# A CASE FOR REGULATION: IMPACTS OF ARTIFICIAL INTELLIGENCE ON THE LGBTQIA+ COMMUNITY

Isadora Valadares Assunção & Bernardo de Souza Dantas Fico

## INTRODUCTION

The LGBTQIA+ community has long been marginalized from society. Granted, there have been positive movements since the Stonewall riots of 1969, but equality is far from being reached and progress is not uniform around the globe. According to the International Lesbian and Gay Association (ILGA), 29 nations allow same-sex marriage, and 58 nations confer protection against sexual orientation-based discrimination. However, 67 nations still criminalize same-sex sexual activity, with six of them prescribing the death penalty.[1]

In this scenario, any large-scale unsupervised processing of information regarding the LGBTQIA+ population can be daunting. Recent applications of artificial intelligence (AI) illustrate how this risk translates into the real world, in different geographic locations. This paper will discuss: (1) the efficiency of LGBTQIA+ recognition, (2) ethical implications of LGBTQIA+ recognition, (3) risks associated with LGBTQIA+ identification, (4) potential benefits, and (5) proposed solutions.

## I. THE EFFICIENCY OF LGBTQIA+ RECOGNITION

"Gaydar" is a popular term used to indicate the ability to identify whether someone is from the LGBTQIA+ community, usually based on appearance or mannerisms. In 2017, Stanford University Professor Michal Kosinski published a paper claiming to have created a deep neural network nicknamed “AI Gaydar.”[2] He purported the AI Gaydar was trained with 35,326 dating website images and could recognise the sexual orientation of men and women with 81% and 74% accuracy, respectively; with five photos of the person to “identify”, the accuracy would rise to 91% for men and 83% for women. The AI Gaydar outperformed human judges who had 61% and 54% accuracy for men and women, respectively. The authors argued the sexual orientation detection was due to facial differences arising from hormone exposure, alongside environmental, developmental and genetic factors. Supposedly, homosexual men had less facial hair

[1] Lucas Mendos et al., *State-Sponsored Homophobia 2020: Global Legislation Overview Update* (ILGA World) (2020), https://perma.cc/FV8X-J4NY (last visited Jan 9, 2022).

[2] Yilun Wang & Michal Kosinski, *Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.*, 114 Journal of Personality and Social Psychology, 246-257 (2018).

and "feminine" faces, while homosexual women had "masculine" faces.

Scientists and the public criticized the paper for its methodology. First, the algorithm had been trained on white people, thereby underrepresenting ethnic groups.[3] This intersects race-related problems, including the fact women with darker faces suffer facial recognition error rates between 20.8% and 34.7%.[4] Second, Todorov *et al.* suggest that the AI Gaydar identified multiple features, including makeup, glasses, selfie angle, sun exposure, and other social characteristics, instead of solely analysing facial structure.[5]

Incrementally, the Automated Gender Recognition (AGR) algorithms have inefficient performance on LGBTQIA+ people. A University of Colorado Boulder study used 2,450 photos with self-identified gender to analyse the accuracy of commercial AGRs. The overall gender recognition accuracy was 98.3% for women and 97.6% for men.[6] Nonetheless, the LGBTQIA+ community had significantly lower accuracy rates, with particularly disproportionate misidentification for trans individuals. AGRs had an accuracy of 87.3% for transwomen, and of only 70.5% for transmen.[7] Giggle, a social relationship app, selects its users by applying similar technology,[8] and Uber verifies user identity the same way.[9] In both cases, transgender people have been—and continue to be—misidentified due to gender stereotypes fed into the AI systems, either via code or via training set biases.

Therefore, even before analyzing ethical issues and risks, using appearance-based AI systems to recognize sexual orientation or gender identity may have an inherent flaw: it has greater inefficiency when applied to the LGBTQIA+ community.

Another means of identifying LGBTQIA+ individuals is by using proxies. In the online world, certain topics can raise a

[3] *Id.*

[4] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, in* Conference on fairness, accountability and transparency, 77-91.

[5] Alexander Todorov et al., *Do algorithms reveal sexual orientation or just expose our stereotypes?* Medium, https://perma.cc/ZXZ5-3Q4B (last visited Jan 9, 2022).

