Rebuilding Trust: Queer in AI Response to the NIST Artificial Intelligence Risk Management Framework

Rebuilding Trust

We applaud the NIST framework's focus on trustworthy AI. AI, machine learning, and data science methods are already pervasive in our society and technology, affecting all of our lives in many subtle ways. Trustworthy AI has become an important topic because that trust has been lost, or was never present in the first place. Researchers, corporations, and governments have long and painful histories of excluding marginalized groups from technology development, deployment, and oversight. As a direct result of this exclusion, these technologies have long histories of being less useful or even harmful to minoritized groups. For example, Google, a major developer and deployer of machine learning where Black people are systematically marginalized and excluded, especially in research groups, deployed AI that labelled Black people as "gorillas" in 2015[1]. Despite widespread condemnation, in 2018 Amazon's computer vision tool matched several members of the Congressional Black Caucus with criminals[2], and in September 2021 Facebook AI labelled videos of Black men as "primates"[3]. This infuriating history illustrates that industry cannot be trusted to self-regulate and why trust in commercial AI systems and development has been lost. We argue that any AI development, deployment, and monitoring framework that aspires to trust must incorporate both participatory design principles and strong, outside, and continual monitoring and testing. Finally, we urge NIST to consider aspects of trustworthiness beyond just transparency, fairness, and accountability, specifically, to consider justice and shifting power to the people and disempowered as core values to any trustworthy AI system[4].

Trust Requires Verification

Monitoring and advocacy by independent organizations has emerged as one of the strongest forces for creating more trustworthy AI[567]. Any standards for trustworthy AI should include methods for publishing information about AI systems to facilitate outside inspection. We

---

[1] https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/

[2] https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

[3] https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html

[4] https://www.nature.com/articles/d41586-020-02003-2

[5] https://www.aclu.org/issues/privacy-technology/surveillance-technologies/face-recognition-technology

[6] https://techinquiry.org/

[7] https://stoplapdspying.org/

strongly recommend starting with the datasheets for datasets[8] and model cards[9] frameworks as the basis for this. In addition, much of this independent monitoring is done on a volunteer basis, or as part of nonprofits that struggle to find ethical funding in a landscape dominated by the very corporations and individuals profiting from harmful AI. Privacy, civil rights, and trustworthy AI organizations and researchers are frequently funded by the companies they are meant to monitor and regulate[10], creating serious conflicts of interest. A meaningful trustworthy AI initiative will include establishment of robust and independent funding mechanisms to support and expand this verification and bias detection work.

Just as important, many groups work to improve inclusion in AI development and empower minoritized individuals and communities. Trustworthy AI cannot be developed by a field that systematically excludes; minoritized communities must be involved at every stage of AI development and empowered to change or stop harmful development. Queer people[11], Black people[12], women[13], and other marginalized groups are systematically excluded and unrepresented in AI. Organizations such as Queer in AI[14], Black in AI[15], LatinX in AI[16], Women in Machine Learning[17], Widening NLP[18], and Diversity in AI[19] have long-running and successful programs for improving representation in AI and making the AI field less hostile to minoritized scientists. Creating trustworthy AI does not start with the development of a specific AI system; it starts by funding, supporting, and empowering groups like these so the field of AI has the diversity and inclusion to credibly and effectively develop trustworthy AI.

Beyond Trust with Participatory Design

We propose that AI systems employ a participatory design process from a Feminist standpoint[20]. As opposed to the western scientific epistemology, it is important that AI systems consider the plurality of experiences in their design process, which is key to designing for marginalised users[21] (could elaborate more if needed). We recommend adhering to the qualities of pluralism, participation, advocacy, ecology, embodiment and self-disclosure as proposed by Prof. Shaowen Bardzell in their seminal work on Feminist HCI[19]. With the use of participatory

---

[8] https://arxiv.org/pdf/1803.09010.pdf
[9] https://arxiv.org/pdf/1810.03993.pdf
[10] https://twitter.com/RMac18/status/1436065518222397444
[11] https://www.science.org/doi/10.1126/sciadv.abe0933
[12] https://towardsdatascience.com/are-there-black-people-in-ai-fb6928166d73
[13] https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/
[14] https://sites.google.com/view/queer-in-ai/
[15] https://blackinai.github.io/#/
[16] https://www.latinxinai.org/
[17] https://wimlworkshop.org/
[18] http://www.winlp.org/
[19] http://www.diverseinai.org
[20] Harding, S. The Feminist Standpoint Theory Reader: Intellectual and Political Controversies. Routledge, 2003.
[21] Shaowen Bardzell. 2010. Feminist HCI: taking stock and outlining an agenda for design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1301–1310. DOI:https://doi.org/10.1145/1753326.1753521

design, researchers and practitioners must also acknowledge various considerations in implementing it in the case of marginalised users.

