Future Humanities

WILEY

future humanities

ORIGINAL PAPER OPEN ACCESS

Operationalization of (Trans)gender in Facial Recognition Systems: From Binarism to Intersectionality

Giovanni Pennisi

Department of Philosophy and Education Sciences, The University of Turin, Turin, Italy

Correspondence: Giovanni Pennisi (g.pennisi@unito.it; giovannipennisi22@gmail.com)

Received: 22 December 2023 | **Revised:** 18 June 2024 | **Accepted:** 18 June 2024

Funding: The author received no specific funding for this work.

Keywords: AGR | gender | intersectional approaches | operationalization | transgender

ABSTRACT

In a paper published in 2018, Os Keyes investigated how the literature on Automated Gender Recognition systems (AGRs) conceived gender, finding that 94.8% of the papers treated it as binary, 72.4% as immutable and 60.3% as a physiological component. In the author's view, this is indicative of an *operationalization* of gender, that is, the assumption that the latter is a discrete and objectively applicable parameter. Keyes claims that such a vision is blind to the performative aspects of gender and particularly dangerous for transgender people. Here I will follow on these remarks, providing several examples that show how AGR systems' failures in recognizing the faces of transgender people are capable of both perpetuating and amplifying gender stereotypes and inequalities. Then, I will introduce the notion of *intersectionality*, which is the idea that humans 'sit at the crossroads' of many physical, social, and political factors, whose combination generates dynamics of discrimination or privilege. I will focus on a subfield of intersectional studies, that is, intersectional stereotyping, which explains how we usually make assumptions and judgments about an individual or group of people based on multiple social identities or categories they belong to, such as their race, gender, sexual orientation, class, religion and ability. I will argue that this area of research provides us with a set of knowledge that might help us rethink and redesign the data sets for AGR. Specifically, I will draw on three key notions of intersectional stereotyping—'perceiver goals', 'category accessibility' and 'category fit'—and use them to envision new ways of collecting images for assessing gender through facial recognition. Finally, I will explicate why my observations call for an urgent integration between computer science and gender studies.

1 | Introduction

In a paper published in 2018, Os Keyes provided a critical review of the literature on a subfield of facial recognition technologies (FRT), that is, Automated Gender Recognition systems (AGRs), focusing on how gender was described and framed. The author found that both gender-centred papers (works that gave a precise definition of gender to implement it in AGRs) and gender non-centred papers (works that used AGR as a test scenario for facial recognition systems employed for more general purposes) relied on a biased conception of gender. Specifically, Keyes showed that 94.8% of the literature treated gender as *binary*, 72.4% as *immutable,* and 60.3% as a *physiological component*. In the author's view, these data prove unequivocally that gender gets *operationalized*, that is, it is intended as an objective and measurable parameter.

For Keyes, the main reason behind this operationalization is the unacceptable confusion between sex and gender, the unawareness that whereas 'sex is a determination made through the application of socially agreed upon *biological* criteria for classifying persons as females or males, [...] gender is the *activity* of managing situated

conduct in light of normative conceptions of attitudes and activities appropriated for one's sex category' (West and Zimmerman 1987, 127, emphasis added). Thus, what the literature on AGR really lacks is the idea that gender is not only a social construct, but one that is best captured by terms that invoke a dynamic process: 'gender proves to be performative—that is, constituting the identity it is purported to be. In this sense, gender is always a doing, though not a doing by a subject who might be said to preexist the deed' (Butler 1990, 34).

Importantly, the collapse of gender into the category of biological criteria produces a misclassification and an under-representation of a wide array of people, such as non-binary, gender non-conforming, and transgender folks. In this paper, I examine the latter case, discussing some examples of misgendering operated by facial recognition systems and then trying to figure out whether it is possible to make the data sets employed by AGRs more inclusive, or if, as Keyes (2018) suggests, 'designers and makers should quite simply avoid implementing or deploying AGR' (p. 14). To do this, I explore a subfield of intersectional studies—intersectional stereotyping. I will focus on three of its key notions—'perceiver goals', 'category fit' and 'category accessibility'—to use them as a basis for proposing new methods of collecting images for assessing gender through facial recognition. Finally, I will explicate why my observations call for an urgent integration between computer science and gender studies.