[6] Morgan Scheuerman et al., *How Computers See Gender: An Evaluation of Gender Classification in* Commercial Facial Analysis Services, 1-33.

[7] *Id.*

[8] Brianna Holt, *Women-only networking app Giggle under fire for transphobia,* Daily Dot, 2020, https://perma.cc/MVX6-MGGY (last visited Jan 9, 2022).

[9] Steven Melendez, *Uber driver troubles raise concerns about transgender face recognition*, Fast Company, 2018, https://perma.cc/WT4E-88Y8 (last visited Jan 9, 2022).

flag to indicate the likelihood that a user is LGBTQIA+. With broad data analysis, one could theoretically identify LGBTQIA+ people even if they are "in the closet." In a study, Kosinski used Facebook likes to predict men's sexual orientation, obtaining 88% accuracy. The study is based on numerous assumptions that correlate certain topics with a given sexual orientation. Nonetheless, it shows how online tracking can unveil and predict information, especially considering the vast amount of data available.[10] Leaving aside whether these kinds of algorithms work correctly, one must ask whether it is ethical to seek identifying LGBTQIA+ individuals, and what the risks are associated with this practice.

## II. Ethical Implications of LGBTQIA+ Recognition

Revealing or concealing personal information during social interactions is essential to individual self-determination.[11] People’s autonomy and self-determination are threatened significantly considering the possibility of LGBTQIA+-related data or facial recognition systems “outing” individuals—making their previously undisclosed sexual orientation or gender identity unintendedly known. From an ethical perspective, revealing this type of private information without consent, and in the absence of strong reasons to do so, is evidently unethical. Moreover, sexual orientation is sensitive personal information under certain laws, like the EU General Data Protection Regulation.[12] Gender and gender identity, however, are usually overlooked categories.

In theory, AGRs could be used to control access to gendered spaces (e.g., bathrooms). This implies AI systems being empowered to enforce a person's gender identity according to the algorithm’s own perception, hindering LGBTQIA+ self-constructed identities and self-determination. According to Keyes, 94.8% of AGRs considered gender binary, 72.4% considered gender immutable, and 60.3% considered gender as purely biological.[13]

Even if the use of sexual orientation and gender identity were not an issue *per se*, contemporary applications of said

[10] Michal Kosinki et al., *Private Traits and Attributes Are Predictable from Digital Records of Human Behavior*, *in* Proceedings of the national academy of sciences, 5802-5805.

[11] Erving Goffman, The Presentation of Self in Everyday Life (London: Harmondsworth) (1978).

[12] Regulation 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119/1) 1 (EU), [hereinafter GDPR].

[13] Os Keyes. *The misgendering machines: Trans/HCI implications of automatic gender recognition*, *in* Proceedings of the ACM on human-computer interaction, 1-22.

technology are questionable. In the current state of surveillance capitalism, people are objectified and labeled with distinctive characteristics that can aid advertising (ad) targeting and data manipulation. These labels are supported by reinforcements and nudges imputed in online interactions.[14] Inferring a person's sexual orientation or gender identity for commercial motives could still imply an unethical use of the information, even without publicly "outing" these individuals. Nonetheless, this same technology financially supports LGBTQIA+ oriented initiatives, like the Trevor Project, mentioned below, for assisting youth to avoid suicide.

## III. Risks Associated with LGBTQIA+ Identification

While the ethical implications of most AI systems will be similar, evolving technology applied in multiple locations will present different risks. In some cases, these risks may overlap or reinforce each other.

### a. Discriminatory risks

As noted above, algorithmic bias may derive from either biased coding or from the use of historically biased datasets, which perpetuate discrimination patterns present in that dataset.[15] Consequently, historical biases can lead to discrimination via the rationale that software engineers apply for creating an algorithm or via the information chosen as a basis for the AI systems. Thus, considered LGBTQIA+ social marginalization, AI systems predicting sexual orientation and gender identity may imply discrimination.

Mindful of the risks of sexual orientation discrimination, Facebook has announced to ban ad targeting based on sexual orientation starting in January 2022.[16] However, gender targeting will continue to be possible, because studies have shown that Facebook's algorithms discriminate based on gender stereotypes. For example, in 2020 Algorithm Watch has demonstrated that Facebook's targeting tool would direct truck ads to 4,864 men and 386 women, while childcare jobs were shown to 6,456 women and 258 men.[17] These issues are relevant to Western societies, where the solution rests in fair development of AI and technology in general, but other cultures face different

[14] Shoshana Zuboff, *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*, Public Affairs (2019).

[15] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev, 671 (2016).

[16] Dan Milmo, *Facebook bans ads targeting race, sexual orientation and religion*, The Guardian, 202, https://perma.cc/7XBJ-PG57 ( (last visited Jan 9, 2022).

[17] Nicolas Kayser-Bril, *Automated discrimination: Facebook uses gross stereotypes to optimize ad delivery*, Algorithm Watch, 2020, https://perma.cc/634F-Q47U (last visited Jan 9, 2022).

problems. Where homosexuality is criminalized, the risk of discrimination can easily turn into a physical safety concern.

### b. Employment risks

Even if one is not legally allowed to discriminate against LGBTQIA+ individuals, studies have shown that human resources staff give lower scores for queer people than for their counterparts.[18] Therefore, any member of the LGBTQIA+ community recognized as such is likely to face higher entrance thresholds, and higher evaluation standards than their peers; that is despite studies that show a more diverse workforce implies better performance for businesses.[19]

From a recruitment perspective, applying an AI system assisting the hiring process may reproduce these biases, either consciously or not. Because AI algorithms are capable of recognizing patterns, it may find similarities between LGBTQIA+ employees and conclude people following that particular pattern need to meet more stringent requirements. While the algorithms themselves may not be biased, the sole fact the previous hiring processes were biased themselves can lead to a continuum of discrimination.

### c. Psychological risks

LGBTQIA+ recognition systems can lead to emotional distress and cause psychological damage. Regardless of sexual orientation and gender identity, people may fear both being correctly recognised or misidentified as part of the LGBTQIA+ community. The unwanted revealing of personal features, even if incorrectly, or the erasing of their identity can cause emotional harm. One may fear being recognised as LGBTQIA+ in an unwelcoming social context, creating a significant chilling effect that makes users police themselves and avoid sharing their identity. Another may fear not being recognized as LGBTQIA+ despite being part of the community; bisexual individuals have repeatedly reported being ostracized by parts of the community, particularly when engaged in heterosexual relationships.

Moreover, knowledge about other risks listed can create in itself significant fear, including: AGR controlled bathrooms that may misidentify transgender individuals, online applications that may leak sexual orientation information, and other issues.

### d. Globalization risks

Once a technology is developed, it may (and it probably will) be used globally. In certain parts of the world, this rapid

[18] Victoria LeCroy & Joshua Rodefer, *The influence of job candidate LGBT association on hiring decisions*, 21 North American Journal of Psychology, 373-386 (2019).

[19] Swinton Hudson, Jr., *Diversity in the Workforce*, 3 Journal of Education and Human Development, 73-82 (2014).

spread may cause technologies that are controversial in some contexts (e.g., sexual orientation facial recognition, *supra*) to become a matter of physical safety. Because of this, technological development needs to bear privacy-oriented designs, ensuring any issues arising from misuse of the technology or from data breaches cause the minimum possible harm. Particularly, any tool that can be used globally needs to consider how different areas will interact with the new technology.

In 2018, the dating network for queer men, Grindr, had a data leak where the HIV status and GPS location of its users were shared without consent.[20] Following that, NBC News reported security breaches exposing users' locations to malicious third parties, even where the user had opted out from sharing their precise location with Grindr in the first place.[21] In 2020, the Norwegian Consumer Council uncovered that Grindr had continued its undue sharing of information.[22] In late 2021, the Norwegian Data Protection Authority fined Grindr approximately 7.2 million euros for violating the GDPR.[23]

Grindr's conduct is questionable, and the consequences of sharing HIV status and GPS location can be harmful anywhere. It is even more so in those (numerous) countries that still criminalizing homosexuality, where the LGBTQIA+ community is exposed to more severe risks. In those countries, LGBTQIA+ people need to conceal their sexual orientation and gender identity to avoid penal sanctions and ostracization. Information leaked from Grindr can be used to identify, track, persecute and prosecute them. For example, in Egypt, where homosexuality is *de facto* criminalized,[24] police used Grindr to

[20] Camila Domonoske, Scott Neuman, *Grindr admits it shared HIV status of users*, The Two-Way, 2018, https://perma.cc/8477-DVZJ (last visited Jan 9, 2022).

[21] Brian Latimer, *Grindr security flaw exposes users' location data*, NBC News, 2018, https://perma.cc/Z2ZN-YNPY (last visited Jan 9, 2022)

[22] Forbrukerrådet, *Out of Control. How consumers are exploited by the online advertising industry* (2020), 2020-01-14 Out Of Control final-v2 (forbrukerradet.no) (last visited Jan 10 2022).

[23] Jan Olsen, Kelvin Chan, *Norway fines dating app Grindr €7.16M over privacy breach*, ABC News, 2021, https://perma.cc/CYQ4-2CSH (last visited Jan 10, 2022).

[24] Lucas Mendos et al., *State-Sponsored Homophobia 2020: Global Legislation Overview Update* (ILGA World) (2020), https://perma.cc/FV8X-J4NY (last visited Jan 9, 2022).

pinpoint the location of users and arrest them.[25] Paul Ohm names these sets of information "databases of ruin."[26]

Given the societal costs of businesses that use LGBTQIA+ recognition systems, their sharing policies, and security, one must assess whether the costs outweigh the benefits.

## IV. Potential Benefits

Although there is a wide range of societal costs involving LGBTQIA+ recognition systems, there can be some positive technical applications for it. Codes are universal tools and might be used for very different purposes: they can be used to either discriminate against LGBTQIA+ people or provide them with better services. The difference lays on the design and application of these tools, and the ethics thereof, not on the tool being developed.

### a. Community development

According to Kirsty Hughes, privacy is the employment of subjective and normative barriers to prevent someone from accessing the individual.[27] As social interactions are based on a balance between the desire to connect with others and to distance oneself from others, privacy fosters social interactions as it allows people to develop their individuality and autonomy to experience different states such as solitude, intimacy, anonymity and reserve.[28] LGBTQIA+ recognition systems can be used to create safe spaces for the community. Data analysis may identify people from different parts of the world with similar tastes and suggest online connections. The Internet has been a means for LGBTQIA+ people to connect, get information, and seek community support; well-intended data analysis can grow the positive impact of such uses.

### b. Healthcare

Certain disease patterns arise in different communities, and AI systems may be able to detect it with greater precision. Once a pattern is identified, studies have shown AI systems can even predict new cases of a certain disease and assist the health system provide better treatment to the population.[29] AI solutions

[25] Gemma Fox, *Egypt police 'using dating apps' to find and imprison LGBT+ people*, Independent, 2020, https://perma.cc/2S3T-TP3L (last visited Jan 10, 2022).

[26] Paul Ohm, *Broken promises of privacy: Responding to the surprising failure of anonymization,* 57 UCLA L. Rev., 1701 (2009).

[27] Kirsty Hughes, *A behavioural understanding of privacy and its implications for privacy law*, 75 The Modern Law Review, 806-836 (2012).

[28] *Id.*

[29] Julia Marcus et al., *Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study*, 6 The Lancet HIV, e688-e695, (2019).

could also be used to identify people with risk of suicide to prioritise treatment and access to assistance initiatives, such as The Trevor Project.[30] The information generated via AI, however, needs to be met with critical thinking from the ones applying it. Uncritical application of the data can lead to bias confirmation, instead of better health planning.

## V. PROPOSED SOLUTIONS

When high-complexity issues arise, society needs to regulate them to allow for ordered development. The use of untested medicine has caused intoxication and deaths in the United States.[31] As a consequence, the US Congress approved the US Food, Drug and Cosmetic Act of 1938, expanding the US Food and Drug Administration's (FDA) powers to include pre-market reviews. With AI, similar steps need to be taken.