The power-distance between the practitioners and users as co-designers must be minimised or overcome before involving them as participants. For instance, anthropologists and ethnographers employ trust-building mechanisms (*rapport*)[22] before probing into the experiences of users, as opposed to traditional HCI research which employs discount ethnographic methods in laboratory settings, which do not acknowledge the relationship between the user and the practitioner.

Different layers of intersectionality should be addressed in terms of representation when it comes to co-design processes. For instance, despite employing participatory design, early ICTs deployed across rural areas of India deepened the inequalities due to differential factors such as caste and gender in terms of accessibility[23]. Hence, practitioners involved in building AI systems for marginalised stakeholders must address (i) the cultural context in which the technology use is situated and (ii) issues of accessibility in their design process.

Against Participation as Exploitation - Ethical Considerations for Participatory Design

It should be mandatory for the researchers and practitioners to self-disclose and make the users aware of the various ways in which the AI system might use their data and affect them before seeking their consent and involving them as co-designers. In this process, they also must document and address concerns of these users through their feedback. We encourage the researchers to be reflexive in the process of making design decisions and include positionality statements, highlighting or accepting potential risks (if any) which might need to be addressed in the future. They are also encouraged to highlight the same when it comes to mitigating these risks within the AI system.

Further, we would like to strongly warn researchers and practitioners against the notion of 'participation-washing'[24] i.e. exploitative and extractive means of community involvement in the name of participatory design. Sloane et. al.[25] suggest (i) developing reward systems to compensate users for their work, (ii) maintaining transparency and explaining the involvement of communities in participatory design and (iii) documenting harms and risks by AI systems and learning from them to ensure long term participation as justice in Machine Learning in an equitable fashion.

---

[22] https://doi.org/10.1080/0951839890020105
[23] https://doi.org/10.1016/j.techsoc.2020.101251
[24] https://www.technologyreview.com/2020/08/25/1007589/participation-washing-ai-trends-opinion-machine-learning/
[25] https://arxiv.org/abs/2007.02423

Grassroots Queer AI Community Building and Activism

Queer in AI[26] is a decentralized, grassroots organization whose mission is to raise awareness of queer issues in AI/ML, foster a community of queer researchers and celebrate the work of queer scientists. Queer in AI started at NeurIPS 2017 when a group of artificial intelligence (AI)/machine learning (ML) researchers—frustrated and angry at the lack of community, representation, and inclusion of queer people in AI—created a queer forum on the NeurIPS conference app and met up for coffee. We spent the next year (hundreds and hundreds of emails and many, many meetings!) organizing, creating our mission and platform, and hosting our first workshop at NeurIPS 2018[27].

Queer in AI works to make the AI community a safe and inclusive place that welcomes, supports, and values *all* LGBTQIA+ people. We build a visible community of queer and ally AI scientists. We also raise awareness of queer issues in the general AI community and encourage and highlight research on and solutions to these problems. We recognize that the "queer experience" is not monolithic, and prioritize uplifting and centering the voices of those who face intersecting vectors of marginality, especially trans, non-binary, and queer people of color and individuals from the global south.

To improve inclusion of queer people and adovate against AI harmful to queer people, Queer in AI organizes a variety of programs:

- Queer in AI hosts workshops, socials, panels, and poster sessions at nearly every AI/ML/natural language processing (NLP) conference each year. Through these events, we empower queer and trans AI researchers and advance research at the intersections of AI and queer and trans identities. In the past, topics have included creating safer spaces for trans and non-binary folks in AI, ensuring trans-inclusive publishing, and AI policy and queer privacy. At each workshop, we also make a call for submissions related to the intersection of LGBTQIA+ representation and AI *or* research produced by LGBTQIA+ individuals. We accept submissions in any media (e.g. research papers, poetry, music, etc.) and any language, to maximize the inclusion of queer and trans perspectives. Queer in AI has hosted six workshops (NeurIPS '18,'19,'20, ICML '19,'20,'21) and dozens of social and meetups at top AI/ ML, NLP, robotics, and cognitive science conferences.
- Queer in AI hosts a large, active online community of over 500 queer AI scientists, which serves as a nexus for commnity, support, and activism.
- Queer in AI administers a demographic survey to identify issues within diverse queer communities and shape the future programs of Queer in AI; all questions are optional and anonymous. Through the survey, we ask our community members about inclusion at conferences, mental health, access to research opportunities, and much more. Our survey finds queer people often feel unwelcome in many AI spaces, experience high

---

[26] https://queerinai.org/about
[27] https://queerai.github.io/QueerInAI/QinAIatNeurIPS.html

rates of mental health issues, and are frequent targets of harassment. We list some statistics from the 2021 demographic survey[28] in Appendix 1.

- We advise conferences on queer and trans issues. We have improved gender and pronouns questions on conference registration forms, implemented rapid systems for name changes, and protected the privacy and safety of queer and trans conference attendees. We have additionally written extensively about how to make virtual conferences queer-friendly[29].
- We operate the graduate school application review and fee aid program[30], a joint effort with oSTEM. We work to eliminate application and test fees as a barrier to queer STEM scholars applying to graduate programs. We additionally guide queer grad applicants in discussing their identities in their personal statements and other application materials by connecting students with queer researchers in their fields.

By creating and empowering a diverse community of queer scientists and activists, Queer in AI provides the basis for both expert critique and monitoring of AI systems from a queer perspective and participatory design approaches including queer people. Queer in AI offers our knowledge, experience, and large, grassroots community during the development of NIST AI Risk Management Framework.

Queer Data and Machine Learning

Through our years of work and advocacy as a Queer AI and data science organization, we have developed expert knowledge around questions of if and how gender, sexuality, and other aspects of identity should be used in AI systems. Below we summarize our main considerations. We would be happy to consult further on specific questions that arise during creation of the AI RMF.

While participatory approaches are essential to the design of human-centered machine learning systems, the feasibility of any system and mitigating its harms should be critically considered.

In 2018, researchers at Stanford attempted to show that deep neural networks are more accurate than humans at detecting sexual orientation from images of people[31]. Sexuality is broadly defined, encompassing one's experience of their romantic, sexual, and emotional behaviors, and is fluid and flexible; therefore, it is inherently not detectable by a human or deep neural network. As such, even a participatory approach is not appropriate for the design of a sexuality detection system; no amount of input from diverse queer participants can compensate for the reality that the premise of the system is based in physiognomy-esque science, assumes the biological essentialism of sexuality and expression, and is fundamentally flawed. Furthermore, a sexuality detection system, regardless of how effective, may be easily

---

[28] https://forms.gle/7B8kRYJJEWSkj8QH9

[29] http://queerinai.org/diversity-guide

[30] https://queerinai.org/graduate-school-application-programs

[31] https://www.gsb.stanford.edu/faculty-research/publications/deep-neural-networks-are-more-accurate-humans-detecting-sexual

weaponized by powerful institutions against queer and straight people alike, with severe representational and allocative harms[32], from violation of privacy to death.

In addition to sexuality, researchers and engineers have proposed systems to detect transness or inclusively detect non-binary genders[33]. But, these systems are also inherently flawed and cannot benefit from participatory design because of 1) their invalid assumption that one's expression can predict their gender identity and 2) their treatment of gender as an immutable, discrete quantity, which grossly mischaracterizes the flexible and fluid nature of one's inner sense of their gender. This is primarily a consequence of research not being conducted by trans folks, which grassroots organizations like Queer in AI help mitigate.

Furthermore, these gender detection systems often poorly, or entirely don't, consider intersectionality: 1) they lack data of non-binary persons facing intervening vectors of marginality 2) they suffer from exnomination, in which researchers implicitly rely on the visual definition of non-binary persons as Western, white, androgynous people with stereotypically queer experiences 3) they treat "non-binary" as a monolithic, third gender, when in reality non-binary genders comprise all genders beyond the binary.

Most machine learning systems, such as those employed in ad targeting and commercial gender recognition, and attempts to mitigate their harms, focus on binary gender. The collection of binary gender data and inference of binary gender forces non-binary individuals to misgender themselves or be misgendered by systems, as well as suffer cyclical erasure[34], in which the assumption of gender as binary is encoded into machine learning models, thereby reinforcing and perpetuating dangerous, hegemonic ideas about gender being binary.

It is further paramount that we expand our understanding of gender biases and harms to include non-binary individuals. However, no matter how inclusive gender categories are made, bucketizing gender and categorizing people will inevitably harm large segments of non-binary individuals, through misgendering, erasure, and underreperesentation[35]. Quoting "Harms of Gender Exclusivity and Challenges in Non-Binary Representation in Language Technologies":

> *Any viable solution cannot simply be a quick fix or patch, but must rely on a bottom-up approach involving affected persons system-wide, such as in annotation and human-in-the-loop mechanisms. Simultaneously, research into monitoring [machine learning models] over time to detect harms against non-binary individuals is critical. It is further paramount to transparently communicate the performance of [models] for non-binary persons and possible harms. In the case of harm, non-binary individuals must be able to obtain valid recourse to receive a more favorable outcome, as well as have the opportunity to provide feedback on the model's output and have a human intervene.*

Towards mitigating machine learning biases and harms faced by non-binary individuals, it is critical to integrate participatory design to survey and understand their lived experiences[36].

---

[32] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899
[33] https://venturebeat.com/2020/07/14/study-describes-facial-recognition-system-designed-to-identify-non-binary-people/
[34] https://arxiv.org/pdf/2108.12084.pdf
[35] https://arxiv.org/pdf/2108.12084.pdf
[36] https://arxiv.org/pdf/2108.12084.pdf

In doing so, it is important to collect demographics in an inclusive manner[36], to capture the intersecting identities of respondents and not over-generalize the experience of harms. We highly recommend consulting HCI Guidelines for Gender Equity and Inclusivity for gender-inclusive design principles[37].

In the next section we present a framework for a gendered approach to artificial intelligence to help foresee, understand, and combat these problems.

A Gendered Approach to Artificial Intelligence

The theory of 'intelligence' and the epistemology operationalized by AI research focuses only on a specific form of knowing. Feminist work has demonstrated that this excludes other epistemologies, including those traditionally gendered as female or feminine[38]. Critical race theory has also analogized its equivalence to the knowledge of the white man[39]. AI, therefore, takes part in a wider socio-technical exclusion or repression of women's knowledge and reifies a gendered and racialized conceptualization of 'intelligence'[40]. Work on AI and epistemology must countenance the possibility that AI and epistemic justice is an illegitimate combination.

Alternatively, AI may provide an opportunity to shift assumptions about male and female epistemology. For example, the narrative of 'hard' and 'soft' intelligence is often gendered as masculine and feminine respectively. Adrian Weller notes how this 'hard' intelligence, often thought to encompass logic and rationality, is much easier to reproduce in technological form. Subsequently, it is reinforced as encompassing all 'intelligence' by the fact it is adopted in these machines, but what has hitherto been thought of as 'soft' intelligence may become more privileged in being harder to encode. As Sarah O'Connor's notes: As machines become better at many cognitive tasks, it is likely that the skills they are relatively bad at will become more valuable. This list includes creative problem-solving, empathy, negotiation and persuasion.[41] She goes on to say that these qualities have historically "been more identified with – and encouraged in – women".

Whether AI is thought to depend upon and epitomize a masculinist epistemology, or whether AI promises to give a feminine epistemology the advantageous position in the job market, AI is perpetuating and reinforcing binary, gendered stereotypes of epistemology.

---

[37] https://www.morgan-klaus.com/gender-guidelines.html
[38] Adam, A. (1998). Artificial knowing: Gender and the thinking machine. London:Routledge.
[39] Ali, Mustafa (2019). "White Crisis" and/as "Existential Risk", or The Entangled Apocalypticism of Artificial Intelligence. Zygon: Journal of Religion and Science, 54(1), pp.207-224.
[40] Davies, S (2019). 'Women's minds matter', [online] Aeon, Available at: https://aeon.co/essays/feminists-never-bought-the-idea-of-a-mind-set-freefrom-its-body
[41] O'Connor, S. (2019). The robot-proof skills that give women an edge in the age of AI. [online] Financial Times. Available at: https://www.ft.com/ content/06afd24a-2dfb-11e9-ba00-0251022932c8

