Challenges to Grassroots Organization Engagement with AI Policy

Jennifer Mickel,* Carter Buckner,* William Agnew, Yanan Long, Michelle Lin, B.V. Alaka, Angelina Wang, Sarthak Arora, Nandhini Swaminathan, Arjun Subramonian
Queer in AI

Abstract

Around the world, policies are being developed to address privacy, economic, intellectual property, energy, and other risks that AI technologies pose. Simultaneously, institutions are creating standards and best practices to further the use of AI. The development of standards and policies involves many well-resourced actors and opaque development processes, often sidelining the needs of marginalized populations, who often lack extensive networks, lobbying capabilities, and other forms of power. In this paper, we present the participatory development of AI policies that meet the needs of queer people through grassroots advocacy. We use collaborative autoethnography to surface granular challenges our organization has faced in doing so, along with factors that assisted us. We conclude with actionable recommendations for empowering marginalized communities to participate in policy development and insights for other marginalized communities working to develop policy to mitigate harms.

1 Introduction

In the United States, there have been numerous AI governance bills introduced at national and local levels within the past two years (50; 82; 41; 8). However, forms of data and technology governance have existed well before the introduction of LLMs to combat data breaches and data misuse (54; 27). These related conversations on data protection, as well as internet use and free speech, have heavily informed newer AI governance approaches. In this work, we present an analysis and critique of our organization's approach to engaging with AI policy and consider future directions to strengthen participation from marginalized communities in large-scale governance efforts.

AI systems cause harm through bias, misinformation, misrepresentation, and privacy infringement (3; 77). Queer and marginalized groups experience AI harms in different forms (3) which may be ill-addressed through universal governance approaches. Throughout history, laws, policies, and court decisions have been enacted that have harmed queer and marginalized communities (76; 46; 68). This precedent necessitates the need for proactive rather than reactive measures in AI governance to minimize the negative impact of AI systems on queer and marginalized communities.

The public plays an important role in addressing AI harms, as they both interact with and feel the impact of these systems. However, the public's direct participation in AI governance is challenging, as policy advancements are a non-linear, complex combination of informal interactions, timing, and people creating a "perfect storm" where policy advancements occur. Organizations that are decentralized, volunteer-led, or directly involved in translating technical jargon and policy can

* These authors contributed equally.
Correspondence to: <queerinai@gmail.com>.

act as a proxy between communities and governance bodies. The successes of large-scale private participation,[2] offer insight into how to scale participatory governance in practice.

With the recent "participatory turn in AI design" (22), calls for empowering data and AI subjects, especially those from marginalized communities, have grown as a means of addressing issues of bias, stereotyping, and other AI harms. Within participatory AI (PAI), there is an ever-evolving effort to align AI to users needs, reduce power imbalances, and reduce harm from AI systems (5; 19). Further, we echo PAI scholars' assertion that communities have essential expertise – and as in PAI, this expertise is essential to AI governance. Achieving large-scale PAI remains challenging, as attempts to broaden participation in AI development frequently risk being shallow and extractive (44; 5; 66). We discuss the importance of participation in governance in Appendix C.

In this paper, we describe the motivations, structure, and outputs of our organization Queer in AI – a queer nonprofit that has spent the past two years engaging in U.S. AI policy[3]. We then use collaborative autoethnography (43; 34) to explore how our attempts at implementing participatory methods in AI policy – both within our organization and in broader policy processes we participated in – succeeded and failed. Beyond the policy contributions of our work, we use the tools of collaborative autoethnography to critically reflect on our experiences to give the broader AI ethics community a grounded look at what participatory AI governance looks like in practice and suggestions to make these methods more practical. We also provide other marginalized communities working in technology and AI spaces with lessons and reflections to aid in their advocacy.

2 AI Policy that Considers Queer People

Understanding the development and deployment of AI technologies requires a sociotechnical perspective that recognizes how these systems affect the people interacting with or subject to them (62). Although significant progress has been made in addressing the negative impacts of AI, involvement from marginalized communities is often absent from these discussions (74).

In particular, a lack of queer perspectives in AI research leads to inherent biases that undermine the inclusivity of these technologies (59; 23). These biases create systems that not only fail to serve queer individuals adequately but can actively enable the weaponization of AI to perpetuate oppression, censorship, and stigmatization, further marginalizing this already vulnerable community (30; 40; 73). Without a more inclusive approach to AI development, these harms will likely continue and proliferate (45). Local groups such as the Detroit Community Technology Project[4] and San Jose (U.S.) Government AI Coalition[5] use participation to develop safeguards and increase access. Civil society organizations and academic research groups also contribute to community data advocacy and AI governance inclusive of marginalized groups. Local legislative efforts have recently created provisions for "AI advisory councils," which offer a mechanism to increase community involvement, though often without input from marginalized, non-expert groups. Similar are efforts to expand consent and privacy power (24). Federally, AI governance efforts have shifted away from regulatory efforts around job protection, civil and data rights, to pro-business growth. We examine this gap in research and advocacy through our organization's involvement in the National Institute for Standards and Technology (NIST) AI Safety Artificial Intelligence Safety Institute Consortium (AISIC) and resultant policy report.

Throughout this work, we refer to "harms" caused by AI systems, which we mean to refer broadly to outputs by AI systems that result in discrimination against or misrepresentation of a social group.

[2] In the U.S., corporate unionization efforts are an example of the success of large-scale private participation (42).

[3] See Appendix A for a limited overview of the U.S. AI policy landscape.

[4] Examples of the projects the Detroit Community Technology Project engages in can be found here: https://detroitcommunitytech.org/?q=projects.

[5] Templates and resources developed by the San Jose Goverment AI Coalition can be found here: https://www.sanjoseca.gov/your-government/departments-offices/information-technology/ai-reviews-algorithm-register/govai-coalition.

3 Engaging in AI Policy as a Grassroots Organization

In this section, we present our organization's experiences engaging in policy work as a grassroots organization, discussing challenges faced and factors that helped. A detailed explanation of our organization's engagement with AI policy and timeline is located in Appendix B.

3.1 Organization Policy Goals

Our organization, Queer in AI, is a 501(c)(3) non-profit professional association that aims to raise awareness of queer issues in artificial intelligence and machine learning. By fostering a supportive and decentralized community of queer researchers, the group champions principles of intersectionality, participatory inclusion, and community-led initiatives (59). We also advocate for AI, data, and digital technology that ensure privacy, cannot be used for surveillance, can represent fluid and complex queer identities, and do not degrade the environment or undermine creative economies. We fight for positive uses of these technologies for queer people, particularly in protecting and improving the online spaces where many queer people find community and knowledge while fighting against the usage of these technologies for surveillance, including facial recognition (64; 14). Our policy approach arose from these beliefs and took shape in the forms of (i) inclusive AI policy development; and (ii) inclusive design and application of AI systems.