2 | How Computers See Gender

The data and reflections on the operationalization of gender were corroborated by a study carried out by Scheuerman, Paul, and Brubaker (2019), who tried to assess how computers see gender by testing the accuracy of gender classification of four popular Face Analysis Systems: Amazon, Clarifai, IBM and Microsoft. The authors constructed a data set made of 2450 pictures taken from Instagram, which were labelled with several hashtags containing seven different genders (#man, #woman, #transman, #transwoman, #agender, #genderqueer, #nonbinary). Then, they calculated the accuracy of each service in assigning the correct label to every image, using the gender hashtags provided by the owners of the Instagram profiles as ground-truth data.

First of all, the authors found that none of the AGRs was able to identify any of the subjects in the pictures as agender, genderqueer or non-binary, since they were all based on a classification system that did not include options outside the binary opposition between male and female. Second, they noticed a consistent and significant discrepancy in the precision of the labelling of pictures portraying a man or a woman as compared to that of pictures portraying a transman or a transwoman. Specifically, they reported an average true positive rate of 87.3% for photographs of transwomen as opposed to 97.6% for photographs of women, and a true positive rate of 70.5% for photographs of transmen as opposed to 98.3% for photographs of men. According to the authors, these results show 'that the training data used to train FA services does not include transgender individuals—at least those who do not perform gender in a cisnormative manner' (p. 15).

I will discuss this issue below. Before that, however, there is a question that must be addressed, which already surfaced from Keyes' discussion on the misgendering power of Artificial Intelligence: given how hard it is to identify a trait that, rather than being immutable and biologically based, is performative and might change over time, should we just quit implementing—if not designing—systems trained to automatically detect gender? More concisely, do we really need AGRs?

3 | Do We Need AGRs?

Even before the failure of AGRs in assessing the gender of transgender people was empirically tested, concerns about the use of such kind of technology were raised by the trans community. Hamidi, Scheuerman, and Branham (2018) conducted a series of semi-structured interviews with transgender individuals, including three technology designers, to understand their reactions and worries about the possibility of being categorized on the basis of gender by algorithms. The authors found that all participants feared that this classification might have reflected the prejudices that they had to face in everyday interactions. More specifically, one of the most dreaded effects of AGR was the reification of the denial of their gender identity, which, in most of the participants' view, would be a byproduct of the perceived objectivity and 'superior intelligence' of computers. As one of the subjects interviewed stated, 'computers are said to be a lot smarter than people ... I would feel a little bit worse if there was a software that looked at everything about me [and misgendered me]' (p. 5).

The debate over the problematic application of technologies that, being built on the same biased categories that regulate daily life, contribute to reinforcing rather than overcoming them, became particularly urgent after a series of distressful episodes involving transgender individuals. For instance, in 2018, a Uber driver had her account suspended[1] because of the Real-Time ID Check system implemented by the company, a system that requires drivers to occasionally pull over and take a selfie to testify the matching between this picture and the photo in the driver's license, could not assess the correspondence due to the changes in physical appearance that the driver was going through because of gender transition. This happened despite the fact that the driver had already documented her transition by periodically uploading new pictures on the Uber catalogue, and was described by Uber support as a glitch that might have happened again in the future. Not by chance, cases of transgender drivers having their account deactivated, if not being banned from the platform, continued to be reported, with some of the profiles being even signalled as 'fraudulent.'[2]

Events like these are to be framed within the broader spectrum of problems related to the automated identification of gender for control and security purposes: the so-called 'securitization of gender' (Currah and Mulqueen 2011), which reveals itself as particularly pernicious for transgender people. Importantly, the issues underlying the securitization of gender are never limited to the recognition of the subject's face, but often involve the possibility of making the whole body fall into a binary categorization. The paradigmatic example of a machine whose functioning is grounded on a biased and deterministic definition

of gender is the body scanner, the use of which often led to infamous episodes of discrimination and public humiliation suffered by transgender individuals, especially during the security procedures at the airport (Clarkson 2019; Beauchamp 2019). As clinical psychologist Katherine Rachlin stated while recalling an incident that occurred to a transgender friend of hers, who was flagged as a potential threat to the security of the airport after a body scanner signalled a 'gonadal anomaly', 'the full-body scanners became news long before they were actually used in local airports and were a major topic in therapy. Patients anticipated that they would be publicly outed by screeners who saw that there was a mismatch between a person's documents and presentation and their body parts … Patients had increased anxiety and even panic attacks just contemplating the possibilities' (Currah and Mulqueen 2011, 565). Again, we are faced with cases that call us to question ourselves about the utility of AGRs and to wonder whether the harm they generate is greater than the benefit society derives from them.