Queer Epistemology: a Proposal

Although feminist theory has often been applied to technological practice[42], approaches to gender in technology have been critiqued by trans writers for their ignorance of trans lives[43]. The use of gender theory needs to be broadened to further apply trans, non-binary, queer, and postcolonial theory to explore the fundamental barriers to equality embedded in the design and purpose of technologies. In addition, research would consider how we could replace these aspects with alternative, inclusive practices, or recommend against the use of AI in certain contexts altogether. Achieving the goal of social equality would be aided by dialogue between gender theorists and technologists. But at present, gender theory and AI practice "are speaking completely different languages"[44]. Susan Leavy points out that, currently, the people who are reading gender theory are not the co-authors of papers such as 'Turning a Blind Eye: Explicit Removal of Biases and Variation from Deep Neutral Networks'[45].

Gina Neff highlights this problem of the growing distance between those who are designing and deploying these systems, and those who are affected by these systems. What will ordinary people do to respond to challenges, adapt and give feedback that will be crucial for the positive evolution of these systems? Neff refers to the importance of the 'social shaping of AI', which would include designing workshops with users and including them in the discussion of how systems could be adapted to work for their benefit[46]. In light of the problems outlined above, some aspects of technology in particular need to be challenged:

- The reproduction of stereotypical gendered and racialized bodies in robotics. Currently, there is a disjunction between the theoretical 'suspension of gender' which could be promising for destroying restrictive gender stereotypes, and the encapsulation of physical gender stereotypes in technology and robotics.
- The use of language, interaction and communication in relation to these systems. This refers to both the use of gendered 'voices' and 'responses' which these systems produce, in addition to the use of gendered pronouns and syntax. Susan Leavy points out that while some recent studies have sought to remove bias from learned algorithms, they largely ignore decades of research on how gender ideology is embedded in language. The mechanisms which reinforce this gendered language in technology include, for instance: the way in which certain genders are named, ordered, and

---

[42] Adam, ibid.

[43] Keyes. O, (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. In Proceedings of the ACM, New York NY, US. on Human-Computer Interaction - CSCW archive. 2 (CSCW) Article 88. [online] 22 pages Available at: https://doi. org/10.1145/3274357

[44] Leavy, S. (2019). Uncovering gender bias in newspaper coverage of Irish politicians using machine learning. Digital Scholarship in the Humanities, 34(1), pp.48-63.

[45] Alvi, M., Zisserman, A., and Nellaker, C. (2018). Turning a Blind Eye: Explicit Removal of Biases and Variation from Deep Neural Network Embeddings. [online] Arxiv. [Preprint] Available at: arXiv:1809.02169

[46] Leavy, S. (2018). Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. In: 1st International Workshop on Gender Equality in Software Engineering (GE). New York NY, US: ACM, pp.14-16.

described, as well as the frequency with which they are referred to, and the metaphors used to describe them[47].

- The reinforcement of societal understandings of gendered epistemology in AI systems. It must challenge such structures and incorporate the insights of queer epistemology into technology.

Alongside theoretical analysis, research would also pursue a multilateral conversation with international stakeholders, technologists and designers. Seeking to understand the conceptions and definitions of gender and race, and why/ how they are being embedded into technological design, would be crucial to assessing how theory can speak to practice. In turn, these would be combined to produce a set of research-based tools which could be employed by designers and technologists to embed pluralism and inclusion into AI systems, or to suggest against the use of specific AI systems altogether. The Stanford University initiative, 'Gendered Innovations', directed by Londa Schiebinger, is focusing on integrating sex and gender research into technology. This inclusion of sex and gender analysis into "basic and applied research produces excellence in science, health & medicine, and engineering research, policy, and practice"[48]. These vital pursuits need to be extended and amplified in relation to AI technologies in order to make the design, marketing, advertisement and the ultimate purpose of these systems work for social justice. Techno-utopianism has been criticized for coming from a place of white privilege[49]. In addition, earlier feminist hopes that the online and virtual world would provide a site for freedom from gender constraints and inequality[50] have, with time, been unfulfilled. However, there is still potential for creative engagement with these AI technologies which might be used to challenge stereotypes surrounding gender[51].

Methodology

1. Theoretical analysis

Queer, trans, feminist, postcolonial, and other nonessentialist gender theory would be used to analyse the design and functionality of current and emerging AI technology. In terms of embodiment, the relationship between transgenderism and embodiment would be explored in

---

[47] Leavy (2019), Ibid.