3.2 Organization Structure

Our organization has a flat organizational structure, with only three formal roles: *members*, those who are in the Slack workspace or email list; *organizers*, those who have or are helping organize an event or initiative; *core organizers*, people who have been organizers for several years. In our Slack workspace, the majority of channels are public, excluding a channel where core organizers occasionally discuss issues involving personally identifiable information (PII) or topics that have a high likelihood of negatively impacting individuals or the organization should they be widely publicized. This structure embodies our organization's stated values of decentralization, allowing any member a high degree of transparency into the organization and a swift path to becoming involved in organizing. In 2022, a policy action group within Queer in AI launched in response to increasing requests for our organization to take policy positions. The policy action group existed as a public Slack channel, a weekly stand-up meeting, and several in-person events. Both the Slack channel and the stand-up meetings were open to all members, leading many members to join the initiative.

3.3 Organization Challenges

Our organization faced several practical challenges in organizing and doing policy advocacy work.

AI policy work is often in-person in a few select geographical centers. Although our members and volunteers are spread throughout the world, with particular concentrations of members in the United States, no member of the policy team was based in cities hosting major AI policymaking institutions or events. These institutions and events were most often held in Washington D.C., London, Brussels, or San Francisco. Our organization could not afford to hire full-time staff or pay for policy team members to travel to AI policy events, which prevented us from attending almost all AI summits, workshops, and other events. Crucially, this also prevented us from engaging in activities surrounding such events (e.g., dinners, side meetings, other networking) that could help us meet allies, improve our policy knowledge, and build our organization's "inside game" (47) (cf. (9)).

AI policy is often opaque. Such physical distances amplified another commonly reported issue: difficulty understanding relevant discussions and ideas within AI policy. Our organizers often reported feeling confused and surprised by developments in AI policy. Specifically, organizers working with the NIST AISIC often felt blindsided by releases of documents for public comments, changes in directions of working groups, and responses to (or lack of response to) their feedback. Organizers discussed feeling uncertain about the purpose and intent of different policy documents: whether a bill was intended to be passed or meant for signaling, or whether a policy document would ever be meaningfully enforced. This uncertainty led to organizers being unsure of where to spend their limited time, and feeling that meaningful and critical conversations were had in meetings and spaces they were not a part of, and that their efforts may have thus been wasted.

Many of Queer in AI's policy team members faced barriers to participating in unpaid policy work. The policy team consisted primarily of students and junior academics, required to balance coursework and pressure to publish with contributing their expertise. Some organizers reported institutional targeting for their activism and organizing work, pressure to simultaneously contribute to multiple organizations advocating for marginalized communities, illnesses, and institutions, and supervisors who were unsupportive of work in AI outside of narrow technical domains. In addition, most recently within the United States, a changing funding landscape has made it extremely difficult to support work focusing on the impact of AI on marginalized communities.

AI, data, and digital technology have vast impacts on queer people, making the scope of focus challenging. While our initial focus was on AI policy, we quickly realized we would be unable to meet our policy goals without considering the full development pipeline. Broadly, AI used for surveillance (37), censorship (16), and other harms can disproportionately impact the queer community. In particular, "lower tech" algorithms such as developing an algorithm to classify images (31; 81) or through gendered algorithms (2; 28) propagate harms early through data collection, use-case definition, and system architecture. We draw a distinction between narrow and strong AI (i.e., GPT and foundation models) to highlight that narrow AI systems are well studied and often have clear pathways to redressal. Foundation models introduce a broader set of explicit and potential harms and knowledge of their effect on queer people is still evolving (32). Together, this creates a broad focus of harms, including innumerable technologies and policy directions (e.g., proliferation of censorship and bias in social media algorithms, online hate speech and targeting, discrimination in algorithmically mediated decisions) through a variety of AI-enabled modalities and techniques. We initially found this broad set of topics too much for our team to effectively cover, and we struggled to prioritize which topics to advocate for. Ultimately, we decided to focus on general values and principles that would be applicable to a variety of policy decisions rather than develop detailed and actionable positions on each issue. We also chose this strategy because of the opaque and rapidly changing nature of AI policy; new policy documents or issues would be raised often and without warning, hindering the ability to provide detailed responses or specific proposals.

Formal communication structures allow limited stakeholder participation. Executive frameworks, which regulate the development of, deployment of, and ethical considerations around AI technologies, often fail to prioritize the concerns of marginalized communities (55). One of the mechanisms intended to address this gap is the Request for Information (RFI) process, used by organizations to solicit feedback from various stakeholders. RFIs are intended to democratize policy development, offering stakeholders, including marginalized communities, a platform to influence decisions that affect them. In theory, RFIs are an essential tool for fostering inclusive dialogue and ensuring that a wide range of perspectives are considered in the regulatory process. Research highlights that when RFIs function as intended, they can help incorporate a variety of voices, leading to more comprehensive and balanced policy outcomes (21). In reality, the RFI process often falls short of achieving genuine inclusion. RFI mechanisms can be opaque and difficult to navigate, especially for marginalized communities that lack the resources or institutional knowledge to effectively participate. Furthermore, while RFIs invite public feedback, they can lack the transparency and accountability needed to ensure that this input leads to real policy change. In a meeting with NIST AISIC leadership, we were informed that our RFC/RFI responses had not been considered in their discussions. Our organization understood the communication norms to be through responses to NIST requests and participation in NIST AISIC activities, though they appeared to have limited impact on NIST AISIC outcomes. Recognizing these limitations, our organization identified the need for more effective way to ensure marginalized voices could drive tangible change. For example, we produced a policy report to fill this gap, offering a clear, accessible framework to further empower public participation in AI governance.[6]

Queer in AI was typically the only queer-focused and one of few civil society organizations in the room, which created high expectations, pressure, and feelings of extraction. Our organization was the only queer-focused organization to be invited to testify in front of the National AI Advisory Committee, and the only queer-focused organization and one of the few civil society members of the 280 members of the NIST AISIC. As an organization, we felt pressure to be both excellent advocates for queer people and to represent all queer people and perspectives. This is analogous

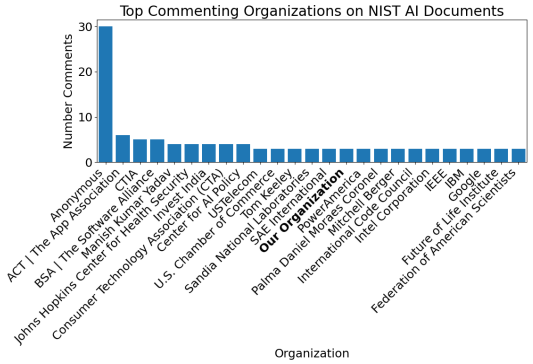
[6]Our policy document is available here: https://www.queerinai.com/policy-research

to what individuals can feel as the only members of a marginalized community in a group (56). This additional pressure contributed to stress and burnout among our organizers. Being the only queer-focused organization also "pigeon-holed" policy discussions. Our organization's positions were interpreted as representative of all queer people, rather than a productive discussion occurring between multiple queer organizations, leading to more nuanced positions representative of more queer people. Our group also struggled with how to relate to others: as the only queer-focused organization, we felt pressure to both boldly and uncompromisingly represent policy stances that would lead to good futures for queer people and simultaneously develop relationships with other organizations and make compromises to be invited to the table.

