The Dangers of Detecting Sexual Orientation From Faces

In February 2018 the Journal of Personality and Social Psychology published Michal Kosinski and Yilun Wang's study entitled, 'Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images'. Wang and Kosinski's study claims they trained an algorithm to reliably predict sexual orientation from facial images with a higher level of accuracy compared to humans.

Wang and Kosinski trained their algorithm to recognize patterns from 35,326 images of self-identified gay and straight individuals. The images were obtained from a dating site where users self-identified as seeking homosexual or heterosexual partners. The extracted patterns were entered into a logistic regression to classify sexual orientation using both fixed (e.g. nose shape) and transient (e.g. grooming style) facial features. The algorithm concluded that gay men and women expressed "gender-atypical" facial morphology, expression, and grooming styles.[1] This gender-atypicality manifested in the form of gay men having more feminine faces and lesbian women having more masculine faces compared to those of their respective heterosexual counterparts. Given two photos, one gay and one straight, human judges correctly guessed each person's sexual orientation 61% of the time for men and 54% of the time for women. Wang and Kosinski's algorithm performed considerably better when given the same test. The algorithm correctly identified each person's sexual orientation 91% of the time for men and 83% of the time for women.

Wand and Kosinski limited the demographics of the sample population that drove their predictive modeling and deep neural network which violates section 1.4 of the ACM Code of

[1] Wang, Yilun, and Michal Kosinski. "Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images.," July 2017. https://doi.org/10.31234/osf.io/hv28a.

Ethics (Be fair and take action not to discriminate).[2] The study only looked at images of white men and women. No people of color were used in their modeling. Even if their software benefited the public, which it did not, the machine learning algorithm learned from a biased sample population which causes the software to perform better with white faces and worse with faces of color.

Beyond the sample bias, the study primitively categorized sexual orientation into two binary categories: straight and gay. Wang and Kosinski's narrow approach to sexual orientation demonstrates a lack of understanding within the field of human sexuality and violates section 2.6 (Perform work only in areas of competence) of the ACM Code of Ethics.[3] The LGBTQ community consists of many different layers from bisexuals to gay transgender women, to straight transgender men The web of combinations stretches far beyond the typical gay and straight categories.

"Gaydar" technologies could pose a threat to the privacy and safety of the LGBTQ community. Section 1.2 of the ACM Code of Ethics states that a computing professional should "avoid harm" by reducing negative consequences caused by their work. The section details that "well-intended actions...may lead to harm. When harm is unintended, those responsible are obliged to undo or mitigate the harm as much as possible."[4] While Wang and Kosinski recognized the backlash their study received, they did nothing to mitigate peoples' qualms. They even went as far as saying they intended to use their research to "warn policymakers and LGBTQ communities about the huge risks they're facing if this technology falls in the wrong

[2] "ACM Code of Ethics." Code of Ethics. Accessed December 14, 2019. https://www.acm.org/code-of-ethics.
[3] ''ACM Code of Ethics"
[4] "ACM Code of Ethics"

hands."[5] However, their "warning" served less as a warning and more as a pedestal thrusting this kind of technology into the spotlight ready to be weaponized against the very people Wang and Kosinski claimed to protect. LGBTQ people remain discriminated against within society. Technology like this jeopardizes the safety of an already alienated group. China has already used a derivative of this facial recognition technology to control and suppress the Uighur population, a Muslim ethnic group, within their country.[6] In countries where one's gayness can be a death sentence, technology that supposedly detects gayness causes harm instead of avoiding it.

While Wang and Kosinski's software generally violates many sections of the ACM Code of Ethics, it does follow article 2.7 (Foster public awareness and understanding of computing, related technologies, and their consequences). By developing a Gaydar, Wang and Kosinski show the power of technology. Publishing this study opened many peoples' eyes to the reach of technology. While the study demonstrates this reach in a negative way, it demonstrates technology's capabilities nonetheless.

Of all the stakeholders in this study, the LGBTQ community suffers the most from the research. By further developing technology that outs sexual minorities as deviant, Wang and Kosinski's research compromises the interests of the LGBTQ community. However, that is not to say that no stakeholders exist to support this technology.

As seen in China with its surveillance of the Muslim Uighur minority, governments who violently suppress sexual minorities could be very interested stakeholders in this kind of technology. Through the increasing use of facial recognition to keep track of people, these

---

[5] Mezzofiore, Gianluca. "Everything That's Wrong with That Study Which Used AI to 'Identify Sexual Orientation'." Mashable, September 11, 2017. https://mashable.com/2017/09/11/artificial-intelligence-ai-lgbtq-gay-straight/.

[6] "ACM Code of Ethics"

governments' interests compete with those of the LGBTQ community. But governments interested in this technology do not publicly act on their desires.

To prevent this problematic case study from occurring in the first place, we must prevent people from training machines to detect sexual orientation at all. Sexual orientation's fluidity and ambiguity renders it impossible for machines to predict it 100% of the time. And when it comes to sexual orientation, misidentifying someone can be even more harmful than identifying someone correctly. Misidentifying someone's sexual orientation could easily lead to unwarranted ridicule and oppression.

The introduction of Wang and Kosinski's Gaydar can be mitigated through educating others on sexual orientation. While it's very possible that Wang and Kosinski did this research in order to spread awareness of the harms that this kind of technology can create, they would've been better just leaving it alone. By stirring the pot, Wang and Kosinski did more harm than good.

# Bibliography

"ACM Code of Ethics." Code of Ethics. Accessed December 14, 2019. https://www.acm.org/code-of-ethics.

Mezzofiore, Gianluca. "Everything That's Wrong with That Study Which Used AI to 'Identify Sexual Orientation'." Mashable, September 11, 2017. https://mashable.com/2017/09/11/artificial-intelligence-ai-lgbtq-gay-straight/.

Wang, Yilun, and Michal Kosinski. "Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images.," July 2017. https://doi.org/10.31234/osf.io/hv28a.