Symposium On Usable Privacy and Security (SOUPS), pp. 1-15 (2012).
2. Davenport, T. H., & Beck, J. C.: The attention economy: Understanding the new currency of business. Harvard Business School Press, Massachusetts (2001).
3. Caliskan, A., Bryson, J. J., & Narayanan, A.: Semantics derived automatically from language corpora contain human-like biases. Science 356(6334), 183-186 (2017).
4. Custers, B. H. M.: Profiling as inferred data: Amplifier effects and positive feedback loops. In: Bayamlioglu, E., Baraliuc, I., Janssens, L., & Hildebrandt, M. (eds.) Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen', pp. 112-116. Amsterdam University Press, Amsterdam (2018).
5. Office of the High Commissioner Human Rights: Gender Stereotypes, https://www.ohchr.org/EN/Issues/Women/WRGS/Pages/GenderStereotypes.aspx (2020).
6. Hentschel, T., Heilman, M. E., & Peus, C. V.: The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. Frontiers in psychology 10(11), 1-19 (2019).
7. Grant, A., Grey, S., & van Hell, J. G.: Male fashionistas and female football fans: Gender stereotypes affect neurophysiological correlates of semantic processing during speech comprehension. Journal of Neurolinguistics 53, 100876, 1-14 (2020).
8. Butler, J.: Gender trouble, feminist theory, and psychoanalytic discourse. In: Nicholson, L. J. (ed.) Feminism/Postmodernism, pp. 324-341. Routledge, New York (1990).
9. Hamidi F., Scheuerman M. K., & Branham S.M. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1-13 (2018).
10. Keyes, O. The misgendering machines: Trans/HCI implications of automatic gender recognition. In: Proceedings of the ACM on Human-Computer Interaction, pp. 1-22 (2018).
11. Fergus, J.: Twitter is guessing users' genders to sell ads and often getting it wrong. Input, https://www.inputmag.com/tech/twitter-guesses-your-gender-to-serve-you-ads-relevant-tweets-wrong-misgendered (2020).
12. Wachter, S.: Affinity profiling and discrimination by association in online behavioural advertising. Berkeley Technology Law Journal 35(2), 1-74 (2020).
13. Bray, F.: Gender and technology. Annu. Rev. Anthropol 36, 37-53 (2007).
14. Schiebinger, L.: Scientific research must take gender into account. Nature 507(7490), 9-9 (2014).
15. Hao, K.: Facebook's ad-serving algorithm discriminates by gender and race. MIT Technology Review, https://www.technologyreview.com/2019/04/05/1175/facebook-algorithm-discriminates-ai-bias/ (2019).

16. Calders, T., and Custers, B. H. M.: What is data mining and how does it work? In: Custers, B. H. M., Calders T., Schermer B., Zarsky T. (eds.) Discrimination and Privacy in the Information Society, pp. 27-42. Springer, Heidelberg (2013).
17. Jernigan, C., & Mistree, B. F.: Gaydar: Facebook friendships expose sexual orientation. First Monday, 14(10), 1-28 (2009).
18. Thorson, K., Cotter, K., Medeiros, M., & Pak, C.: Algorithmic inference, political interest, and exposure to news and politics on Facebook. Information, Communication & Society, 1-18 (2019).
19. Wachter, S., & Mittelstadt, B.: A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Colum. Bus. L. Rev. 2019(2), 494-620 (2019).
20. Zuboff, S.: Big other: surveillance capitalism and the prospects of an information civilization. Journal of Information Technology 30(1), 75-89 (2015).
21. Kosinski, M., Stillwell, D., & Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. In: Proceedings of the National Academy of Sciences, 110(15), pp. 5802-5805 (2013).
22. Roosendaal, A.: Facebook tracks and traces everyone: Like this! Tilburg Law School Legal Studies Research Paper Series No. 03/2011, 1-10 (2010).
23. Matz, S. C., Menges, J. I., Stillwell, D. J., & Schwartz, H.: A. Predicting individual-level income from Facebook profiles. PloS one 14(3), 1-13 (2019).
