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. We conclude current trends on the use of gender in our research community are scientifically incorrect and constitute a form of algorithmic bias with harmful effects. We propose ways for correcting these trends and pose novel research questions.

1 INTRODUCTION

References to sex and gender can be found all throughout the Computer Graphics research literature: a dataset is said to contain images of men and women, user study participants are reported to have a certain male/female ratio, 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 greatly 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 sex, hormonal sex, reproductive sex, ...*) 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 [Butler 2003] or arbitrary social organizational structures that segregate people in different public bathrooms and even decide who can access education or participate in public life [Lorber 1994]. By all these contemporary definitions, gender is non-binary, fluid and culturally-specific. Furthermore, assuming outdated binary definitions of sex and gender is not just scientifically incorrect, but can also be shown to be harmful to those who conform the least to this artificial binary [UNHCHR 2015].

Despite this, we observe that the treatment of sex and gender in Computer Graphics research still answers to a traditional binary understanding of it that excludes intersex and many transgender and gender non-conforming people. In what follows, we will use an algorithmic fairness perspective to argue that our community's current use of gender is imprecise, contradictory and detrimental to our scientific integrity. We will examine the harmful real-world consequences of the algorithmic bias introduced by our modeling choices with respect to gender on how gender non-conforming people interact with our technology in their daily lives. We will advocate for reexamining our treatment of gender and show that this will not only correct worrying trends in our community, but also open the door to whole new avenues of research.

2 SURVEY

Inspired by the work of Keyes [2018], we conducted a survey of all technical papers presented at SIGGRAPH North America and SIGGRAPH Asia since 2015 (see *Supplemental Material*). We observed references to gender routinely throughout, varying in nature from demographic information reported about user study participants to gender-specific algorithms. Whenever gender is used explicitly as a variable, it is always as a binary one. Gender is never given a precise definition in all the reviewed Computer Graphics literature, and appears to be used implicitly as a proxy for anything from body proportions to facial expression to voice inflection in speech.

An analysis of the above reveals worrying trends about the current use of gender as a variable in Computer Graphics, both scientifically and ethically. As we mention examples of works that perpetuate these trends, we stress that we do not associate any malicious intent to any. Rather, we wish to show how seemingly neutral, well-established practices in our community can lead to us unwittingly perpetuating forms of algorithmic bias.

3 ALGORITHMIC FAIRNESS ANALYSIS

Our survey shows that the current use of gender in the Computer Graphics literature is at best ill-defined, and at worst incorrect. In this section, we apply the framework of Suresh and Guttag [2021], which categorizes different types of 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 throughout our surveyed work and show that these are *technical* limitations that impede our community's goal of producing precise, high-quality and reproducible research.

Representation bias. A part of a population may be poorly represented by a dataset, for example, because the sampling procedure is biased not to include people of non-binary genders (*sample selection bias*) or because no care is taken to ensure algorithms perform equally well in groups of *underrepresented* sex or gender. Despite the prevalence of these individuals in the general population, we could 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 accidentally 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 underrepresented genders (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 a society 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. We observed that many works use sex or gender as imprecise *proxies* (O4) for attributes like *commonly co-occuring bodily characteristics* or *speech characteristics*, where it is possible that the authors would be better served using other less abstract features (e.g., hair length, or voice

*Joint First Authors

pitch). We even observed works that combine several of these proxies into one; for example, conversational agents that use gender to refer to both voice pitch *and* culturally acquired speech inflections.

Furthermore, when gender was chosen as a feature or target variable, it was always (O1) through an *inaccurate method of measurement* (treating gender as a binary variable, non-binary individuals cannot be captured by design even if they are in the dataset) and often (O7) through an *incorrect method of measurement* (using image-based gender classifiers as opposed to self-identification, transgender and non-binary individuals may be misidentified).