Our organizers often described worrying about why policymakers were including our organization in meetings and programs, citing concerns that they were included to superficially boost the diversity of participants *without challenging the underlying power structures* and that our positions and expertise had not been taken seriously. These concerns validate ongoing discussions of the pitfalls of participation within AI ethics literature (e.g. (66; 5)). Our organizers continued to participate in policy efforts even when they had concerns about the intentions of engagement because they knew they were the only group invited to represent the needs of queer people, and refusing to participate would mean queer people would not have any voice.

Furthermore, our organizers often reported feeling that civil society and marginalized communities were sidelined in U.S. policymaking. They frequently referenced the membership of the NIST AISIC (52), where for-profit corporations comprise the majority of the membership. They also cited the industry-heavy compositions of high-profile AI policy events (4), and the industry backgrounds of key AI appointees (20) as evidence for this belief. We collected all documents from `regulations.gov` posted by NIST with the keyword `Artificial Intelligence`, and all comments on those documents, finding that the top commentors have been mostly industry and lobbying groups (see Figure 1). There were a few civil society and academic organizations, including our organization, among the most frequent of the over 500 organizations commenting on NIST AI policy. Our organizers also noted that NIST and other organizations seemed eager to form deep partnerships with industry AI developers (60), but the role of civil society was much less clearly defined.

Figure 1: 30 entities that commented the most frequently on NIST documents with the keyword "Artificial Intelligence".



3.4 Factors that Helped Organization Amid Challenges

Mentorship and collaboration from larger, more experienced organizations. Our organizers have built a relationship with a larger civil society organization that provided them with extensive mentorship and collaboration. Staff from this organization met frequently with us and provided feedback on document drafts and advice in shared Slack spaces. At the start of our policy work, we had very limited policy experience. The mentorship from this organization provided crucial help in understanding norms in policy spaces, including how to format different responses, what to expect of and how to engage with different government entities, and how AI policymaking compares to previous successful and unsuccessful policy initiatives. This mentorship also helped build trust and solidarity between us, the organization mentoring us, and other organizations associated with the mentor organization. This helped build our policy network, aiding us in overcoming barriers we faced as an organization not physically located near government centers and composed of people unfamiliar with policy spaces.

Many volunteers, open meetings. Our organizers also cited the large number of volunteers they were able to involve in policy efforts, and the underlying values and structures that enabled this, as a key factor that enabled them to engage in many policy efforts. As of writing, our policy Slack

channel had 78 members, including dozens of people with or pursuing advanced degrees in computer science, law, or other relevant fields, showing we had a multidisciplinary team. We used several strategies to effectively engage with this team. Our organization has established a large, active Slack space with over 1,500 members, and by repeatedly advertising the policy team and its meetings, particularly when members organically brought up policy concerns, many people joined. Barriers to joining our policy team were also minimized; the Slack channel was kept public and anyone could join the meetings. Our meeting time was frequently spent orienting people joining for the first time, or rejoining after absences due to work, school, or other reasons. Although this took away from time spent on direct policy work, this structure ensured there were enough active volunteers. We hosted a policy workshop co-located with NeurIPS in 2024 that helped us engage with more community members, and all workshop participants were invited to join the Slack channel and policy team. Many organizers noted that this openness is in unresolved tension with another key value of online queer spaces: privacy and security. Organizers worried that the openness of the policy team could lead to malicious actors joining and disrupting proceedings or doxxing participants, mirroring frequent concern for online queer communities (1).

4 Recommendations

We offer several recommendations for policymakers, funders, and civil society organizations to increase the participation and impact of marginalized communities in AI policymaking.

Moving away from scarcity mindsets in AI policy programs. Mentorship by a larger and more experienced organization was vital to our efforts. However, obtaining this mentorship required years of building our organization's reputation and network, which in turn relied on many of our members having the privilege of being inside academia. Our members applied for many formal policy workshops, fellowships, and mentorship programs, but they were space-limited, and we were rejected from all of them. If we had relied on being accepted by larger organizations or labs to which we lacked deep connections, our policy work would have never occurred. The scarcity model many AI policy mentorship opportunities are built on excludes marginalized communities and stands in contrast to our demonstrated values of welcoming and mentoring anyone and everyone we can.

Organizing and mobilizing for power. We observed that civil society groups were already disempowered, and we believe that coming together and building collective power is the best way to counter the vast interests opposing effective AI regulation. Nonetheless, we frequently heard about other meetings and groups similar to ours, generally all opaque, secret, and invite-only. While there are reasons for some discussions to be private, we believe the default opacity of AI policymaking needlessly impedes mobilizing people and organizations, building solidarity, and taking collective actions. Our organization again demonstrates the possibility of being highly transparent, and welcoming, mentoring, and empowering each person who wishes to join.

All gatherings should be virtual or strong hybrid. Physical distance from important policy meetings and centers consistently excluded our organization from advocating for our positions, networking, and gathering information. Making all gatherings either entirely virtual or strong hybrid would greatly reduce this barrier. Here *strong* means that virtual participation does not play second fiddle to the in-person components and that adequate infrastructures (e.g., internet, virtual meeting, A/V) are in place. Allowing public participation will prevent these events from becoming exclusive gatherings of elite academics, top government officials, and industry, as would providing public meeting notes or recordings.

