Ethics in Academic Publishing

Dhruval Bhatt
May 24th, 2020

The article, "Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation from Facial Images", published in the Journal of Personality and Social Psychology caused a major backlash. This paper's abstract suggests that "findings [of their work] advance our understanding of the origins of sexual orientation and the limits of human perception" and "expose a threat to the privacy and safety of gay men and women" (Wang and Kosinski (WAK) 246). LGBTQ rights groups, HRC and GLAAD, refute the paper by stating that "technology cannot identify someone's sexual orientation" and fault their method as being inaccurate and limited (Anderson). While an academic, Greggor Mattson, criticizes this paper for not utilizing the findings from the fields, such as, sociology, cultural anthropology and LGBTQ studies to understand the diversity in human sexual orientation which limits their work to only detect stereotypical, overt physical characteristics. In addition to critiques of their method and results, this paper poses a major question of the ethics of using such targeted data and exposing a method that could endanger the very population of study. In face of such criticism, it is hard not to wonder, how did this article get published? If I were in the editor's shoes, what would I have done?

An editor's role is to review good scholarly work and publish ideas that are novel, credible, and contribute to the advancement of the field of publication. The promise of big data and increasing allure of using artificial intelligence methods in social science research, makes this paper a good example in this new direction. The paper does follow the established standards of ethics and methods for scientific research: the authors have received approval from the Institutional Review Board (IRB); have explicitly stated the limitations of their subset of data and warn readers of interpretation; and contrary to popular critique have their work peer reviewed. Within their constraints, the authors do demonstrate that even common neural nets can be trained to identify a person's self-identified sexual orientation from non-standardized photographs. This interesting outcome demonstrates the power of technology and possibly help generate hypotheses about the connection of facial features and sexual orientation. This study presents a potential for better understanding of human sexuality using large scale, non-laboratory methodology.

However, I would not have published the article as is and would have required a major revision that uses their results to improve theory and clarify the purpose of the paper. This paper feels like a data science project that is thinly veiled as a social science research through weak references to debatable theory and a superficial consideration of its impact. The authors state that "this work examines whether an intimate psycho–demographic trait, sexual orientation, is displayed on human faces beyond what can be perceived by humans" (WAK, 248). Does it though? To claim that, the authors need strong evidence that humans or machines are in fact, relying facial features to make that judgement not social context and stereotypical clues coded in while labeling. This is not done well. They state that "[p]revious empirical evidence provides mixed support for the gender typicality of facial features of gay men and women" but dismiss it as a result of small samples and human inability (WAK 247). However, could it be human ability causing this inaccuracy? Humans judging may know of arguments that Mattson raises, such as, "gender atypicality may cause gayness as much as gayness causes gender atypical behaviors" and the shifting nature of sexuality and are not "trained" to look for reoccurring clues to attest labels.

The theoretical justification using Physiognomy or prenatal hormone theory (PHT) is ridden with "mays" making themselves questionable yet the method is not setup to verify these assumptions before concluding that "our faces contain more information about sexual orientation than can be perceived or interpreted by the human brain" (WAK 254). As an editor of a social science journal, I would find that unacceptable. While technology enthusiasts would support a new approach, where "[c]orrelation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all", social scientific community should be cognizant that correlation driven AI cannot be used to explain the why (Anderson). If the goal was to just build a sexual orientation classification tool used on dating profiles of white Americans, the authors succeeded but it is missing the rigor of a repeatable, controlled study to expand that into theoretical contribution to social science.

Finally, this paper presents a new question of ethics in technology era. As an editor, it would be hard to penalize academics who followed the existing norms of due diligence, but I would recommend the authors to revisit this paper idea with ethical concerns in the forefront. Informed consent and ensuring that the research does not endanger the subjects is imperative. Given that dating profile pictures are publicly available and people are not personally identifiable,

it seems to meet the minimum requirement, but it could include explicit consent without compromising the results. In addition, the authors are aware of the potential harm of this technology and do warn the readers of its implications but they could mitigate the risk by reconsidering how they present their methodology or who they share the details with. Instead of conflating the social scientific claims and technology warning, it might have been better to focus on the power and problems of this technology and work to help those in perceived danger. Then, it may not be suitable for this journal, but it would be work that would be necessary.

To address the ethical dilemmas of researchers, in Bit by Bit, Salganik urges researchers to see "the IRB is a floor, not a ceiling; put yourself in everyone else's shoes; and think of research ethics as continuous, not discrete". His recommendation to evolve from just adhering to guidelines to embracing principals of ethics would empower researchers to tackle difficult subjects from new data sources. Even though I would not publish this work, I would encourage the researchers to explore the interesting topic but deeply consider theory and ethics in this AI driven research.

Words: 1019

References

"Chapter 6: Ethics." Bit by Bit.

Wang, Y., & Kosinski, M. (2018). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of personality and social psychology, 114(2), 246.

Mattson, G. "AI Can't Tell if You're Gay. . . But it Can Tell if You're a Walking Stereotype."

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired.