



HAL
open science

Queer Identities and Machine Learning

Mélanie Gornet, Tiphaine Viard

► **To cite this version:**

| Mélanie Gornet, Tiphaine Viard. Queer Identities and Machine Learning. 2023. <hal-04763352>

HAL Id: hal-04763352

<https://hal.science/hal-04763352v1>

Preprint submitted on 1 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

QUEER IDENTITIES AND MACHINE LEARNING

Mélanie Gornet

i³, SES,
Télécom Paris
Institut Polytechnique de Paris
melanie.gornet@telecom-paris.fr

Tiphaine Viard

i³, SES,
Télécom Paris
Institut Polytechnique de Paris
tiphaine.viard@telecom-paris.fr

1 Introduction

The marginalization of queer¹ individuals by society is not a new problem. However, technological advances, particularly in artificial intelligence and machine learning (AI/ML), have accelerated this trend. For a few years now, initiatives have been developed to raise awareness of queer issues in AI/ML and foster a community of queer researchers [11]. Yet the problems associated with machine learning for queer identities remain largely ignored.

When systems are developed for the masses, divergent identities that do not fit the model are excluded from the technology and therefore from the society that revolves around it. This is even truer for machine learning systems, based on data that, for queer people, does not always exist; which learn patterns that, statistically, only work for larger groups; and whose sole purpose is to fit individuals into boxes, assigning them labels that often do not correspond to queer identities. For the most part, these problems are inherent to the nature of machine learning systems, which clashes with that of queer identities. We reach here the difference in ontological categories between computer science and social science – we have to deal with two levels of theoretical construction: building constructs from individuals’ constructs, whom give interpretive meaning to their word.

In this work, we summarize the challenges faced by machine learning practitioners in including considerations for queer identities in system design. We also provides insights into why the subject is not being researched more, and how the field of machine learning might adapt to meet these challenges.

2 What are the problems

Fitting into labels One of the main goals of *supervised* machine learning is to predict a certain class for a certain input point. To that end, the system is trained on a variety of data for which their respective labels are known. Labels, whether for determining different classes or assigning them to data instances, are decided by humans. They are sometimes called “ground truth” or “gold standard” to reflect their presumed impartiality. Yet the “right answer” in a given context is always decided by a subjective human being [14]. Therefore we often see classifications that do not fit queer individuals, either because there is not a correct box – such as the classification between male and female, dismissing all non-binary identities [12]; or because the wrong label is assigned – as with trans people, often misgendered by their human counterparts.

Plurality of queerness The difficulties in labels also comes from the multiplicity and dynamic nature of queer identities. In the words of Mary L. Gray, “there is no one way of being queer, [...] our identity isn’t a staple”, while ML systems try to emulate well defined and static decisions [24]. Therefore, even if there is a label that corresponds to the individual, such as “queer”, the heterogeneity of the data *within* this label will pose problems for a ML system, with the risk of a Simpson’s paradox [4] appearing if the label is too large.

Underrepresentation ML systems perform poorly on small groups, such as dark-skinned women [5] and non-binary people [19], among others: intersectional experiments show that the smaller the group, the lower the performance [5]. It is a question of both data quantity and statistical methods. Indeed, ML systems adopt a one-size-fits-all approach that can pose problems for learning and evaluating the performance of underrepresented groups [22].

¹We define “queer” as individuals who have a non-normative sexual orientation, gender identity and/or sexual characteristics. We aim to encompass all LGBTQ+ individuals – Lesbian, Gay, Bi, Trans, Queer and beyond.

Lack of data In one of the few studies on the impact of AI onto queer communities, [21] addresses the problem of “unobserved characteristics”, *i.e.* the absence of certain attributes in datasets that could nevertheless benefit queer communities. This is the case, for example, with sexual orientation and gender identity, which would be useful to know when making health decisions. [21] explains this lack of attributes by over-protective data regulations – as in Europe, where it is forbidden to process “sensitive” data – and by the non-measurable nature of some of these attributes. Similarly, several researchers, such as [9], advocate the collection of data on queer individuals, in order to make them visible and serve as a lever for action. This lack of data is even truer in parts of the world with less access to the internet, or where being queer is more penalized, legally and socially.