Another sensitive issue involving the trans community and the application of AGRs is privacy. In 2017, controversy sparked[3] after the discovery that many images taken from YouTube videos featuring individuals undergoing hormone replacement therapy (HRT) were grabbed without their consent to train an artificial intelligence to track potential terrorists who might have been using HRT for fooling facial recognition systems. When interviewed, one of the subjects whose pictures were forcefully included in the sample expressed all her concerns, claiming that those who work in computer science should be aware that a large segment of the trans population would not be happy in partaking in the improvement of systems that help people figure out how to 'spot the trans person'.

Considering all the shortcomings that AGRs have shown in categorizing individuals who challenge traditional ways of performing gender, and recognizing how hard it is to reduce gender to a list of labels good for algorithmic knowledge, some authors argue that the only solution is to ban this type of technology. In an interview with The Verge,[4] policy analyst at digital rights group Access Now Daniel Leufer said that a society that is dedicated to safeguarding human rights should prohibit AGRs, since their functioning clashes with the principle of allowing individuals to freely live and express their gender identity beyond the conventional male–female binary or contrary to the sex they were assigned at birth. Similarly, Barrett (2020) stands up against AGR, because 'the erasure of transgender identity by coding a refusal to acknowledge it into facial recognition systems is dehumanizing and regressive for the transgender and non-binary people whose identities the system ignores' (p. 251). The reasons behind the calls for a ban of AGRs—namely, protecting the groups 'that were already more vulnerable to surveillance and over-policing' (p. 247)—echo the reasons supporting the appeal to getting rid of FRT employed for control and security purposes (Selinger and Hartzog 2020; Dauvergne 2022). In turn, the anti-FRT stances inspired activists, artists and citizens to find creative strategies for escaping these forms of vigilance (see Thibault and Buruk 2021; Delupi 2023).

At this point, it seems pretty obvious that there are many experts who would answer 'no' to the question I opened this section with. It is fathomable that they would argue that the technical difficulties that need to be overcome to make AGR more precise (at least with transgender individuals) are disproportionately higher than the advantage we would gain by having gender designated as one of the criteria for letting people in or out of public places. However, from a philosophical perspective, it might be interesting to take up the challenge anyway, and explore possible ways for rendering the data sets employed by AGRs more accurate and inclusive. This is where intersectionality comes into play.

4 | Theoretical and Methodological Applications of Intersectional Research

The term *intersectionality* was coined by critical race theory leading scholar Kimberlé Crenshaw (1989). It refers to the idea that humans 'sit at the crossroads' of many physical, social, and political factors, whose combination determines the intensity of dynamics of privilege or oppression. Crenshaw discussed many cases in which this interaction was overlooked, such as the one that culminated with the lawsuit filed in 1976 by Emma DeGraffenreid and four other African American women against General Motors. In short, the lawsuit was filed because, during the 1970s recession, the company decided to lay off employees on a seniority basis, thereby cutting out especially African American women, who had been hired only after the enactment of the Civil Rights Act in 1964. The lawsuit was eventually lost by the five women, because the Court rejected both the charges of sexual and racial discrimination, since no White women nor Black men were affected by the wave of redundancies. This showed unequivocally that legal solutions in lawsuits could only tackle one form of marginalization at a time, completely ignoring intersectional aspects of discrimination (see Ryan 2019).

The intersectional framework has proven to be particularly useful for assessing the peculiar nature of the acts of intolerance that transgender people often have to deal with, and for providing a more sophisticated understanding of multiplicity in gender identification and its interaction with other identity traits. For example, Diamond and Butterworth (2008) gave an intersectional interpretation to the results of a 10-year longitudinal study of sexual identity development in four transgender women, finding that 'women who began to explore multiplicity and fluidity with respect to their gender identity became progressively more aware of multiplicity and fluidity in their erotic attractions as well' (p. 371).