Omitted variable bias. The success of using a certain feature may be overemphasized if it correlates with another important feature that has been omitted from the model (see e.g., [Clarke 2005]). For example, *gender* is likely not as discriminative as a variable when the result is also conditioned on *hair length*, *hip width* or *mean voice frequency.* In survey, where gender's use was justified because of an assumed improvement in model accuracy (O4), we found no effort to indentify if the success was due to omitted variables.

Evaluation bias. For example, we observed works in body modeling that provide binary gender-segregated parametric models (O3). We observed these being used to evaluate *other* works with contributions orthogonal to body modeling, like virtual try-on or motion capture. If the computer graphics community settles on benchmarks with biased data, the development of models that conform to those biases is encouraged.

Deployment bias. The effects of algorithmic bias do not stop at the publication stage; rather, harm is introduced as a consequence of a model being deployed or published in the real world. For example, the publication exclusively of papers with a binary understanding of sex and gender incentivizes researchers (and reviewers) to conform to that definition (O1). This can also lead to *feedback loops*: if, as we have observed, gender non-conforming people are not included in a virtual clothing try-on system, they are less likely to use them, causing the data about the performance of such a system to be skewed to include fewer gender non-conforming people. Finally, a biased system can cause further bias by nudging users to artificially change their behaviour: for example, a trans person might feel the need to change the pitch of their voice in order to not get misgendered by an algorithm, thus also skewing the collected data.

3.1 Real world harm

The discussed technical limitations of the reviewed algorithms are not only academic in nature, but can and do lead to real world harm. As Computer Graphics is being increasingly applied to other fields, such as for processing geometric data 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 potentially causing harm in ways we did not intend.

The algorithmic fairness literature disambiguates between allocative and representational harms [Barocas et al. 2019].

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

Representational harms. These encompas the perpetuation of harmful stereotypes or cultural norms that subject individuals to denegration. For example, it is well documented that airport body scanners (which, despite their specifics being proprietary, one could imagine being trained on synthetic geometric data generated with algorithms from our survey) routinely subject transgender passangers to public humilliation [Beauchamp 2019].

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

4 WHERE DO WE GO FROM HERE?

Silvia: to-do

In general, we have found the gender-related language to be imprecise, which hinders the clarity of the presentation. Oftentimes, we found that *gender* and *sex* are used seemingly interchangebly, and most of the times it is not even clear from context which one the authors intended to use. In fact, we argue that it is impossible to know a-priori if a trained model has picked up on *gender* characteristics, and not on the characteristics of *sex* or *gender expression*, without looking at algorithmic fairness metrics accross different subgroups.

A worrying trend we noticed is that none of the reviewed papers provide an analysis of the algorithmic biases that they potentially introduce. While different real-world constraints might not make it realistic for a research group to successfully mitigate certain sources of bias, the potentially introduced biases should at the very least be acknowledged. For example, none of the reviewed papers included any of the various algorithmic fairness metrics (for a summaries, see [Mehrabi et al. 2021; Pessach and Shmueli 2020]) into their evaluation, nor did they even include a discussion of the potential harm their methods could be causing.

Algorithmic fairness evaluations and discussions are left out of computer graphics papers not because they happen to be difficult or time-consuming, but because they are deemed unnecessary by the people who would at large be either unaffected or positively affected by the introduced biases.

However, we argue that the problems introduced in these methods are not only potentially harmful to under-represented populations, but also often *technically limited ways which are not well researched.* If a method cannot model a class of humans by design, or if a production system fails for a subsection of the population, these are fundamental *technical* limitations. The question is then, why does our community prioritize solving these limitations disproportionatly less then other technical problems?

5 WHERE DO WE GO FROM HERE?

Ana: Some suggestions for this section:

- **Algorithmic fairness should not be an afterthought, but something we plan for from the beginning, as it**

can be introduced in every step of algorithm development.