24. Custers, B.H.M.: Data dilemmas in the information society. In: Custers, B. H. M., Calders, T., Schermer, B., Zarsky, T. (eds.) Discrimination and Privacy in the Information Society, pp. 3-26. Springer, Heidelberg (2013).
25. O'Neil, C.: Weapons of math destruction: How big data increases inequality and threatens democracy. Crown, New York (2016).
26. Kamiran, F., Calders, T., & Pechenizkiy, M.: Techniques for discrimination-free predictive models. In: Custers, B. H. M., Calders, T., Schermer, B., & Zarsky, T. (eds.) Discrimination and Privacy in the Information Society, pp. 223-239. Springer, Heidelberg (2013).
27. Zliobaite I. & Custers B.: Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. Artificial Intelligence and Law 24, 183-201 (2016).
28. Rai P., & Khanna P.: Gender classification techniques: A review. In: Wyld D., Zizka J., Nagamalai D. (eds) Advances in Computer Science, Engineering & Applications, Advances in Intelligent and Soft Computing, vol. 166, pp. 51-59. Springer, Berlin, Heidelberg (2012).
29. Lin, F., Wu, Y., Zhuang, Y., Long, X., & Xu, W.: Human gender classification: A review. International Journal of Biometrics 8(3/4), 275-300 (2015).
30. Khan, S. A., Ahmad, M., Nazir, M., & Riaz, N.: A comparative analysis of gender classification techniques. Middle East Journal of Scientific Research 20, 1-13 (2014).
31. Li, B., Lian, X.-C., & Lu, B.-L.: Gender classification by combining clothing, hair and facial component classifiers. Neurocomputing 76(1), 18-27 (2012).
32. Mathivanan, P., & Poornima, K.: Biometric authentication for gender classification techniques: A review. J. Inst. Eng. India: Ser. B 99, 79–85 (2018).
33. Kumar, D., Gupta, R., Sharma, A., & Saroj, S. K.: Gender classification using skin patterns. In: Proceedings of 2nd International Conference on Advanced Computing and Software Engineering, pp. 472-476 (2019).
34. Park, S. & Woo, J.: Gender classification using sentiment analysis and deep learning in a health web forum. Applied Sciences 9, 1249, 1-12 (2019).
35. Corney, M., De Vel, O., Anderson, A., & Mohay, G.: Gender-preferential text mining of e-mail discourse. In: Proceedings of the 18th IEEE Annual Computer Security Applications Conference, pp. 282-289 (2002).

36. Nieuwenhuis, M., & Wilkens, J.: Twitter text and image gender classification with a logistic regression n-gram model. In: Proceedings of the Ninth International Conference of the CLEF Association, pp. 1-9 (2018).
37. Garibo-Orts, O.: A big data approach to gender classification in Twitter. In: Proceedings of PAN at CLEF, pp. 1-5 (2018).
38. Lopes Filho, J. A. B., Pasti, R., & De Castro, L.: Gender classification of Twitter data based on textual meta-attributes extraction. In: Rocha Á., Correia A., Adeli H., Reis L., Mendonça Teixeira M. (eds.) New Advances in Information Systems and Technologies, Advances in Intelligent Systems and Computing, vol. 444, pp. 1025-1034. Springer, Cham (2016).
39. Fink, C. & Kopecky, J. & Morawski, M.: Inferring gender from the content of tweets: A region specific example. In: Sixth International AAAI Conference on Weblogs and Social Media. Association for the Advancement of Artificial Intelligence (AAAI), pp. 459-462 (2012).
40. Willson, M.: Algorithms (and the) everyday. Information, Communication & Society 20(1), 137-150 (2017).
41. Noble, S. U.: Algorithms of oppression: How search engines reinforce racism. NYU Press, New York (2018).
42. Ito, J.: Supposedly 'fair' algorithms can perpetuate discrimination. MIT Media Lab. https://www.media.mit.edu/articles/supposedly-fair-algorithms-can-perpetuate-discrimination/ (2019).