The connection between gender and what the authors call the other 'identity statuses' was the focus of much research on transgender experience. Take, for instance, Chou's and Feagin's (2008) and Han's (2009) accounts of Asian transgender men-targeted stigmatization, from which distinctive patterns of discrimination based on the interlocking of gender, sexual and ethnic attributes emerge. These patterns can be compared to those generated by other identity stratifications, such as transgender Black men, to look for informative insights into the different forms of intersectional prejudice: as De Vries (2012) notes, whereas Asian transgender men are usually portrayed as feminine, submissive, harmless, and socially invisible, transgender Black men are deemed as exotic and

threatening and are treated as a hyper-visible menace accordingly.

Another way in which the intersectional framework can be used to highlight the intricacy of discrimination is suggested by Scheuerman, Branham, and Hamidi (2018). In this study, the authors assessed the level of safety that transgender and non-binary people perceived while using digital social platforms, reporting that most of the participants felt in danger on a regular basis online as well as offline and that all the negative comments they received could be read through an intersectional lens—for instance, trying to figure out whether one was targeted as a transgender woman of colour, as a transgender sex worker, or both. In the authors' view, the intersectional model is a precious theoretical and methodological tool not only because it provides a key to understanding complex and multifaceted phenomena (see also Scheuerman, Pape, and Hanna 2021), but also by virtue of its power to change the language used to describe them: it is the case, for example, of misogyny aimed at transgender women, and of misogyny aimed at transgender Black women, which gained respectively, in 2007 and 2015, the terms *transmisogyny* and *transmisogynoir* (Scheuerman, Branham, and Hamidi 2018, 22; see also Jones 2019).

Finally, an area of application where intersectionality showed its efficacy to frame problems in a nuanced way is HCI, and Automated Facial Analysis in particular. For some years now scholars have advocated for those who work in HCI to acknowledge intersectionality, and to implement it in software design and the construction of the data sets employed for facial classification and recognition (Rankin and Thomas 2019; Schlesinger, Edwards, and Grinter 2017; D'Ignazio and Klein 2016). To date, however, only one study (Buolamwini and Gebru 2018) has empirically tested the accuracy of facial analysis algorithms in recognizing identity traits, such as gender and race, following an intersectional approach. In the preliminary phase of this pioneering experiment, the authors carried out an in-depth examination of the data sets of two facial analysis benchmarks, IJB-A (a US government benchmark released by the National Institute of Standards and Technologies in 2015, which was made up of 500 facial images of public figures) and Adience (a gender classification benchmark released in 2014, which was made up of 284 unique individuals' facial images), finding that they were overwhelmingly composed of lighter-skinned subjects (79.6% for IJB-A, 86.2% for Adience). Therefore, Buolamwini and Gebru collected 1270 facial images of female (44.6%) and male (54.4%) parliamentarians from three African and three European countries, creating their own benchmark, named Pilots Parliaments Benchmark (PPB). In the last stage of the experiment, they used the PPB data set to assess the precision in gender classification of three commercial AGRs (Microsoft, Face++ and IBM), obtaining the results reported in the following table.

As we can see, the overlapping between female gender and darker skin tone produces outcomes that are way worse than any other, suggesting that, similarly to the cases I previously discussed, the misgendering power of AGRs might stem from an underrepresentation of certain social categories, such as people of colour, and women of colour in particular. Therefore, as the authors state, 'future work should explore intersectional error analysis of facial detection, identification and verification [...]. Inclusive benchmark data sets and subgroup accuracy reports will be necessary to increase transparency and accountability in artificial intelligence' (p. 11). To find solutions to the biases intrinsic to the functioning of facial analysis systems and AGR, thus, engineers and designers should open themselves to other fields of research, exploring areas where intersectionality has already left its mark. My proposal is to draw indications from the literature on intersectional stereotyping.

4.1 | Intersectional Stereotyping

The term *intersectional stereotyping* refers to the act of making assumptions and judgments about an individual or a group of individuals based on multiple social identities or categories they belong to, such as race, gender, sexual orientation, class, religion, ability and so forth. This phenomenon is well captured by Khader's 'intensification thesis' (2013), which 'supposes that gender oppression subjects all women to qualitatively similar harms and that race and class oppression increase the severity of those harms' (p. 68). Similar examples can be done by taking other possible intersections, such as gender, ethnicity and religion (Erentzen et al. 2023) into account, and analyzing how the interplay between these factors contributes to the strengthening of people's prejudices and discriminatory attitudes.