- **Data collection should account for how much representation is necessary from each marginalized group.**
- **Algorithmic fairness metrics exist and are easy to implement. Use them, and report them in your research.**
- **In our review, we found that gender or sex could have often been replaced by another variable and it would have been more accurate.**
- **We argue that discussions around algorithmic fairness need to become front-and-center within our own community, instead of being relegated to other venues or "future work".**

Ana: This paragraph might be better suited for this section: It bears mentioning that our research community's entrenchment in the traditional gender binary is a rare example of Computer Graphics research lagging behind the needs of our partner industries. *Metahuman*, the latest photorealistic character modeller by Unreal Engine [2021] has no mention of gender; Google [2020] removed all gender references from its Cloud Vision API; video games as diverse as *Animal Crossing: New Horizons*,and *Forza Horizon 5* completely decouple attributes like hairstyle, body proportions, voice pitch and prononouns from one another.

Ana: This might also be better for the conclusion: As Computer Graphics researchers, we must consider our role in shaping whose stories get to be told and who gets to seem themselves represented in the entertainment culture.

We believe the reasons above to be enough to make us reevaluate the role of gender in our community's scientific literature.

For example, the reporting of gender among other demographic information in user study participants and dataset collection subjects answer to a scientifically positive goal (experimental transparency) as well as an ethical one, to safeguard against the "male default" that plagues science and has plagued it since its infancy. However, we found instances in our survey of participants being reported as of "unknown gender", which may indicate that their gender is being assumed post facto by researchers as opposed to self reported, leading to the potential misidentification and exclusion of gender non-conforming individuals or of those from certain ethnicities (see e.g., [Buolamwini and Gebru 2018; Santamaría and Mihaljević 2018]). Therefore, we would argue it is still advisable to include this kind of data, as long as it is self reported by participants who are given a breadth of gender options not restricted to the traditional binary ones.

On the other hand, the scientific and ethical harm caused by gender-segregated algorithms is likely too significant to offset any possible benefits. At the very least, these choices should be justified and their consequences in terms of excluding gender non-conforming individuals should be examined and clearly stated. Eventually, we hope that our field evolves to address these limitations and move beyond the outdated gender binary. We trust that our fellow researchers share our scientific excitement in this new frame of reference and the potential novel research directions it opens; for example:

- What is a complete parametric model for the human body that is decoupled from gender and accurately represents the diverse bodies of all humans, regardless of whether they conform to traditional gender norms?
- How can our research inform or contrast more modern understandings of gender? Can data-based methods be used to evaluate cultural differences in gender presentation?
- How can we evaluate our algorithms for bias towards the gender binary? What tools are needed to obtain or synthesize data that covers more diverse experiences of gender?

We acknowledge that our proposed break with tradition may bring with it effort and difficult conversations, but these are challenges worth facing in the interest of scientific advancement as well as producing a fairer, more inclusive future.

REFERENCES

Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org. http://www.fairmlbook.org.

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

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

Melanie Blackless, Anthony Charuvastra, Amanda Derryck, Anne Fausto-Sterling, Karl Lauzanne, and Ellen Lee. 2000. How sexually dimorphic are we? Review and synthesis. *American Journal of Human Biology: The Official Journal of the Human Biology Association* 12, 2 (2000), 151–166.

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

Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*.

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

Richard J Chen, Ming Y Lu, Tiffany Y Chen, Drew FK Williamson, and Faisal Mahmood. 2021. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering* (2021), 1–5.

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

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

Batya Friedman and Helen 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://diversity.google/story/ethics-in-action-removing-gender-labels-from-clouds-vision-api/. Online; accessed 20 January 2022.

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

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

Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (jul 2021), 35 pages. https://doi.org/10.1145/3457607

John Money and Anke 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.

Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data* 2 (2019), 13.

Dana Pessach and Erez Shmueli. 2020. Algorithmic Fairness. *CoRR* abs/2001.09784 (2020). arXiv:2001.09784 https://arxiv.org/abs/2001.09784

Lucía Santamaría and Helena Mihaljević. 2018. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* (2018).

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

Harini Suresh and John 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. https://www.unrealengine.com/en-US/digital-humans/. Online; accessed 20 January 2022.