LGBTQIA+ recognition systems have been increasingly causing different harms to people, including to their physical well-being. AI systems' complexity and inherent opacity make it harder to gauge potential risks and harms. Thus, it is of the utmost importance that a central and independent national supervisory authority, like the US FDA[32] or the UK Medicines and Healthcare products Regulatory Agency (MHRA), is established for algorithmic regulation, including that of LGBTQIA+ recognition systems.

This authority needs to have the power to classify algorithms according to their complexity and potential risks, applying different standards of obligations for each type: low-risk, medium-risk, and high-risk, based on technical and contextual risk management assessments, such as those proposed in Art. 9 of the Draft EU Artificial Intelligence Act.[33]

The standards would vary according to each risk category: (i) low-risk algorithms could be pre-authorized to roll out, provided they comply with the explainability standards and existing data protection legislation; (ii) medium-risk algorithms ought to be assessed *ex-ante* on a case-by-case basis to determine whether developers took reasonable steps to prevent discrimination, and the algorithms can be launched in the market, or if additional assurances are needed; and (iii) high-risk algorithms may lead to unacceptable risk, and should not be

[30] Anagha Srikanth, *Google, the Trevor Project are using AI to help LGBTQ+ teens at risk of suicide*, Changing America, 2021, https://perma.cc/RD8H-5NPJ (last visited Jan 10, 2022).

[31] Kristin Jarrell, *Regulatory history: Elixir sulfanilamide*, 16 Journal of GXP Compliance, 12-15, (2012).

[32] Andrew Tutt, *An FDA for algorithms*, 69 ADMIN. L. REV., 83, (2017).

[33] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final (April 21, 2021).

launched without extensive scrutiny to ensure the observance of non-discrimination obligations and human dignity. If developers are unable to meet the necessary safeguards, these algorithms should not be launched into the market.

This case-by-case analysis ought to be a holistic process of weighing several factors, such as the intended purpose of the system, its technical shortcomings or strengths, whether vulnerable groups are involved, the results of the system in its testing phase, the possible societal benefits and costs of implementing the algorithm, and whether the product seeks to resolve an urgent situation. Particularly regarding the LGBTQIA+ community, assessments need to be contextual, extrapolating the strictly technical aspects of the algorithm. Gender identity and sexual orientation are social concepts that evolve through time and change across locations. This multifaceted regime can support innovation while giving due consideration to ethical issues, potential risks, and disproportionate effects vulnerable groups may suffer from the development of new technology.

Some will inevitably argue that regulating AI will slow down technological progress, and this proposition is likely true. Nonetheless, while progress is desirable, it is not the only relevant factor when deciding whether, and how, AI will be regulated. The safety of individuals during the development of new technologies ought to be considered. Additionally, when taking into consideration the necessity of developing a particular product, the regulation could adopt different protocols for urgent cases, mirroring what national health authorities across the world did with the timeline of vaccine testing when the COVID-19 pandemic was declared.[34] Regulating the development AI can be the middle ground that allows for innovation without losing sight of the protection of vulnerable groups.

## CONCLUSION

LGBTQIA+ recognition systems have been used increasingly in research and the market. Despite the different efficiency rates of facial recognition and data inference systems, both have been implemented without due consideration of the ethics and risks involved, which can include behavioral manipulation, discrimination, marginalization, physical, emotional and relationship harms, and others. Thus, regulation

[34] U.S. Food and Drug Administration, *Emergency Use Authorization for Vaccines to Prevent COVID-19: Guidance for Industry* (2021), https://www.fda.gov/media/142749/download (last visited Jan 10 2022); Medicines and Healthcare Products Regulatory Agency, *Conditions for authorisation for emergency supply under Regulation 174 for COVID-19 Vaccine AstraZeneca* (amended Sept. 9, 2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1016207/AZ_Conditions__-_9SEPT21.pdf (last visited Jan 10 2022).

is necessary to ensure an ethical processing of information, and to prevent risks from materializing. Considering the large-scale effects of AI and its rapid capacity of spreading around the world, algorithms should be subjected to an *ex-ante* scrutiny by an independent national authority. The pace of innovation may decrease, but technology will continue to develop, and the byproducts thereon will respect the particularities of different groups affected by the novel AI systems.