In recent years, the model built upon the intensification hypothesis has been challenged by another model, which apparently seems at odds with the principles of intersectional research: the lens-based account of intersectional stereotyping put forward by social psychologist Cristopher Petsko and colleagues (Petsko, Rosette, and Bodenhausen 2022; Petsko and Bodenhausen 2019, 2020). In this paradigm, intersectional prejudice is understood as a compartmentalized phenomenon, that is, as a way of stereotyping by targeting exclusively one single identity (or one intersection of identities) at a time, in response to social contingencies. Therefore, the lens-based approach is grounded on two key premises: that 'perceivers have a repertoire of functionally independent lenses in their minds that they use as frameworks for thinking about others' (Petsko, Rosette, and Bodenhausen 2022, 764), and that perceivers 'use one lens at a time in a given social context' (Petsko, Rosette, and Bodenhausen 2022).

Gender classifier	Darker male (%)	Darker female (%)	Lighter male (%)	Lighter female (%)	Largest gap (%)
Microsoft	94	79.2	100	98.3	20.6
Face++	99.3	65.5	99.2	94	33.6
IBM	88	65.3	99.7	92.9	34.4

At first glance, it would appear that the lens-based account fits within those models for which there is always an identity trait that is more salient than another for a certain social category—say, gender with respect to race for women of colour. However, as I extensively discussed, there are many scholars who reject this 'either/or thinking', and claim that women of colour experience 'both/and', 'multiple oppressions/jeopardies' or a 'racialized gendered' identity in combination (see De Vries 2012, 51). However, if one looks more closely at the issue, finds that the model proposed by Petsko and colleagues is not competing with, but rather complementary to more traditional intersectional accounts of discrimination. Whereas the latter are meant specifically at addressing and dismantling the systematic, hegemonic, and historical conditions that determine marginalization, the former does not deny them but simply focuses on more situational aspects, which often characterize everyday interactions. As Petsko and Bodenhausen (2020) thoroughly explain:

> *There may be some contexts in which Black women are stereotyped as Black, other contexts in which they are stereotyped as women, and still other contexts in which they are stereotyped not as Black or as women per se, but as Black women specifically. Correspondingly, there may be moments when perceivers exhibit racial bias (but not gender bias) against Black women, moments when perceivers exhibit gender bias (but not racial bias) against Black women, and moments when perceivers exhibit intersectional bias (but not more general forms of racial bias or gender bias) against Black women, instead.*
>
> (p. 5)

Delving into the details of such a debate goes beyond the purposes of this paper. The reason why I am interested in the lens-based account, in fact, is more methodological than theoretical, and has to do with how its proponents describe the unfolding of those mental processes that lead a person to perceive a certain category as more salient for stereotyping than another. According to the authors, this cognitive mechanism is triggered by four factors: perceiver goals, category accessibility, category fit and category distinctiveness (Petsko and Bodenhausen 2020). I will focus on the first three and argue that they might be transformed into constructs useful for imagining new solutions to the problem of misgendering in AGR, suggesting what would happen if we replaced the human perceiver with an artificial intelligence in Petsko's and Bodenhausen's model.

> *Perceiver goals: 'the more a particular social category serves a perceiver's goals, the more likely a perceiver is to use that category over alternatives for construing social targets'.*
>
> (p. 6)

If we were to substitute the human perceiver with an artificial intelligence, then we might wonder: why would the latter be 'interested' in knowing the gender of an individual? This is not a rhetorical question. Most of the scholars who advocate for a prohibition of AGRs—if not of FRT as a whole—argue that in most of the occasions when AGRs are employed, gender does not constitute relevant information, but is treated as such because it is considered a biological datum. However, if we uphold the theory of gender performativity (Butler 1990), we are entitled to wonder why gender should be included within biometric criteria for mass surveillance, and why it would serve an automated system's ultimate goals, such as ensuring security or allowing people in public places. For many, the mere idea of having to answer such questions is reason enough to call for a ban or a very restricted application of AGRs. However, even if someone were able to provide justifications for the use of AGRs that satisfy even the most sceptical, the problem of rendering the data sets employed by these systems more accurate and inclusive would remain. The following points serve to illustrate possible strategies to achieve this purpose.