[48] European Commission. Directorate-General for Research Innovation, (2013). Gendered innovations: How gender analysis contributes to research: Report of the expert group 'Innovation through gender' (EUR (Luxembourg), 25848). Luxembourg: Publications Office.

[49] Ali, Mustafa (2019). "White Crisis" and/as "Existential Risk", or The Entangled Apocalypticism of Artificial Intelligence. Zygon: Journal of Religion and Science, 54(1), pp.207-224.

[50] Haraway, D. (1985). Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. Socialist Review, 80, pp.65-108.
Haraway, D. (1997). Modest_Witness@Second_Millenium.FemaleMan[C]_ Meets_OncoMou seTM: Feminism and technoscience. Journal of the History of Biology, 30(3) pp. 494-497.
Wajcman, J. (2006) The Feminisation of Work in the Information Age? In: M. Frank Fox, D. Johnson and S. Rosser, eds., Women, Gender, and Technology. Champaign, Ill.: University of Illinois Press, pp. 80-97.
[51] Ferrando, F. (2014). Is the post-human a post-woman? Cyborgs, robots, artificial intelligence and the futures of gender: A case study. European Journal of Futures Research, 2(1), pp.1-17.

relation to AI[52]. It would also be enlightening to consider how the female body can be constructed and perceived, and the role that power and politics plays in structuring and training bodies into certain gendered roles and behaviours[53]. Focusing on intersectionality, it would be useful to concentrate on how racial ideology shapes gender narratives and discourse[54] and, in addition, focus on the intersection of race, queer and/or trans identity, and how relevant these historical narratives are in contemporary society. These intersections and relationships could be used to reflect on how such narratives could be considered in AI technology[55]. The queer concept of continual disruption of repetitively reinforced, historically meaningful gender norms could also be explored, especially in relation to how this could be reflected in AI systems[56].

[52] Elliot, P. and Roen, K. (1998) Transgenderism and the Question of Embodiment: Promising Queer Politics? GLQ: A Journal of Lesbian and Gay Studies, 4(2), pp. 231-261.
Stanley, E. and Smith, N. (2011). Captive genders: Trans embodiment and the prison industrial complex. Edinburgh: AK Press
[53] Butler, J. (1990). Gender Trouble: feminism and the subversion of identity. New York: Routledge.
Foucault, M. (1991 [1975]). Discipline and punish: The birth of the prison. London: Penguin Books.
Grosz, E. (1994). Volatile bodies: Toward a corporeal feminism. Bloomington: Indiana University Press.
[54] Collins, P. H. (2005). Black Sexual Politics: African Americans, Gender, and the New Racism. New York: Routledge.
Oyěwùmí, O. (1997) Invention of Women: Making an African Sense of Western Gender Discourses. Minneapolis; London: University of Minnesota Press.
Spivak, G. (1988) 'Can the Subaltern Speak?', in Nelson, C. and Grossberg, L. (eds.) Marxism and the Interpretation of Culture. Urbana: University of Illinois Press, pp. 271-313. Spivak, G. (1988) In Other Worlds: Essays in Cultural Politics. New York: Routledge.
[55] hooks, b. (1996). Reel to Real: Race, sex, and class at the movies. New York, London: Routledge.
Puar, J (2007). Terrorist assemblages. Durham: Duke University Press.
Snorton, C. R. (2017). Black on Both Sides: A Racial History of Trans Identity. Minneapolis: University of Minnesota Press.
Stone, S. (1992) The Empire Strikes Back: A Postranssexual Manifesto. Camera Obscura, 10(2 29), pp. 150-176.
Stryker, S. (2008). Transgender history. Berkeley, CA: Seal Press
Whittlestone, J. et al. (2019). Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research. London: Nuffield Foundation
Ware, S. M. (2017). All Power to All People? Black LGBTTI2QQ Activism, Remembrance, and Archiving in Toronto. TSQ, 4 (2), pp.170–180.
[56] Ahmed, S. (2006). Queer Phenomenology. Durham: Duke University Press.
Barad, K. (2011). Nature's Queer Performativity. Qui Parle: Critical Humanities and Social Sciences, 19(2), pp.121-158.
Butler, J. (1993). Critically Queer GLQ: A Journal of Lesbian and Gay Studies, 1(1), pp.17-32.
Campbell, A. and Farrier, S. (2015). Queer Practice as Research: A Fabulously Messy Business. Theatre Research International, 40(1), pp. 83-87.
Eng, D. L. with Halberstam, J. and Munoz, J.E. (2005) What's Queer about Queer Studies Now? Social Text, 23(3-4), pp. 84-85.
Freeman, E. (2011). Time Binds: Queer Temporalities, Queer Histories. North Carolina: Duke University Press.
Kember, S. (2016). iMedia. The gendering of objects, environments and smart materials. UK: Palgrave MACM, New York NY, USA. illan.
Warner, M. (1999) The Trouble with Normal: Sex, politics and the ethics of queer life. Cambridge, MA: Harvard University Press.