5 Conclusion

In this work, we provide an analysis of organizational challenges resulting from our participation with the U.S. NIST AI Safety Institute and Consortium and creation of a policy report detailing inclusive AI practices and public participation. To the former, we find the highly centralized, opaque, and frenetic nature of federal governance a barrier to increased participation by academics and civil society organizations advocating for marginalized groups. Further, the federal policy landscape can be diversified through mentorship and support – especially from large, well-funded organizations – aimed at smaller advocacy groups, coalition building, and less emphasis on geographic locality

to policy hubs. In this work, we also provide the catalyst for creating a policy report. While we were initially invited to participate in NIST AISI, as the only queer-focused organization and one of few civil society participants, we noticed a sharp early shift away from civil and data rights-based approaches. This felt especially at odds with stances taken within the Biden administration (i.e., Blueprint for an AI Bill of Rights). The policy report was our organization's response to that dissonance and a way to draw importance to (i) what inclusive AI stances must be considered; and (ii) why participation in AI development and governance is essential. Our findings offer insight into how participation and PD theory can interact with the existing mechanisms of federal governance.

Positionality Statement

The authors of this paper have formal training as researchers and practitioners in machine learning, natural language processing, privacy, security, fairness, ethics, and legal research. Many authors also identify as members of marginalized communities and have expertise in other areas. The authors also have varying levels of experience in policy design, participatory governance, participatory AI design, advocacy, unionizing, and activism. Our shared experiences heavily influence how we write about our policy work. The focus of this paper is on policy conversations in the U.S. though, the authors are from North America, Asia, and Europe and are knowledgeable about related policy conversations in these locations.

References

[1] ALI CONDON. Meta told to do more to protect LGBTQ+ users from doxxing, outing, and abuse, 2024. https://www.thepinknews.com/2024/01/24/meta-protect-lgbtq-users-outing-abuse.

[2] AMERICAN CIVIL LIBERTIES UNION. Four ways the tsa is making flying easier for transgender people, 2022.

[3] BAROCAS, S., CRAWFORD, K., SHAPIRO, A., AND WALLACH, H. The problem with bias: Allocative versus representational harms in machine learning. In *Proceedings of SIGCIS* (Philadelphia, PA, 2017).

[4] BIDEN WHITE HOUSE. Readout of White House Roundtable on U.S. Leadership in AI Infrastructure , 2024. https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/09/12/readout-of-white-house-roundtable-on-u-s-leadership-in-ai-infrastructure.

[5] BIRHANE, A., ISAAC, W., PRABHAKARAN, V., DIAZ, M., ELISH, M. C., GABRIEL, I., AND MOHAMED, S. Power to the people? opportunities and challenges for participatory ai. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (2022), pp. 1–8.

[6] BOMMASANI, R., ARORA, S., CHOI, Y., FEI-FEI, L., HO, D. E., JURAFSKY, D., KOYEJO, S., LAKKARAJU, H., NARAYANAN, A., NELSON, A., PIERSON, E., PINEAU, J., VAROQUAUX, G., VENKATASUBRAMANIAN, S., STOICA, I., LIANG, P., AND SONG, D. A path for science- and evidence-based ai policy, 2024. https://understanding-ai-safety.org/.

[7] BOSSEN, C., TALHOUK, R., AND VINES, J. Scaling Participatory Design. In *Routledge International Handbook of Contemporary Participatory Design*, 1 ed. Routledge, London, Dec. 2024, pp. 27–50.

[8] BRENNAN CENTER FOR JUSTICE. Artificial Intelligence Legislation Tracker | Brennan Center for Justice, Oct. 2025. https://www.brennancenter.org/our-work/research-reports/artificial-intelligence-legislation-tracker.

[9] BØDKER, S., DINDLER, C., AND IVERSEN, O. S. Tying Knots: Participatory Infrastructuring at Work. *Computer Supported Cooperative Work (CSCW) 26*, 1 (Apr. 2017), 245–273.

[10] BØDKER, S., DINDLER, C., IVERSEN, O. S., AND SMITH, R. C. *Participatory Design*. Synthesis Lectures on Human-Centered Informatics. Springer International Publishing, Cham, 2022.

[11] BØDKER, S., AND KYNG, M. Participatory Design that Matters—Facing the Big Issues. *ACM Trans. Comput.-Hum. Interact. 25*, 1 (Feb. 2018), 4:1–4:31.

[12] CARPENTER, D., AND EZELL, C. An fda for ai? pitfalls and plausibility of approval regulation for frontier artificial intelligence. In *AAAI/ACM Conference on AI, Ethics, and Society* (2024).

[13] CASPER, S., KRUEGER, D., AND HADFIELD-MENELL, D. Pitfalls of evidence-based ai policy. *ArXiv abs/2502.09618* (2025).

[14] CHEREPANOVA, V., GOLDBLUM, M., FOLEY, H., DUAN, S., DICKERSON, J., TAYLOR, G., AND GOLDSTEIN, T. Lowkey: Leveraging adversarial attacks to protect social media users from facial recognition. *arXiv preprint arXiv:2101.07922* (2021).

[15] CLEMENS, B., LEFORT-BESNARD, J., RITTER, C., SMITH, E., VOTINOV, M., DERNTL, B., HABEL, U., AND BZDOK, D. Accurate machine learning prediction of sexual orientation based on brain morphology and intrinsic functional connectivity. *Cerebral cortex* (2022).

[16] COBBE, J. Algorithmic censorship by social platforms: Power and resistance. *Philosophy & Technology 34*, 4 (2021), 739–766.

[17] COMMITTEE, N. A. A. Naiac year 1 report, May 2023. https://www.ai.gov/wp-content/uploads/2023/05/NAIAC-Report-Year1.pdf, urldate = 2025-05-21.

[18] COMMITTEE, N. A. A. Naiac year 2 report, May 2024. https://ai.gov/wp-content/uploads/2024/06/National-Artificial-Intelligence-Advisory-Committee_Year-Two-Insights-Report.pdf.

[19] CORBETT, E., DENTON, E., AND ERETE, S. Power and public participation in ai. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (2023), pp. 1–13.

[20] DAN PRIMACK. What to know about Sriram Krishnan, whose AI appointment sparked MAGA outrage, 2024. https://www.axios.com/2024/12/27/sriram-krishnan-maga-civil-war-backlash.

[21] DARBISHIRE, H. Proactive transparency: The future of the right to information?, 2009.

[22] DELGADO, F., YANG, S., MADAIO, M., AND YANG, Q. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (2023), pp. 1–23.

[23] DEV, S., MONAJATIPOOR, M., OVALLE, A., SUBRAMONIAN, A., PHILLIPS, J., AND CHANG, K.-W. Harms of gender exclusivity and challenges in non-binary representation in language technologies. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (Online and Punta Cana, Dominican Republic, Nov. 2021), M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds., Association for Computational Linguistics, pp. 1968–1994.

[24] ELECTRONIC PRIVACY INFORMATION CENTER (EPIC), AND CONSUMER REPORTS. The State Data Privacy Act: A Proposed Compromise by Consumer Advocates. https://epic.org/wp-content/uploads/2025/04/State-Data-Privacy-Act-EPIC-CR-April2025.pdf, Apr. 2025.

