Synthetic Gaydar: The Moral Implications of Using AI to Detect Sexuality
Hannah Johnson

## Introduction

In the past ten years there has been a large change in the acknowledgement of non-hetero sexualities in society. Now, a person's sexuality is considered a major part of their personality. But how can you tell a person's sexuality unless you have been told? In pop culture, gaydar refers to the phenomenon of observing visual presentations, auditory cues, and behavior of individuals to determine if that individual is attracted to members of the same sex. The scientific validity of gaydar seems still to be debated, however, current research is pointing towards the possibility that it is real.

But what if instead of people, computers were the ones doing the judging? In this paper, I claim that the use of AI to detect sexuality is pro tanto morally impermissible. I will use a humanity kantianism argument to show that AI gaydar violates the rational autonomy of individuals. I will then consider several objections to this argument and reply to them. Finally, I will discuss other scenarios in which this argument could be applicable.

In 2017, Michal Kosinski and Yilun Wang published research in which AI was more accurate in predicting if someone was gay or straight than the human participants. In this paper, I will refer to this technology as synthetic gaydar. The AI was created by applying neural networks to guess the sexuality based off of images of faces. Many people were upset by this research, but Kosinkski claimed that they conducted this research to spark a discussion and highlight the possible dangers. Others were worried about the validity and accuracy of this research. Regardless of these worries, this technology will very likely be possible in the near future and it will have various applications, some of which could have catastrophic impact. Consider, for example, if a country had legislation in which being gay is illegal, then the AI algorithm could be run against all the images of citizens without their consent. This would essentially "out" people in an environment where it is not safe to be open with their sexuality. These people could then be imprisoned, targeted, or even killed — negative consequences that they did not foresee by giving their picture to the government.

## The Argument

I claim that the use of synthetic gaydar is morally impermissible, and therefore this type of technology should not be developed further and should not be used. The following is my argument in premise conclusion form:

(P1) An action is morally impermissible if it violates the rational autonomy of one or more individuals.

(P2) Synthetic Gaydar violates the rational autonomy of those whose faces it is used on.

(C1) Synthetic Gaydar is morally impermissible.

P1 is a statement of Humanity Kantianism, which states that an action is permissible if it respects the rationality and autonomy of moral patients. This means that moral patients have the ability to make informed decisions that are not coerced. Now why might Kantianism be a good moral theory to apply to this situation? Sexuality is a very intimate part of someone's identity, and therefore by attacking someone's sexuality you are deeply attacking their person. Since sexuality is so closely tied with the moral patient, it is logical to consider if synthetic gaydar respects their autonomy.

P2 is the main premise that must be justified. P2 states that synthetic gaydar violates the rational autonomy of the people to whose faces it is applied. To do this, we must show that participants can rationally consent to users applying synthetic gaydar on them. To rationally consent, participants must know the users' plan and the maxim behind that plan. The result of applying synthetic gaydar is that the user of the AI will know the sexuality of the participant, but the important part is why they want this information. The following are a few examples of maxims for why the user can be using synthetic gaydar. Maybe the user wants to know the sexuality of the participant because they want to confirm their suspicions. Maybe the user wants to know the sexuality of the participant because they do not trust the participant to reveal it truthfully. It is even possible the user wants to ask out the participant, but does not know the participant's sexuality, and does not want to embarrass themselves by asking. However, the user could also want the sexuality of the participant to report it to local authorities because they think being gay is wrong. The point is there are many different possible maxims for the application of this technology. Some are neutral, some may harm the participant, and some may be beneficial to the participant. Unless the user truthfully tells the participant that they would like to use synthetic gaydar on them and the maxim behind it, the participant cannot rationally consent to it because they do not know the full story. Thus in the case when the user does not explain their maxim, synthetic gaydar is impermissible.

Now what if the user asks the participant permission and fully explains what they will do with the information? Notice that all the maxim examples had a common theme of the participant not knowing the technology was being applied to them and by who. By

directly asking to apply the technology, the user might just as well ask the participant their sexuality. In addition, there is no way a participant could rationally consent to the AI algorithm. This is because AI algorithms are very complex and the inner workings are often hidden from humans. Thus there is no way the participants could know all the information about the AI. Therefore, it is impossible for the participant to consent rationally to the AI.

