Gender and Sex in the Computer Graphics Research Literature

Ana Dodik[*] Meta Platforms
Silvia Sellán[*] University of Toronto
Theodore Kim Yale University
Amanda Phillips Georgetown University

ABSTRACT

We survey the treatment of sex and gender in the Computer Graphics research literature from an algorithmic fairness perspective. The established practices on the use of gender and sex in our community are scientifically incorrect and constitute a form of algorithmic bias with potential harmful effects. We propose ways of addressing these trends as technical limitations.

1 INTRODUCTION

Sex and gender are referenced in the Computer Graphics literature: a dataset is said to contain images of men and women, user study participants are reported with certain male/female ratios, a body modeling algorithm trains two different gendered models, a voice modification method is said to work on male and female voices, etc.

The scientific consensus around the concepts of sex and gender has evolved in the past decades (see, e.g., [Nature Editorial Board 2018]). As surveyed by Fausto-Sterling [2012], *sex* is not one but a combination of many biological classifications (*chromosomal, hormonal, reproductive, ...*) which cannot be unambiguously assigned in a binary way to as much as one in 50 people [Blackless et al. 2000]. *Gender*, on the other hand, is used to refer to an individual's self-identity [Money and Ehrhardt 1972], their performance of certain acts as shaped by social expectations [Butler 2003] and arbitrary organizational structures that can segregate people in different public bathrooms and even decide who can access education [Lorber 1994]. In these contemporary definitions, gender is non-binary, fluid and culturally-specific. Assuming outdated binary definitions of sex and gender is not just scientifically incorrect, but is also harmful (see [UNHCHR 2015]) to those who conform the least to this artificial binary (e.g, *intersex*, *transgender*, *non-binary* people), to whom we will refer to as *gender non-conforming* individuals.

The treatment of sex and gender in SIGGRAPH Technical Papers still adheres to a traditional binary understanding, excluding intersex, transgender, and gender non-conforming people. Furthermore, it makes research lag behind the needs of industry. The latest character modeller for Unreal Engine [2021] and the Cloud Vision API by Google [2020] have removed references to sex and gender. *Animal Crossing* and *Forza Horizon* completely decouple attributes like body proportions, voice pitch, hairstyle and pronouns.

We will use an algorithmic fairness lens to argue that this binary understanding adds algorithmic biases detrimental to scientific integrity. We will examine the real-world harms caused these biases in how gender non-conforming people interact with our technology. We advocate for a reexamination of our treatment of gender, and show that correcting problematic practices in our community will open the door to new avenues of research.

2 SURVEY

Inspired by Keyes [2018], we survey the technical papers from SIGGRAPH North America and SIGGRAPH Asia since 2015 (see supplement). We observed references to sex and gender, varying in nature from demographic information regarding study participants, to gender-specific algorithms. Whenever gender or sex is used explicitly as a variable, it is always binary. Sex and gender are never given a precise definition, and are used as an proxies for features such as body proportions, facial expressions, or patterns in speech.

From an algorithmic fairness perspective (§2.1), our analysis reveals a worrying status quo in the use of gender and sex as variables. When we mention specific examples that perpetuate these trends, we stress that we are not associating any malicious intent. Rather, we are showing how seemingly neutral, well-established practices can unwittingly perpetuating forms of algorithmic bias.

2.1 Algorithmic Fairness Analysis

Our survey shows that the current use of gender and sex in Computer Graphics is at best ill-defined, and at worst incorrect. We apply the framework of Suresh and Guttag [2021], which categorizes bias according to the stages of a system's lifecycle (see also [Friedman and Nissenbaum 1996; Mehrabi et al. 2021; Olteanu et al. 2019]) We give examples of how different types of bias occur, and show that these are *technical* limitations that impede the development of precise, high-quality, reproducible research.

Representation bias. Portions of populations may be poorly represented by a dataset, e.g., because the sampling procedure did not include people of non-binary genders (*sample selection bias*) or because algorithm performance was not evaluated on groups of *underrepresented* sex or gender. Despite the prevalence of these individuals in the general population, we did not identified a single paper (O3) that explicitly mentioned them as part of datasets (O5) or user study participants (O6). The sampling procedure may have been unintentionally designed to exclude these individuals, or it might be due to measurement bias. We did not identify any work that explicitly analyzed any type of representation bias experienced by gender non-conforming individuals (O1).

Historical bias. Data, despite being abundant and perfectly sampled, may encode existing prejudice. For example, a *gender classifier* (O7) trained on portrait image data collected in an environment where social norms dictate gender expression might learn that "wearing a dress" means woman, and "short hair" means man.

Measurement bias. Bias may be introduced through the selection and measurement of features and target variables. Many works use sex or gender as imprecise *proxies* (O4) for attributes like *commonly co-occuring bodily* or *speech characteristics*, in lieu of less abstract features like hair length or voice pitch. Some works even combined proxies, e.g., conversational agents that use gender for voice pitch *and* culturally acquired speech inflections.

When gender or sex was chosen as a feature or target variable, it was always (O1) through an *inaccurate method of measurement*, such as treating gender as a binary variable that excludes non-binary individuals by design. Alternatively (O7) *incorrect methods of measurement* were used, such as image-based gender classifiers in

[*]Joint First Authors

lieu of self-identification, which can cause gender non-conforming individuals to be misidentified.

Omitted variable bias. A successful feature may correlate with an important feature that has been omitted from the model (see e.g., [Clarke 2005]). *Gender* or *sex* are likely not as discriminative when the result is also conditioned on *hair length*, *hip width* or *mean voice frequency*. When the use of gender or sex was justified because of an assumed improvement in accuracy (O4), we found no attempt to identify if the success was due to omitted variables.

Evaluation bias. These are biases occur during evaluation of an algorithm, such as body modeling works that provide binary segregated parametric models (O3). These are then used to evaluate *other* works with orthogonal contributions, like virtual try-on or motion capture. If our community codifies biased benchmarks, we encourage the development of models that conform to those biases.

Deployment bias. Real-world harm is introduced when graphics models are published or deployed. The exclusive publication of papers with a binary understanding of sex and gender incentivizes researchers (and reviewers) to conform to that definition (O1). This leads to *feedback loops*: if gender non-conforming people are not included in a virtual clothing try-on system, they are less likely to use it, skewing the system's performance data to include them even less. Finally, a system can impose its biases onto user behavior: a trans person may need to change the pitch of their voice in order to not get misgendered by an algorithm, further skewing the data.

2.2 Real world harm

The technical limitations of the reviewed algorithms can lead to real world harms. As Computer Graphics is increasingly applied to other fields, such as geometric data processing in medicine, or for synthetic dataset generation in computer vision, with numerous downstream applications [Behzadi 2021; Brewer 2020; Chen et al. 2021], it is paramount to understand that our algorithms can and will be used in novel ways that can cause unintended harms. The algorithmic fairness literature disambiguates between *representational* and *allocative* harms [Barocas et al. 2019].

Representational harms encompass the perpetuation of stereotypes or cultural norms that subject individuals to denigration. For example, airport body scanners routinely subject gender non-conforming passengers to public humillation [Beauchamp 2019].

Allocative harms are when certain groups are denied access to a resource because of algorithmic bias. For example, a virtual try-on experience based on biased algorithms might exclude the precise people with non-normative bodies who are most in danger in traditional physical changing rooms [Silver 2017].

Finally, ignoring the existence of gender non-conforming individuals in our research (O3) creates an alienating and exclusionary environment for these exact members of our research community, directly contravening SIGGRAPH's goal to be *a model of inclusion, equity, access and diversity for all.*

3 WHERE DO WE GO FROM HERE?

Our analysis reveals that the common use of sex and gender in Computer Graphics can pepper our research with algorithmic bias. Our disambiguated study shows bias throughout the modeling process: algorithmic fairness cannot be an afterthought but must present at all stages of our research. We have focused on sex and gender, but hope our work broadens conversations about algorithmic fairness.

Various real-world constraints may make it unrealistic for specific research groups to mitigate certain sources of bias, but potentially introduced biases should still be acknowledged. For example, none of the surveyed papers included algorithmic fairness metrics (for a summary, see [Mehrabi et al. 2021; Pessach and Shmueli 2020]) in their evaluation, nor did they include a discussion of how their treatment of sex and gender could cause potential harm.

The issues raised by our survey often reveal *scientific* limitations. If a method cannot model a class of humans, or a production system fails for a subsection of the population, these are fundamental *technical* limitations, and should be discussed as such. Gender and sex can have a place in our research. It would be beneficial to report them among demographic statistics of datasets or user study participants (self-reported and non binary, in agreement with the scientific consensus) to safeguard against the "male default" that plagues the sciences. In most cases we observed sex or gender being used as features or targets, they should have been replaced by other, more accurate, variables. Finding these ommitted variables and disaggregating the attributes that have been traditionally crammed into sex and gender constitute important open research problems.

Our proposed break with tradition requires effort, and difficult conversations. These are challenges worth facing if we want scientific advances to produce a fairer, more inclusive future.

REFERENCES

S. Barocas, M. Hardt, and A. Narayanan. 2019. *Fairness and Machine Learning.*

T. Beauchamp. 2019. *Going Stealth: Transgender Politics and U.S. Surveillance Practices.* Duke University Press, Chapel Hill, NC.

Y. Behzadi. 2021. Synthetic data to play a real role in enabling ADAS and autonomy. *Automotive World* (2021).

M. Blackless, A. Charuvastra, A. Derryck, A. Fausto-Sterling, K. Lauzanne, and E. Lee. 2000. How sexually dimorphic are we? *Am. J. Hum. Biol.* 12, 2 (2000).

T. Brewer. 2020. DHS Awards $1 Million to Support Machine Learning Development for Airport Security. *Synthetik Applied Technologies Blog* (2020).

Judith Butler. 2003. Gender trouble. *Continental feminism reader* (2003), 29–56.

R. J. Chen, M. Y. Lu, T. Y. Chen, D. FK Williamson, and F. Mahmood. 2021. Synthetic data in machine learning for medicine and healthcare. *Nat. Biom.* (2021), 1–5.

K. A Clarke. 2005. The phantom menace: Omitted variable bias in econometric research. *Conflict management and peace science* 22, 4 (2005), 341–352.

A. Fausto-Sterling. 2012. *Sex/gender: Biology in a social world.* Routledge.

B. Friedman and H. Nissenbaum. 1996. Bias in Computer Systems. *ACM Trans. Inf. Syst.* 14, 3 (jul 1996), 330–347. https://doi.org/10.1145/230538.230561

Google. 2020. Ethics in Action: Removing Gender Labels from Cloud's Vision API. https://bit.ly/3JkihSH. Online; accessed 20 January 2022.

O. Keyes. 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction* (2018).

J. Lorber. 1994. *Paradoxes of gender.* Yale University Press.

N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6 (2021).

J. Money and A. Ehrhardt. 1972. Man and woman, boy and girl: Differentiation and dimorphism of gender identity from conception to maturity. (1972).

A Nature Editorial Board. 2018. US proposal for defining gender has no basis in science. *Nature* 563, 7729 (11 2018), 5.

A. Olteanu, C. Castillo, F. Diaz, and E. Kıcıman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2 (2019), 13.

D. Pessach and E. Shmueli. 2020. Algorithmic Fairness. (2020). arXiv:2001.09784

L. Silver. 2017. Topshop Refused To Let A Trans Person Into An All-Gender Changing Room. *BuzzFeed News* (2017).

H. Suresh and J. Guttag. 2021. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Equity and Access in Algorithms, Mechanisms, and Optimization* (Oct 2021). https://doi.org/10.1145/3465416.3483305

UNHCHR. 2015. Discrimination and violence against individuals based on their sexual orientation and gender identity. (2015).

Unreal Engine. 2021. Digital Humans | Metahuman Creator.