[25] ESCOBAR, A. *Designs for the pluriverse: radical interdependence, autonomy, and the making of worlds*. New ecologies for the twenty-first century. Duke University press, Durham, 2018.

[26] EUROPARL. Eu ai act: first regulation on artificial intelligence, 2024. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[27] EUROPEAN PARLIAMENT, AND COUNCIL OF THE EUROPEAN UNION. Regulation (EU) 2016/679 of the European Parliament and of the Council.

[28] FOR TRANSGENDER EQUALITY, N. C. Whole body imaging faq, June 2009.

[29] GLOBAL PARTNERSHIP ON ARTIFICAL INTELLIGENCE, AND OECD. Oecd ai principles overview, 2024. https://oecd.ai/en/ai-principles.

[30] GOMES, A., ANTONIALLI, D., AND DIAS-OLIVA, T. Drag queens and artificial intelligence: Should computers decide what is toxic on the internet? http://www.internetlab.org.br/en/freedom-of-expression/drag-queens-and-artificial-intelligence-should-computers-decide-what-is-toxic-on-the-internet, 2019. Internet Lab.

[31] GRANT, N., AND HILL, K. Google's photo app still can't find gorillas. and neither can apple's. image, 2023.

[32] HAIMSON, O. L., MAYWORM, S. R., INGBER, A. S., AND ANDALIBI, N. Ai attitudes among marginalized populations in the u.s.: Nonbinary, transgender, and disabled individuals report more negative ai attitudes. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2025), FAccT '25, Association for Computing Machinery, p. 1224–1237.

[33] HENDRYCKS, D., MAZEIKA, M., AND WOODSIDE, T. An overview of catastrophic ai risks. *ArXiv abs/2306.12001* (2023).

[34] HUGHES, S. A., AND PENNINGTON, J. L. *Autoethnography: Process, product, and possibility for critical social research*. Sage Publications, 2016.

[35] HUYBRECHTS, L., TELI, M., SAAD-SULONEN, J., AND SCIANNAMBLO, M. Collectives and participatory design. *Routledge International Handbook of Contemporary Participatory Design* (2025).

[36] JAKESCH, M., BUÇINCA, Z., AMERSHI, S., AND OLTEANU, A. How different groups prioritize ethical values for responsible ai. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (2022).

[37] KALLURI, P. R., AGNEW, W., CHENG, M., OWENS, K., SOLDAINI, L., AND BIRHANE, A. The surveillance ai pipeline. *arXiv preprint arXiv:2309.15084* (2023).

[38] KATIE LANGIN. LGBTQ Ph.D. graduates will soon be counted in key U.S. survey, 2024. https://www.science.org/content/article/lgbtq-ph-d-graduates-will-soon-be-counted-key-u-s-survey.

[39] KENSING, F., AND GREENBAUM, J. Heritage: having a say. In *Routledge international handbook of participatory design*, J. Simonsen and T. Robertson, Eds. Routledge, 2013, pp. 21–36.

[40] KEYES, O. The misgendering machines: Trans/hci implications of automatic gender recognition. *Proceedings of the ACM on human-computer interaction 2*, CSCW (2018), 1–22.

[41] KIBBY, C. US State Privacy Legislation Tracker, 2025.

[42] KNEPPER, M. From the fringe to the fore: Labor unions and employee compensation. *Review of Economics and Statistics 102*, 1 (2020), 98–112.

[43] LAPADAT, J. C. Ethics in autoethnography and collaborative autoethnography. *Qualitative inquiry 23*, 8 (2017), 589–603.

[44] MAAS, J., AND INGLÉS, A. M. Beyond participatory ai. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2024), vol. 7, pp. 932–942.

[45] MICKEL, J., DE-ARTEAGA, M., LIU, L., AND TIAN, K. More of the same: Persistent representational harms under increased representation. *arXiv preprint arXiv:2503.00333* (2025).

[46] MISSISSIPPI SUPREME COURT. Plaxico v. michael. Southern Reporter, 1999. 735 So. 2d 1036 (Miss. 1999).

[47] MOSER, R. What is the inside/outside strategy? *COUNTERPUNCH* (2016).

[48] NAACP. Artificial Intelligence in Predictive Policing Issue Brief | NAACP, Feb. 2024. https://naacp.org/resources/artificial-intelligence-predictive-policing-issue-brief.

[49] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (US). Artificial intelligence risk management framework : generative artificial intelligence profile. Tech. rep., National Institute of Standards and Technology, Gaithersburg, MD, July 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

[50] NCSL. National conference of state legislatures artificial intelligence 2024 legislation, 2024. https://www.ncsl.org/technology-and-communication/artificial-intelligence-2024-legislation.

[51] NEWSOM, G. Sb 1047 veto message. Office of the Governor of California, September 2024. https://www.gov.ca.gov/wp-content/uploads/2024/09/SB-1047-Veto-Message.pdf.

[52] NIST AISIC. AISIC Members, 2024. https://www.nist.gov/aisi/artificial-intelligence-safety-institute-consortium/aisic-members.

[53] NOZZA, D., BIANCHI, F., LAUSCHER, A., AND HOVY, D. Measuring harmful sentence completion in language models for LGBTQIA+ individuals. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion* (Dublin, Ireland, May 2022), B. R. Chakravarthi, B. Bharathi, J. P. McCrae, M. Zarrouk, K. Bali, and P. Buitelaar, Eds., Association for Computational Linguistics, pp. 26–34.

[54] OF AUSTRALIA, C. Privacy act 1988, 1988. `https://www.legislation.gov.au/Series/C2004A03712`.

[55] PARTHASARATHY, S., AND KATZMAN, J. Bringing communities in, achieving ai for all. *Issues in Science and Technology 10* (2024).

[56] POWELL, G. N., AND GRAVES, L. Women and men in organizations. *Newbury Park, CA* (1989).

[57] QUEER IN AI. Ai harms and proposed policy interventions, 2024. Accessed: 2025-05-08.

[58] QUEERINAI, DENNLER, N., OVALLE, A., SINGH, A., SOLDAINI, L., SUBRAMONIAN, A., TU, H., AGNEW, W., GHOSH, A., YEE, K., PERADEJORDI, I. F., TALAT, Z., RUSSO, M., AND DE JESUS DE PINHO PINHAL, J. Bound by the bounty: Collaboratively shaping evaluation processes for queer ai harms. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (2023).