Privacy abuses While there is a lack of data on queer people, protecting data and especially sensitive attributes is also a pressing issue. Indeed, disclosure of a person’s sexual orientation or gender identity can be a matter of life and death in many countries. A few years ago, [23] sought to recognize a person’s sexual orientation from their facial features. In addition to being scientifically flawed [2, 8], seeking to display these attributes can be harmful to queer individuals, putting them in danger if their identity is exposed without their consent.

Biased representations A large portion of biases in model outputs originate from the data, whether due to historically skewed distributions, collection and sampling, measurement or preprocessing [20]. A skewed representation of data can have harmful consequences, as illustrated by the COMPAS system in the U.S., which indicates a higher risk of recidivism among black offenders than among white offenders [1], as black people are more often incarcerated. Similar results are likely to emerge in the future for queer individuals, who are statistically more likely to be incarcerated [7], have less higher education [18], have low incomes [25], be homeless [13] or face health disparities [3].

3 Why is the topic not more researched

The academic fields of AI/ML fairness and queer studies have grown considerably in recent years, and the subject is being explored by para-scientific communities, such as artists [15] and the creation of support and discussion groups [11]. Yet at the academic level, the intersection between the two remains largely uncharted.

This lack of interest is primarily due to the fact that queer communities represent only a small percentage of the population. There is, therefore, less incentive to push these issues forward, especially for AI/ML solutions that need to work for the greatest number of people.

As well as being few in number, queer individuals are not well represented in the computer science community either, or in STEM² in general. They are also likely to have a more negative experience of the workplace [6], leading to a loop situation in which the field is even less attractive to queer individuals. This can result in issues specific to queer communities being ignored, unintentionally, by non-queer individuals working on AI/ML systems, because they feel less concerned.

4 Hopes and future research

The easiest fix to queer issues in AI/ML is probably the technical dimension. There is already a vast literature dealing with the imbalance of datasets and underrepresented groups [10], and bias mitigation [17], but it still needs to be fleshed out. Training methods and evaluation metrics are currently being developed for underrepresented groups [22]. Research is also being carried out into privacy protection methods for ML, although these are still in their infancy [16].

However, problems persist when it comes to taking into account the particularity of queer issues and not just looking at the big picture of AI/ML fairness. There is still work to be done in terms of visibility of queer issues and awareness-raising. For instance, ML practitioners need to be more concerned about labeling issues, defining population groups or simply not developing harmful systems overall. Better integration of queer individuals into the computer science community and greater visibility for the work of queer researchers will be essential to shed light on these issues.

The impact of AI/ML on queer communities needs special attention, because of the specific challenges it poses. If, as a society, we want to progress technologically, we cannot do so by leaving people behind. Meeting those challenges will thus require a shift in point of view, to consider the few behind the many. A complete change in how we design and implement AI/ML systems will be necessary, reconsidering fundamentals, such as the simple use of labels and groups, as well as reconnecting with the social and ethical dimension of these systems.

²Science, Technology, Engineering and Mathematics

About the authors

Mélanie Gornet is a PhD student at Télécom Paris – Institut Polytechnique de Paris, on the ethics and regulation of AI systems. Her current projects focus on the interplay between social, regulatory and technical issues through the operational implementation of ethical criteria for AI systems, such as explainability and fairness. She holds an engineering degree in data science and AI, and studied gender/queer theory at SciencesPo. She self-identifies as queer, has volunteered for queer associations, and has conducted research on diversity in LGBTQ+ activism.

Tiphaine Viard is a *maîtresse de conférences* (roughly equivalent to tenured assistant professor) of AI and society in the social sciences laboratory at Télécom Paris – Institut Polytechnique de Paris. Her work and expertise are at the crossroads of graph modelling for real-world phenomena and explicating the social world of artificial intelligence. She self-identifies as queer, and has volunteered for queer associations since 2017.