Now we will discuss the conclusion and how it follows from P1 and P2. I conclude that synthetic gaydar is morally impermissible. This is because humanity kantianism states that an action is morally impermissible if it violates the rational autonomy of an individual. I have given reasons as to why synthetic gadar violates the rational autonomy of participants whose faces are put through the algorithm. And thus synthetic gaydar is impermissible.

The Objections

The first objection that we will discuss is what if an individual wanted to use the gaydar on themselves. One might want to do this if they wanted to find out if they were gay. Since they are voluntarily doing it to themselves, it would not violate their rational autonomy.

However, this is not the case. Recall that humans would not be able to understand all the complex inner workings of a neural network. We can separate the AI algorithm from the user and consider the algorithm as a moral agent. The individual cannot conceptualize all the hidden work the AI is doing, and therefore the individual could not rationally consent to the AI. In addition, it is unlikely that the use of this technology could be enforced to only be for personal use. Finally, there is unique value in the journey that individuals go through to discover their sexuality. By outsourcing this discovery to AI, the process of discovering sexuality would diminish in worth for the person.

The second objection that we will discuss is that individuals' faces are a public commodity. People walk around with their faces exposed, and by doing so they are implicitly consenting to their face being looked at. If they did not want their face to be observed, they could hide it. Thus it makes sense for faces to be considered as public, therefore rendering rational consent to be useless in this situation. In addition, people judge others by their appearance all the time, including judging their sexuality. So why is it morally wrong for AI to do the same?

The first response to this is that humans' gaydar on average is less accurate than the synthetic gaydar. SInce there is a level of uncertainty in that judgment, personal

information is not being revealed as it would be with AI. In addition, people who are queer tend to have a more fine-tuned gadar. This is relevant because that person is more likely to be safe and understanding, so even if they have more certainty of a person's sexuality, they do not introduce more harm in that person's life. Finally, when people walk around with their face exposed, they do not expect that an AI will be used to guess their sexuality. The implicit consent that is given is for others to learn surface level information, like that person's eye color. The person does not give consent to learn deeper information like sexuality or fears. Thus the implicit consent for people to gaze on their face does not extend to synthetic gaydar.

Other Applications

The technology that Kosinski and Wang created can be modified and applied to other aspects of a person's identity, such as psychological traits, IQ, and a person's disposition to commit crime. In addition, we already have technology that is able to determine information about a person without being directly told. For example, TikTok is well-known for its algorithm at suggesting videos to users that they like. Thus, TikTok is already equipped to determine if a person is queer, as queer people are more likely to watch queer-related content. Many other technologies and services that we use today have similar abilities. Therefore, I think it is important to discuss the morality of gaining secondhand information through technology.

Conclusion

In this essay, I argue that using AI to determine or predict sexuality is morally impermissible. I motivate this through a humanity kantianism argument, claiming that the synthetic gaydar violates the autonomy of individuals. There are two objections which I consider. The first is that an individual could use the technology on themself, and the second is that faces are a public commodity. Both of these objections fail because rational autonomy is still not respected in either of these situations. Finally, I discuss some related moral situations that this argument might be applicable towards, which are other technologies already doing the same work as a synthetic gaydar without being as direct as the technology created by Kosinki and Wang.

References

Lehmiller, Justin." *The Science of "Gaydar": How Well Can We Detect Other People's Sexual Orientation?*",
https://kinseyinstitute.org/news-events/news/2017-12-18-gaydar.php.

Levin, Sam. "New AI Can Work out Whether You're Gay or Straight from a Photograph." *The Guardian*, Guardian News and Media, 7 Sept. 2017, https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph.

Lewis, Paul. "'I Was Shocked It Was so Easy': Meet the Professor Who Says Facial Recognition Can Tell If You're Gay." *The Guardian*, Guardian News and Media, 7 July 2018, https://www.theguardian.com/technology/2018/jul/07/artificial-intelligence-can-tell-your-sexuality-politics-surveillance-paul-lewis.

Rachels, James. "Why Privacy Is Important." *Philosophy & Public Affairs*, vol. 4, no. 4, Wiley, 1975, pp. 323–33, http://www.jstor.org/stable/2265077.

Quach, Katyanna. "The Infamous AI Gaydar Study Was Repeated – and, No, Code Can't Tell If You're Straight or Not Just from Your Face." *The Register® - Biting the Hand That Feeds IT*, The Register, 12 Mar. 2019, https://www.theregister.com/2019/03/05/ai_gaydar/.