[59] QUEERINAI, O. O., OVALLE, A., SUBRAMONIAN, A., SINGH, A., VOELCKER, C., SUTHERLAND, D. J., LOCATELLI, D., BREZNIK, E., KLUBICKA, F., YUAN, H., J, H., ZHANG, H., SHRIRAM, J., LEHMAN, K., SOLDAINI, L., SAP, M., DEISENROTH, M. P., PACHECO, M. L., RYSKINA, M., MUNDT, M., AGARWAL, M., MCLEAN, N., XU, P., PRANAV, A., KORPAN, R., RAY, R., MATHEW, S., ARORA, S., JOHN, S., ANAND, T., AGRAWAL, V., AGNEW, W., LONG, Y., WANG, Z. J., TALAT, Z., GHOSH, A., DENNLER, N., NOSEWORTHY, M., JHA, S., BAYLOR, E., JOSHI, A., BILENKO, N. Y., MCNAMARA, A., GONTIJO-LOPES, R., MARKHAM, A., DONG, E., KAY, J., SARASWAT, M., VYTLA, N., AND STARK, L. Queer in ai: A case study in community-led participatory ai. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2023), FAccT '23, Association for Computing Machinery, p. 1882–1895.

[60] RICHARD PRESS. U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation With Anthropic and OpenAI, 2024. `https://www.nist.gov/news-events/news/2024/08/us-ai-safety-institute-signs-agreements-regarding-ai-safety-research`.

[61] SADEK, M., CALVO, R. A., AND MOUGENOT, C. Closing the socio–technical gap in ai: The need for measuring practitioners' attitudes and perceptions. *IEEE Technology and Society Magazine 43* (2024), 88–91.

[62] SARTORI, L., AND THEODOROU, A. A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics and Information Technology 24*, 1 (Jan. 2022), 4.

[63] SCHEUERMAN, M. K., PAPE, M., AND HANNA, A. Auto-essentialization: Gender in automated facial analysis as extended colonial project. *Big Data & Society 8* (2021).

[64] SHAN, S., WENGER, E., ZHANG, J., LI, H., ZHENG, H., AND ZHAO, B. Y. Fawkes: Protecting privacy against unauthorized deep learning models. In *29th USENIX security symposium (USENIX Security 20)* (2020), pp. 1589–1604.

[65] SLATTERY, P., SAERI, A. K., GRUNDY, E. A. C., GRAHAM, J., NOETEL, M., UUK, R., DAO, J., POUR, S., CASPER, S., AND THOMPSON, N. The ai risk repository: A comprehensive meta-review, database, and taxonomy of risks from artificial intelligence. *ArXiv abs/2408.12622* (2024).

[66] SLOANE, M., MOSS, E., AWOMOLO, O., AND FORLANO, L. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Arlington, VA, USA, Oct. 2022), EAAMO '22, Association for Computing Machinery.

[67] SUBRAMONIAN, A., AND GAUTAM, V. Factoring the matrix of domination: A critical review and reimagination of intersectionality in ai fairness. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (2023).

[68] SUPREME COURT OF THE UNITED STATES. Masterpiece cakeshop, ltd. v. colorado civil rights commission, 2018.

[69] THE WHITE HOUSE. Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government , 2020. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.

[70] THE WHITE HOUSE. Advancing Racial Equity and Support for Underserved Communities Through the Federal Government , 2021. https://www.federalregister.gov/documents/2021/01/25/2021-01753/advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government.

[71] THE WHITE HOUSE. Preventing and Combating Discrimination on the Basis of Gender Identity or Sexual Orientation, 2021. https://www.federalregister.gov/documents/2021/01/25/2021-01761/preventing-and-combating-discrimination-on-the-basis-of-gender-identity-or-sexual-orientation.

[72] THE WHITE HOUSE. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 2023. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

[73] THIAGO, D. O., MARCELO, A. D., AND GOMES, A. Fighting hate speech, silencing drag queens? artificial intelligence in content moderation and risks to lgbtq voices online. *Sexuality & culture 25*, 2 (2021), 700–732.

[74] TREWIN, S., BASSON, S., MULLER, M., BRANHAM, S., TREVIRANUS, J., GRUEN, D., HEBERT, D., LYCKOWSKI, N., AND MANSER, E. Considerations for ai fairness for people with disabilities. *AI Matters 5*, 3 (Dec. 2019), 40–63.

[75] UNITED STATES. WHITE HOUSE OFFICE, I. B. *White House toolkit on transgender equality.* The White House, Washington, D.C, 2021.

[76] U.S. SUPREME COURT. Plessy v. ferguson. United States Reports, 1896. 163 U.S. 537.

[77] WANG, A., MORGENSTERN, J., AND DICKERSON, J. P. Large language models that replace human participants can harmfully misportray and flatten identity groups, 2025.

[78] WANG, Y., AND KOSINSKI, M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. *Journal of Personality and Social Psychology 114* (2018), 246–257.

[79] WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY. Blueprint for an ai bill of rights: Making automated systems work for the american people, 2022. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/.

[80] ZAHLSEN, O. K., PARMIGGIANI, E., AND DAHL, Y. Challenges of Scaling Participatory Design: A Systematic Literature Review. In *OzCHI '22* (New York, NY, USA, Apr. 2023), Association for Computing Machinery, pp. 143–159.

[81] ZHANG, M. Google photos tags two african-americans as gorillas through facial recognition software, 2015.

[82] ZWEIFEL-KEEGAN, R. S. C. US State AI Governance Legislation Tracker, 2025. https://iapp.org/resources/article/us-state-ai-governance-legislation-tracker/.

A AI Policy in the United States

We offer a brief context of AI governance efforts around 2024. AI policy in the U.S. is not necessarily reflective of the rest of the world, but it is relevant to understanding the scope of policy landscapes we engaged in.