References

- [1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine Bias. *ProPublica*, May 2016. URL: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- [2] Blaise Agüera y Arcas. Do algorithms reveal sexual orientation or just expose our stereotypes? *Medium*, January 2018. URL: <https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477>.
- [3] Thomas M. Barwick. NIH Recognizes LGBTQ Community as ‘Health Disparity Population’. *NBC News*, October 2016. URL: <https://www.nbcnews.com/feature/nbc-out/nih-recognizes-lgbtq-community-health-disparity-population-n661161>.
- [4] Colin R. Blyth. On Simpson’s Paradox and the Sure-Thing Principle. *Journal of the American Statistical Association*, 67(338):364–366, 1972. American Statistical Association, Taylor & Francis, Ltd.
- [5] Joy Buolamwini and Timnit Gebru. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, pages 77–91. PMLR, January 2018.
- [6] Erin A. Cech and Michelle V. Pham. Queer in STEM Organizations: Workplace Disadvantages for LGBT Employees in STEM Related Federal Agencies. *Social Sciences*, 6(1), March 2017. Multidisciplinary Digital Publishing Institute.
- [7] Carolyn Crist. LGBT individuals more likely to be incarcerated. *Reuters*, December 2016. URL: <https://www.reuters.com/article/us-health-lgbt-incarceration-usa-idUSKBN14C1ZI>.
- [8] Andrew Gelman, Greggor Mattson, and Daniel Simpson. Gaydar and the fallacy of objective measurement. February 2018.
- [9] Kevin Guyan. *Queer Data - Using Gender, Sex and Sexuality Data for Action*. Bloomsbury Publishing, January 2022.
- [10] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239, May 2017.
- [11] Hetvi Jethwani, Arjun Subramonian, William Agnew, MaryLena Bleile, Sarthak Arora, Maria Ryskina, and Jeffrey Xiong. Queer in AI. *XRDS: Crossroads, The ACM Magazine for Students*, 28(4):18–21, June 2022.
- [12] Os Keyes. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–22, November 2018.
- [13] Carrie Kovalick. LGBTQ+ Youth Homelessness, September 2023. URL: <https://nn4youth.org/lgbtq-homeless-youth/>.
- [14] Cassie Kozyrkov. What is “Ground Truth” in AI? (A warning.), October 2021. URL: <https://towardsdatascience.com/in-ai-the-objective-is-subjective-4614795d179b>.
- [15] Blog Writing Guide Lines. Coding queer bias: An artist’s response to the lack of diversity in machine learning.
- [16] Bo Liu, Ming Ding, Sina Shaham, Wenny Rahayu, Farhad Farokhi, and Zihuai Lin. When Machine Learning Meets Privacy: A Survey and Outlook. *ACM Computing Surveys*, 54(2):31:1–31:36, March 2021.
- [17] Dana Pessach and Erez Shmueli. A Review on Fairness in Machine Learning. *ACM Computing Surveys*, 55(3):51:1–51:44, February 2022.
- [18] Dario Sansone. LGBT students: New evidence on demographics and educational outcomes. *Economics of Education Review*, 73, December 2019.
- [19] Morgan Klaus Scheuerman, Jacob M. Paul, and Jed R. Brubaker. How Computers See Gender: An Evaluation of Gender Classification in Commercial Facial Analysis Services. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–33, November 2019.
- [20] Harini Suresh and John Guttag. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO ’21, pages 1–9, New York, NY, USA, October 2021. Association for Computing Machinery.

-
- [21] Nenad Tomasev, Kevin R. McKee, Jackie Kay, and Shakir Mohamed. Fairness for Unobserved Characteristics: Insights from Technological Impacts on Queer Communities. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 254–265, Virtual Event USA, July 2021. ACM.
 - [22] Angelina Wang, Vikram V Ramaswamy, and Olga Russakovsky. Towards Intersectionality in Machine Learning: Including More Identities, Handling Underrepresentation, and Performing Evaluation. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 336–349, Seoul Republic of Korea, June 2022. ACM.
 - [23] Yilun Wang and Michal Kosinski. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology*, 114(2):246–257, 2018. American Psychological Association.
 - [24] Jamie Wareham. Why Artificial Intelligence Is Set Up To Fail LGBTQ People. *Forbes*, March 2021. URL: <https://www.forbes.com/sites/jamiewareham/2021/03/21/why-artificial-intelligence-will-always-fail-lgbtq-people/>.
 - [25] Bianca D M Wilson, Lauren J A Bouton, M V Lee Badgett, and Moriah L Macklin. LGBT Poverty in the United States - Trends at the Onset of COVID-19. Technical report, William Institute UCLA School of Law, February 2023.