> *Category accessibility: 'the more accessible a social category, the more likely a perceiver is to use that category over alternative categories for construing targets [...]. The factors that can increase a category's accessibility can be situational or chronic'.*
>
> (Petsko and Bodenhausen 2020, 5)

If we take gender as the category in question and imagine to replace the human perceiver with an artificial intelligence, then we might wonder: is it possible to train an artificial intelligence to recognize the situational aspects of gender? To narrow it down, are there ways to create data sets that reflect the changes in the perceived gender that many people—and especially transgender individuals—undergo over time? The main problem with AGRs, in fact, is that they are built upon the foundations of gender operationalization, and are a byproduct of the idea that gender is an immutable, chronic identity trait. However, what would happen if we implemented a situational versus chronic framework in the constitution of the data sets, for example, using images of the same person over time?

A similar approach was adopted by Kumar et al. (2016). First, they compiled a data set by collecting images from 38 YouTube videos featuring individuals who documented their gender transition; then, the authors incorporated an appearance factor and a transformation factor caused due to HRT for recognition. The results that they achieved (82.36% recognition accuracy) were promising for the times and might stimulate further research in this direction. Specifically, one could design an experimental setup in which, instead of images taken from the internet, photos provided by transgender individuals who consent to their use are employed. These images should depict the faces of the subjects before, during, and after HRT. Since the photos collected before and during hormone therapy 'may contain common appearance features that are stable and consistent even after the therapy' (p. 3), the whole set of images could be used to instruct AGRs to accurately recognize gender even in the presence of continuity between the diachronic dimension—how the face appeared before—and the synchronic dimension of the images—how the face appears now.

Given the resistance that transgender individuals have already shown towards AGRs (see Section 3), one might think that obtaining their consent for image processing could be complicated. However, since the experiment described here would aim

at improving the accuracy of systems struggling to recognize changes in a person's perceived gender, any positive results could be presented as a step towards validating the individual's 'new' gender. Essentially, efforts to refine the gender recognition process could be framed as serving the desire to promote 'gender euphoria', that is, 'the rush of joy a trans person feels when their external appearance aligns with their perceived gender identity' (Piluso 2024, this issue). In this case, the perceived objectivity and superior intelligence of computers (see Hamidi, Scheuerman, and Branham 2018) could serve as mechanisms of affirmation rather than discrimination against the individual.

> *Category fit: 'the more a category appears to "fit" a social context, the more likely a perceiver is to use it for construing social targets. In the social realm [...], fit comes in one of two forms: normative fit and comparative fit'.*
>
> (Petsko and Bodenhausen 2020, 6)

The concept of category fit can be used as a methodological tool good for imagining new ways of constructing inclusive data sets, in which attributes usually associated with one gender (normative fit) are shown to be distributed across all genders (comparative fit). For instance, features such as makeup and long hair are largely treated by AGRs as feminine traits: a comparative fit approach might instruct the technology about the fact that they are not inherently tied to one gender. As Scheuerman, Paul, and Brubaker (2019) argue, 'performative markers like makeup would actually be more relevant to beauty product advertisers than gender classification, because they could then capture all genders who wear makeup' (p. 23). Such a shift in the categorization of certain facial features would be significant not only for transgender people but also for everyone who does not enact gender in traditional ways.

In this context, the notion of gender performativity provides a great contribution to the discussion. It is only within this theoretical framework that one can envision developing a methodology for constructing data sets that challenge those 'gender ontologies [that] always operate within established political contexts as normative injunctions' (Butler 1990, 203). The strategy of feeding the technology with images where specific markers—makeup, hairstyles, jewellery, facial hair and so forth—are comparatively displayed as gender-neutral would serve exactly this purpose. Moreover, the function fulfilled by this approach would be perfectly aligned with the role that Butler assigns to the 'repetitive practices of signification' (see p. 202) that '*displace* the very gender norms that enable the repetition itself' (p. 203): by showing that certain facial features are shared by all genders, comparative fit might make AGRs reliant on less fixed, biased and conventional parameters.

However, the possibility of implementing strategies like that is conditional upon a call for an integration between computer science and human sciences, which, to date, seems to be problematic. As Os Keyes stated in an interview with Vice,[5] in fact, one of the main issues with computer science is that those who study it in college are not expected to take a gender studies class nor an ethics class, and that 'it'd be good if they did'. The moral dilemmas posed by facial and gender recognition technologies discussed thus far, along with the inherent biases of these tools, demonstrate the validity of such an appeal. Should AGRs continue to be developed and implemented, the opportunity to bridge computer science and human sciences would become a genuine necessity. In this scenario, models like the one inspired by the literature on intersectional stereotyping represent just one of the potential points of contact from which we can begin to build this long-overdue dialogue.