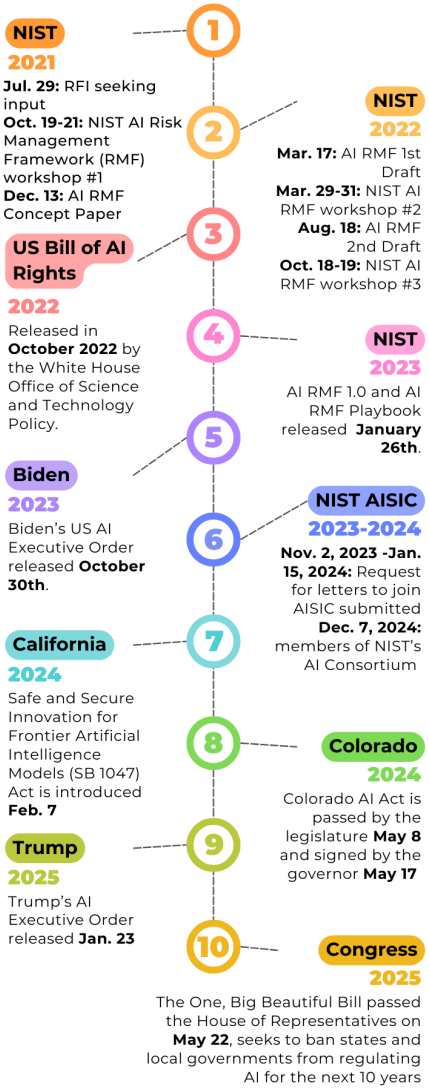
U.S. public AI governance efforts have led to a few significant trends (see Figure 2). In late 2022, the White House Office of Science and Technology released the Blueprint for an AI Bill of Rights (79), signaling a desire to establish rights for individuals who directly use AI systems or whose personal data is collected for use in AI systems. It was met with varying criticism, but was quickly followed by congressional bills (8) and the establishment of the National Institute for Standards and Technology (NIST) AI Safety Institute (AISI) and Consortium (AISIC) in 2024. Three approaches emerged during this time: rights-based, risk-based, and evidence-based.

Rights-based approaches appear explicitly in the Blueprint for an AI Bill of Rights, though their adoption across executive agencies remains inconsistent. This approach is also supported by the U.S. alignment with the intergovernmental Organization for Economic Cooperation and Development's AI Principles for Trustworthy AI (29) and previous Biden executive orders on AI use in the Federal government (69) and civil rights (70; 71).

Risk-based approaches are introduced in the Biden Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence and the resultant creation of a National AI Advisory Council, NIST AISI, and AISIC. Each of these entities works in tandem to codify risk in the NIST Risk Management Framework (RMF) and respond to mandates within the executive order. These mandates explicitly apply to executive agencies, with wider adoption by private entities being voluntary. The NIST AISIC members came from academia, civil society, and private companies. NIST AISIC members participated in working groups designed to codify risks and harm, as well as in red-teaming activities (i.e., empirical risk exploration) and guidance creation.

A wider coalition forms among academia and government for evidence-based approaches (13; 6; 51). Calls for empirical approaches appear to be in reaction to existential-risk conversations that were gaining traction at the time (33). Supporters of evidence-based approaches desired that empirical evidence be shown regarding AI risk and harm prior to policy considerations.

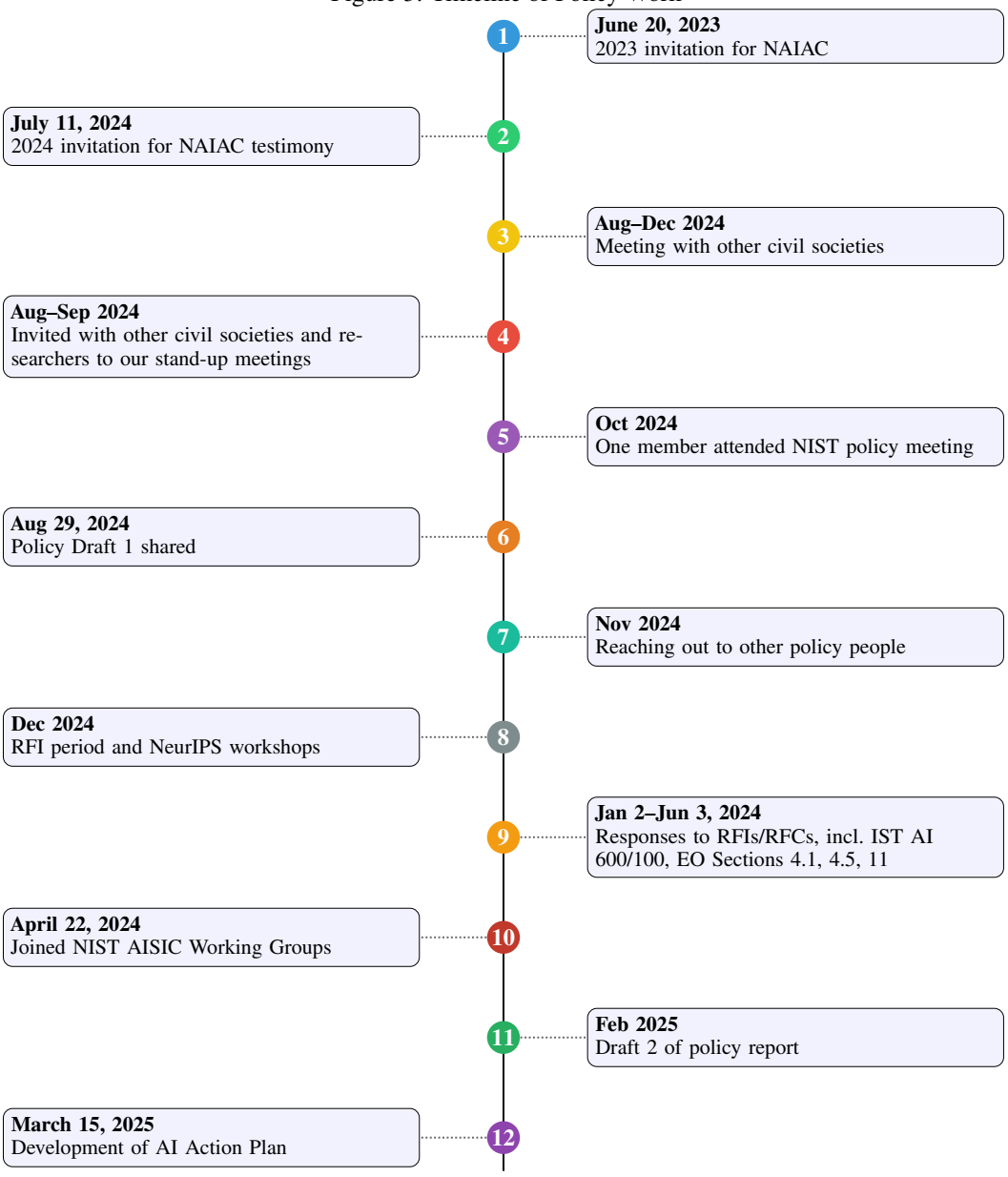
Figure 2: Brief overview of U.S. AI Governance efforts between 2021 and 2025.