43. Gomes, A., Antonialli, D., & Dias-Oliva, T.: Drag queens and artificial intelligence: Should computers decide what is 'toxic' on the internet? Internet Lab Blog, https://www.internetlab.org.br/en/freedom-of-expression/drag-queens-and-artificial-intelligence-should-computers-decide-what-is-toxic-on-the-internet/ (2019).
44. Torralba, A., & Efros, A. A.: Unbiased look at dataset bias. In: Proceedings of Computer Vision and Pattern Recognition 2011, pp. 1521-1528 (2011).
45. Buolamwini, J., & Gebru, T.: Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, 81, pp. 77-91 (2018).
46. Font, J. E., & Costa-jussà, M. R.: Equalizing gender bias in neural machine translation with word embeddings techniques. In: Proceedings of the First Workshop on Gender Bias in Natural Language Processing, pp. 147–154 (2019).
47. McDuff, D., Ma, S., Song, Y., & Kapoor, A.: Characterizing bias in classifiers using generative models. arXiv 1906.11891, 1-11 (2019).
48. Nosek, B. A., Banaji, M. R., & Greenwald, A. G.: Harvesting implicit group attitudes and beliefs from a demonstration web site. Group Dynamics: Theory, Research, and Practice 6(1), 101-115 (2002a).
49. Nosek, B. A., Banaji, M. R., & Greenwald, A. G.: Math= male, me= female, therefore math≠ me. Journal of Personality and Social Psychology 83(1), 44-59 (2002b).
50. Zhao, J., Wang, T,. Yatskar, M., Ordonez, V., & Chang, K-W.: Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2979-2989 (2017).
51. Zhou, P., Shi, W., Zhao, J., Huang, K-H., Chen, M., Cotterell, R., & Chang, K-W,: Examining gender bias in languages with grammatical gender. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 5279-5287 (2019).
52. Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K-W., & Wang, W.: Mitigating gender bias in natural language processing: Literature review. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1630-1640 (2019)

53. Campa, S., Davis, M., & Gonzalez, D. Deep & machine learning approaches to analyzing gender representations in journalism. In: Proceedings of CS224n: Natural Language Processing with Deep Learning, pp. 1-8 (2019).
54. Ashley, F.: No, pronouns won't send you to jail: The misunderstood scope of Bill C-16. Medium, https://medium.com/@florence.ashley/no-pronouns-wont-send-you-to-jail-43c268cffd55 (2017).
55. Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., . . . Stern, M. J.: Digital inequalities and why they matter. Information, Communication & Society 18(5), 569-582 (2015).
56. Article 29 Working Party: A29WP Opinion 06/2014 on the notion of legitimate interest of the data controller under Article 7 of Directive 95/46/EC. 844/14/EN, WP 217, adopted 9 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf (2014).
57. Information Commissioner's Office: Legitimate interests, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/ (2020).
58. Harari, Y. N.: The world after coronavirus. Financial Times, https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75 (2020).
59. McLemore, K. A.: Experiences with misgendering: Identity misclassification of transgender spectrum individuals. Self and Identity 14(1), 51-74 (2015).
60. Brown, G.: Thinking beyond homonormativity: Performative explorations of diverse gay economies. Environment and Planning 41(6), 1496-1510 (2009).
61. Keating, A., & McLoughlin, D.: Understanding the emergence of markets: A social constructionist perspective on gay economy. Consumption, Markets and Culture 8(2), 131-152 (2005).
62. Um, N. H.: Seeking the holy grail through gay and lesbian consumers: An exploratory content analysis of ads with gay/lesbian-specific content. Journal of Marketing Communications 18(2), 133-149 (2012).
63. Jenkins, H., & Ito, M.: Participatory culture in a networked era: A conversation on youth, learning, commerce, and politics. John Wiley & Sons, New Jersey (2015).
64. Lambrecht, A., & Tucker, C. E.: Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads. Management Science 65(7), 2966-2981 (2018).