5 | Conclusion

The examination of AGRs illuminates significant challenges and ethical considerations inherent in the intersection of technology and gender identity. Across the discussed literature, a pervasive theme emerges: the inadequacy of current AGRs in capturing the complexity and fluidity of gender identity, particularly for transgender, non-binary and gender non-conforming individuals. This critique underscores the urgent need for more inclusive and nuanced approaches to automated facial analysis.

One promising avenue for addressing these shortcomings lies in the adoption of intersectional stereotyping models within AGRs. As delineated in the discourse, intersectionality offers a robust theoretical framework for understanding the multifaceted nature of identity and discrimination. The application of intersectional stereotyping models can facilitate the development of more inclusive and contextually sensitive methodologies for data set construction and algorithmic training. By recognizing the intersectional nature of gender identity and discrimination, researchers can design data sets that better reflect the diversity of human experience and ensure equitable representation across all gender identities. In essence, the discourse on AGRs underscores the importance of adopting inclusive and socially responsible approaches in technological development. By embracing intersectional perspectives and leveraging models, such as intersectional stereotyping, researchers can advance the field of automated facial analysis while promoting greater equity, respect and inclusion for all individuals. As AGRs continue to evolve, it is imperative that they reflect the complexity and diversity of human identity, and intersectional approaches offer a promising pathway forward in achieving this goal.

Data Availability Statement

The data cited in this study are derived from previously published sources, which are all referenced in the bibliography. Detailed information about the data sets can be found in the respective cited articles.

Endnotes

[1] See https://www.cnbc.com/2018/08/08/transgender-uber-driver-suspended-tech-oversight-facial-recognition.html.

[2] See https://www.businessinsider.com/uber-block-deadname-some-transgender-drivers-account-photo-update-app-2021-12?r=US&IR=T.

[3] See https://www.theverge.com/2017/8/22/16180080/transgender-youtubers-ai-facial-recognition-dataset.

[4] See https://www.theverge.com/2021/4/14/22381370/automatic-gender-recognition-sexual-orientation-facial-ai-analysis-ban-campaign.

[5] See https://www.vice.com/en/article/7xnwed/facial-recognition-software-regularly-misgenders-trans-people.

References

Barrett, L. 2020. "Ban Facial Recognition Technologies for Children and for Everyone Else." *Boston University Journal of Science and Technology Law* 26: 223–285.

Beauchamp, T. 2019. *Going Stealth: Transgender Politics and US Surveillance Practices*. Durham: Duke University Press.

Buolamwini, J., and T. Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial GenderClassification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in Proceedings of Machine Learning Research*, 81, 77–91.

Butler, J. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.

Chou, R. S., and J. R. Feagin. 2008. *The Myth of the Model Minority: Asian Americans Facing Racism*. Boulder: Paradigm Publishers.

Clarkson, N. L. 2019. "Incoherent Assemblages: Transgender Conflicts in US Security." *Surveillance & Society* 17, no. 5: 618–630.

Crenshaw, K. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum* 1, no. 8: 139–168.

Currah, P., and T. Mulqueen. 2011. "Securitizing Gender: Identity, Biometrics, and Transgender Bodies at the Airport." *Social Research: An International Quarterly* 78, no. 2: 557–582.

Dauvergne, P. 2022. "Facial Recognition Technology for Policing and Surveillance in the Global South: A Call for Bans." *Third World Quarterly* 43, no. 9: 2325–2335.

De Vries, K. M. 2012. "Intersectional Identities and Conceptions of the Self: The Experience of Transgender People." *Symbolic Interaction* 35, no. 1: 49–67.

Delupi, B. 2023. "Concealment of the Face and New Physiognomies." *Chinese Semiotic Studies* 19, no. 3: 511–521.

Diamond, L. M., and M. Butterworth. 2008. "Questioning Gender and Sexual Identity: Dynamic Links Over Time." *Sex Roles* 59: 365–376.

D'Ignazio, C., and L. F. Klein. 2016. "Feminist Data Visualization." In *Proceedings of the Workshop on Visualization for the Digital Humanities (VIS4DH)*, Baltimore, October 23.

