Introduction

It really saddens us that the LGBTQ rights groups, HRC and GLAAD, who strived for so many years to protect the rights of the oppressed, are now engaged in a smear campaign against us with a real gusto.

They dismissed our paper as "junk science" based on the opinion of a lawyer and a marketer, who don't have training in science. They spend their donors' money on a PR firm that calls journalists who covered this story, to bully them into including untruthful allegations against the paper. They lie to people that "Stanford has distanced itself from our results." They sent a press release full of counterfactual statements.

They assured people that "Technology cannot identify someone's sexual orientation," but did not explain how they arrived at this conclusion. In the very next paragraph, they contradict themselves by demanding that this technology could be used to "*support a brutal regime's efforts to identify and/or persecute people they believed to be gay.*" Both statements cannot be true at the same time.

Let's be clear: Our paper can be wrong. In fact, despite evidence to the contrary, we hope that it is wrong. But only replication and science can debunk it—not spin doctors.

If our paper is indeed wrong, we sounded a false alarm. In good faith.

But what if our findings are right? Then GLAAD and HRC representatives' knee-jerk dismissal of the scientific findings puts at risk the very people for whom their organizations strive to advocate.

Also see LGBTQ nation's comment on this issue.

Our response to HRC and GLAAD's press release

In their press release, GLAAD and HRC asserted that our research findings are wrong and that the methodology is flawed. Their press release contains statements that are incorrect, misleading, or without merit. Irresponsibly, it was based on poorly researched opinions of non-scientists. Comments asserting that an AI algorithm "*cannot identify someone's sexual orientation*" and dismissing our work as "*junk science*" are not only groundless, but also distract from the main implication of our research: that readily available technology can—and may already—be misused.

We think that this shows premature judgment by the individuals behind this press release. They do a great disservice to the LGBTQ community by dismissing our results outright without properly assessing the science behind it, and hurt the mission of the great organizations that they represent.

Let's be clear: Our findings could be wrong. In fact, despite evidence to the contrary, we *hope* that we are wrong. However, scientific findings can only be debunked by scientific data and replication, not by well-meaning lawyers and communication officers lacking scientific training.

If our findings are wrong, we merely raised a false alarm. However, if our results are correct, GLAAD and HRC representatives' knee-jerk dismissal of the scientific findings puts at risk the very people for whom their organizations strive to advocate.

Our study shows that widely available tools can be used to detect sexual orientation from images of people's faces. The paper can be found here; the author notes are available here. Below, let us address each of the profoundly problematic points raised in their press release:

"The study was not peer reviewed."

This is incorrect, as this study was peer reviewed *and* accepted for publication in the *Journal of Personality and Social Psychology*, the leading academic journal in psychology. The statement on peer review was noted in the first sentence on the first page of the manuscript.

In addition, before it was sent for a formal peer review, the manuscript was reviewed by over a dozen experts in the fields of sexuality, psychology, and artificial intelligence.

"The study did not independently verify crucial information including age and sexual orientation, and took at face value information appearing online."

We put much effort into ascertaining that our data was as valid as possible, and there are no reasons to believe that there are gross inaccuracies. Our approach was no different than in other similar studies. More than a dozen scholars who have reviewed this work did not see any issues in how we handled those variables.

"The study assumed there was no difference between sexual orientation and sexual activity, which is incorrect."

This is incorrect; we did not assume that there was no difference between sexual orientation and sexual activity.

We assumed that there was a correlation between sexual activity and sexual orientation, in that people who said they were looking for partners of the same gender were homosexual. As we discuss in the paper, this is not always correct, and may add noise to the training data, which might result in underestimating the actual classifier accuracy.

"The study assumed there were only two sexual orientations -- gay and straight -- and does not address bisexual individuals."

That is incorrect; we did not assume that there were only two sexual orientations.

We did not make any claims related to how many sexual orientations there are. Our study focuses on just two—straight and gay—which were best represented in our dataset.

As we discuss in the paper, it is possible that some users categorized as straight or gay had a different sexual orientation (e.g., were bisexual). Correcting such errors, however, would likely boost the classifier's accuracy.

"The study did not look at any non-white individuals."

True, but this does not invalidate the findings of the study in any way.

Yes, results based exclusively on white individuals are mostly informative about the risks faced by white individuals. Unfortunately, however, as we discuss in the paper, our results suggest that people of other ethnicities are also at risk.

The reason for focusing on white individuals? Non-white individuals were not represented in sufficiently large numbers in our dataset. We hope that other studies will look at faces of people of other ethnicities in the future.

"The study only looked at out gay men and women who are white, of a certain age, and are on dating sites. It is not surprising that gay people (out, white, similar age) who choose to go on dating sites post photos of themselves with similar expressions and hairstyles (one of the characteristics according to the study)."

This does not in any way invalidate our study. We looked at white men and women in a broad age range (between 18 and 40). The data was collected from a dating website as well as from Facebook profile images.

As we discuss at length in the paper, we were worried that dating website images might be particularly revealing of sexual orientation. However, this was not the case.

First, we tested our classifier on an external sample of Facebook profile images. It achieved comparable accuracy, suggesting that dating website images were not more revealing than Facebook profile images.

Second, we also asked humans to judge the sexual orientation of these same faces, and their accuracy was no better than in past studies where humans judged sexual orientation from carefully standardized images taken in a lab. This suggests that the dating website images were not particularly revealing of sexual orientation—at least, not to humans.

Finally, as mentioned before, the deep neural network used here was specifically trained to focus on fixed facial features that cannot be easily altered, such as structural facial elements. This reduced the risk of the classifier discovering a superficial and not face-related difference between facial images of gay and straight individuals used in this study.

"The research states: "Outside the lab, the accuracy rate would be much lower" (the lab = certain dating sites) and is 10 points less accurate for women.
The study claims to detect gay men from the pool of photos on the dating sites with 81% accuracy. Even if this were true given the aforementioned flaws, it still means that heterosexual men could therefore be identified as gay nearly 20% of the time.
The study reviewed superficial characteristics in the photos of out gay men and women on dating sites such as weight, hairstyle and facial expression."

Yes, this is what we reported. We are not sure what issue GLAAD and HRC see with these facts.

To summarize: The concerns raised by GLAAD and HRC were incorrect, misleading, lack merit, or were clearly addressed in our work. Worryingly, GLAAD and HRC shared with us their commentary before releasing it, but did not give us time to respond and ignored our invitation to discuss their concerns. Consequently, they irresponsibly called into question the validity of the scientific findings that, if correct, should be urgently addressed by technology companies, policymakers, and the public.

We would be delighted to work with GLAAD, HRC, and other organizations representing the LGBTQ community to reach an understanding of our results and their potential implications, and work toward the urgent common goal of protecting the rights and well-being of the LGBTQ community. We would be also delighted to address any criticism that they might have. Any scientific findings can be wrong, but dismissing them and their implications without due consideration could be dangerous and ill-informed.