2. Trans-sectoral communication

Conducting a multi-lateral conversation with international stakeholders, technologists and designers would encourage the researcher to understand contemporary non-western conceptions of what defines and frames gender and race and how these definitions inform technological practice. It must assess how these fit into data, computing, design and implementation of AI systems that have, or will have, a global impact, despite the regional focus of their development. There is significant theory available that can be mobilised to address technological challenges. Researchers literate in this theory need to engage with those who are designing and implementing AI in order to develop dialogue, understanding and progress.

3. Synthesis and recommended tools

Bridging the gap between gender and queer theory and AI practice would require synthesizing the theoretical work and the communicative work to produce research-based, practical tools. The aim would be for these tools to be employed and incorporated into the way that these technologies are designed and used in society. These tools could inform the technological process at all stages, as well as the more political aspects of technological creation. This would include data gathering; algorithm design; the purpose of technology; technological use; implementation and distribution. In addition, it would explore where systems should not be deployed or would be inappropriate given the goal of social justice.

Challenges of this Approach

- There may be challenges when it comes to synthesizing the theoretical work and the material gathered from conversing with technologists. This multi-source approach, however, will also introduce a valuable insight into the nature of any disjunctions between theoretical and practical definitions of gender and race in AI systems.
- There may be challenges in deciding how these tools can be framed in order to be easily implemented by technologists.

Value of a Queer Methodology and Epistemology

- This will foster international collaboration and networks, looking to create unity around an ambition for social equality and justice in relation to technology and its implementation.
- It will gather an entirely new, unique evidence base of cultural understandings about how technology intersects with gender and race.
- These technologies are often exclusive, restrictive, and controlling in relation to gender and race. This could expose the ways in which this is happening and seek to open up pluralistic and inclusive ways that technology can be developed.

Warmly,

Arjun Subramonian
William Agnew
Ashwin
Juan Pajaro
Queer in AI Organizers
queerinai@gmail.com

Appendix 1: Selected 2021 Queer in AI Demographic Survey Results

Selected results from the 2021 Queer in AI Demographic Survey[57] (n=142).

Demographics of survey respondents:
- 30% non-cis folks
- 21.5% BIPOC folks
- 10% mentioned disabilities
- 50% graduate students
- 20% have changed their name or plan to do so

Feelings of belonging:
- 85% reported lack of role models
- 80% reported lack of community
- 3.4 (1-5, 5 high) rating on how comfortable queer folks feel at conferences
- 4.6 (1-5) rating on how welcome queer folks feel at Queer in AI events

Feelings of inclusion:
- 2.6 (1-5) rating on satisfaction with gender forms by trans folks
- 2.8 (1-5) rating on name change policy satisfaction
- 3.5 (1-5) rating on deadnames not being cited

Mental health:
- 55% have had ability to conduct research or participate in classes impaired by mental health issues
- 31% felt their mental health was lowest within the last year
- 26% have seriously considered or attempted suicide
- 20% have harmed themselves

Discrimination and Harassment
- 29% have been targets of innuendos and/or jokes
- 27% have been deliberately ignored or excluded
- 27% have been singled out as the "resident authority" due to their identity
- 25% have been the target of derogatory comments (written or in person)
- 20% have felt intimidated/bullied
- 16% have feared for their physical safety
- 15% have received unsolicited physical contact

---

[57]
https://docs.google.com/forms/d/e/1FAIpQLSes-IzwkKHruQrAmH3Tnz1tJsTUl-YP51V8wDtHbfb8Z9FoNg/viewform