Erentzen, C. A., V. N. Bergstrom, N. Zeng, and A. L. Chasteen. 2023. "The Gendered Nature of Muslim and Christian Stereotypes in the United States." *Group Processes & Intergroup Relations* 26, no. 8: 1726–1749.

Hamidi, F., M. K. Scheuerman, and S. M. Branham. 2018. "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 System (CHI '18)*, 1–13. New York, NY: Association for Computing Machinery.

Han, C. 2009. "Asian Girls Are Prettier: Gendered Presentations as Stigma Management Among Gay Asian Men." *Symbolic Interaction* 32, no. 2: 106–122.

Jones, S. D. 2019. "Ain't I a Woman, Too? Depictions of Toxic Femininity, Transmisogynoir, and Violence on STAR." USF Tampa Graduate Theses and Dissertations.

Keyes, O. 2018. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." In *Proceedings of the ACM on Human-Computer Interaction*, edited by K. Karahalios, A. Monroy-Hernandez, A. Lampinen, and G. Fitzpatrick, 2, 1–22. New York, NY: CSCW, Association for Computing Machinery.

Khader, S. J. 2013. "Intersectionality and the Ethics of Transnational Commercial Surrogacy." *IJFAB: International Journal of Feminist Approaches to Bioethics* 6, no. 1: 68–90.

Kumar, V., R. Raghavendra, A. Namboodiri, and C. Busch. 2016. "Robust Transgender Face Recognition: Approach Based on Appearance and Therapy Factors." In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), Sendai, Japan, 2016*, 1–7.

Petsko, C. D., A. S. Rosette, and G. V. Bodenhausen. 2022. "Through the Looking Glass: A Lens-Based Account of Intersectional Stereotyping." *Journal of Personality and Social Psychology* 123, no. 4: 763–787.

Petsko, C. D., and G. V. Bodenhausen. 2019. "Racial Stereotyping of Gay Men: Can A Minority Sexual Orientation Erase Race?" *Journal of Experimental Social Psychology* 83: 37–54.

Petsko, C. D., and G. V. Bodenhausen. 2020. "Multifarious Person Perception: How Social Perceivers Manage the Complexity of Intersectional Targets." *Social and Personality Psychology Compass* 14, no. 2: e12518.

Piluso, F. 2024. "Dis-Playing Gender: From Ludic Reconfiguration to Utopic Outcomes in Gender-Swap Apps." *Future Humanities* 2, no. 1–2: e8.

Rankin, Y. A., and J. O. Thomas. 2019. "Straighten Up and Fly Right: Rethinking Intersectionality in HCI Research." *Interactions* 26, no. 6: 64–68.

Ryan, C. 2019. "Identity and Discrimination in the Work Place: An Intersectional, Legal History." *FAU Undergraduate Law Journal* 6: 172–180.

Scheuerman, M. K., S. M. Branham, and F. Hamidi. 2018. "Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm With Transgender People." In *Proceedings of the ACM on Human-Computer Interaction*, edited by K. Karahalios, A. Monroy-Hernandez, A. Lampinen, and G. Fitzpatrick, 2, 1–27. New York, NY: CSCW, Association for Computing Machinery.

Scheuerman, M. K., M. Pape, and A. Hanna. 2021. "Auto-Essentialization: Gender in Automated Facial Analysis as Extended Colonial Project." *Big Data & Society* 8, no. 2: 20539517211053712.

Scheuerman, M. K., J. M. Paul, and J. R. Brubaker. 2019. "How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services." In *Proceedings of the ACM on Human-Computer Interaction*, edited by A. Lampinen, D. Gergle, and D. A. Shamma, 3, 1–33. New York, NY: CSCW, Association for Computing Machinery.

Schlesinger, A., W. K. Edwards, and R. E. Grinter. 2017. "Intersectional HCI: Engaging Identity Through Gender, Race, and Class." In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems *(CHI '17)*, 5412–5427. New York, NY: Association for Computing Machinery.

Selinger, E., and W. Hartzog. 2020. "The Inconsentability of Facial Surveillance." *Loyola of Los Angeles Law Review* 66: 33.

Thibault, M., and O. Buruk. 2021. "Transhuman Faces in the Transurban City: Facial Recognition, Identity, Resistance." *Lexia: Rivista di semiotica* 37–38: 179–195.

West, C., and D. H. Zimmerman. 1987. "Doing Gender." *Gender & Society* 1, no. 2: 125–151.