B Methodology

Our organization's policy work is young in that our involvement in executive-level policy work came after the introduction of Biden-era executive orders on preventing discrimination on the basis of gender identity or sexual orientation (71); safe, secure, and trustworthy AI (72); and resultant creation of the NIST AI Safety Institute (NIST AISI) and Consortium (AISIC). Biden Executive Order (EO) 13988, was introduced in early 2021 and instructed executive agencies to extend Title VII of the 1964 Civil Rights Act to include sexual orientation and gender identity as protected classes (71). This had far reaching effects for anyone accepting federal funding and was also accompanied by a toolkit to codify equality efforts for transgender and LGBTQIA+ individuals (75). Biden EO 14110 was introduced in late 2023 to start AI governance conversations and announce long-term goals for the administration (72). The scope of the EO was noticeably wide and resulted in the creation of two groups, the NIST AI Safety Institute (NIST AISI) and AI Safety Institute Consortium (AISIC), with mandates to produce policy guidance for executive agencies and those receiving executive-level funding. Biden EO 14110 was introduced after the AI Bill of Rights (79) by the Office of Science and Technology Policy (OSTP), which broadly defined individual "rights" of Americans to be protected when using AI. These documents show a tension in the previous U.S. administration between rights-based and risk-management approaches.

Figure 3: Timeline of Policy Work



B.1 National AI Advisory Committee (NAIAC) Briefings

In 2023 and 2024, our organization was invited by members of the National AI Advisory Committee (NAIAC) to give testimony on harms caused by AI systems to queer people and related topics (e.g., company data use policies, scope of pre- and post-deployment evaluation). Our group's initial 2023 testimony was given with a group of similarly focused research-based, AI affinity organizations and civil-society organizations, ranging in AI expertise. The 2023 session allowed each group two minutes for an initial statement and limited discussion after. The 2024 session allowed five minutes for an initial statement and a longer discussion. Discussion in both sessions were limited to members of the NAIAC and invited organization, and NAIAC committee attendance appeared optional. The NAIAC reported to the Secretary of Commerce, and some of our organization's appeals do appear in the year 1 and 2 reports (17; 18), as well as resultant recommendations to NIST RMF. Of note is that the NAIAC's power was limited in scope and largely focused on supporting the NIST RMF and AI implementation in executive agencies. Much of the communication structures described above appeared to our organization's members to hinder open discussion of policy, but can also be described as policy norms unknown to the group.

In crafting our 2023 and 2024 testimonies to NAIAC, our organization gathered initial stances internally and from the ethics research community. These initial stances largely considered AI harms towards queer people, as detailed within the academic community, and did little to engage the broader queer community. The compilation of this research work was done by members in our organization, who often identified as members of queer and other marginalized communities. This heavily dictated our organization's stance towards intersectional approaches that situate civil rights. There were also discussions with researchers-turned policy advisors in efforts to further validate these stances and better communicate their work to the NAIAC. As the only queer-focused group to provide a testimony to the NAIAC, the scope of our testimonies often did not show a clear plan for describing the needs of queer people in the U.S. The practical reason for this is that doing so exceeded the scope of what our volunteer policy team could accomplish. We believe that more concrete plans for our organization to facilitate larger engagement should have been considered. Finally, the NAIAC briefings are delivered to an audience composed primarily of academics, industry, and civil society. The group appeared to have many members coming from policy backgrounds. As our expertise largely was not in policy, we needed to shift our language to match our audience. Often, this meant translating our research in the context of agency-specific applications and cognizable claims.

Our organization's policy stances in these two briefs remained similar, with some progress made in the interim. In both 2023 and 2024, our briefs highlighted that AI models produced queer-focused hate speech, misrepresented and misgendered gender minorities, and was currently being used to give legitimacy to ill-defined tasks (e.g., gender and sexual orientation prediction (40; 78)). We asked for increased funding for queer scientists, more accurate statistics of the number of queer-identifying scientists, stronger consent practices before using AI on personal information, increased training for those using AI, and increased engagement with other queer-focused organizations. Some of our calls were answered as executive agencies began collecting statistics on queer scientists to facilitate research on systemic discrimination in science (38).

B.2 NIST AI Safety Institute Consortium

Our organization was also invited to join the NIST AI Safety Institute Consortium (AISIC), a group with members from academia, industry, government, and civil society, contributing to working groups on practical AI research expansion, guidelines creation, robust design and evaluation framework design, and other topics. We joined AISIC to implement the calls in our 2023 and 2024 NAIAC testimonies. We held many internal meetings to gather initial thoughts and develop strategies focused on advancing rights-based governance strategies. As the only queer-focused organization, the group decided to be vocal about queer-specific problems such as forced "outing" of queer people in unsafe or malicious contexts, misgendering, consent, and excessive or insufficient data collection and use. To further amplify our points, our organization planned to partner with more established policy groups and large queer-focused organizations outside of AI and technology spaces. There were discussions on the disconnect between the classification of harms in ethics research (65) and in the NIST RMF (49), which chose to focus on issues such as bio-warfare and not prioritize model bias and discrimination against queer people. There were also discussions on the general focus towards harms perpetuated by generative AI design and deployment, and a lesser focus on harms from "narrow AI"

with specific use cases. More broadly, the dominant conversations with NIST AISIC focused on a version of AI safety that did not fully reflect the range of harms known by the queer and AI ethics communities.

This realization was reflected in our organization's responses to NIST in formal meetings and in responses to requests for comments or information. We felt we engaged consistently and critically, but without signals to gauge whether our concerns were being considered. This can be viewed as a shortcoming of existing policy communication structures with executive agencies, of which our organization may have been unaware of. Our organization met with AISIC leadership and discovered that, while their participation had been consistent, our comments had not necessarily been read by current NIST staff. Our organization noted that our involvement with NIST felt "opaque" at many points. We did however continue to submit responses to requests for information.

B.3 Meeting with Other Civil Society Groups

In policy conversations, our organization sought to employ learnings from our existing work in community engagement and inclusive AI research. In crafting policy directions, we had discussions with long-established civil society organizations and technology policy experts. We detail some of the directions from these conversations here.

While our organization was the only queer-focused group in AISIC, there were other AI ethics, civil society groups, and similarly-focused academics in AISIC. Our organization also reached out to other civil society and ethics-focused groups active in AI ethics and sociotechnical policy spaces. Our organization was able to meet with some of them and participate in regular group meetings. We note that the members in this coalition maintained regular contact with NIST AISI leadership, AISIC working group activities, and NIST AISI requests for information and comment. Coalition building with civil society organizations in AISIC allowed us to consolidate policy asks. Our meetings were multi-purpose, serving as a place to vent and openly discuss problems each group was facing and actively connect thoughts that each group had into shared stances to communicate in our responses to RFCs. Retrospectively, a third theme can be seen in this coalition's desire to create a "counter-power" – working to shift the power dynamics within AISIC and prioritize sociotechnical approaches in AI policy. There was also a focus in this coalition on advocating for transparency from the NIST AISI to combat frustrations many in the group had communicating with AISI leadership.

The makeup of the group was interdisciplinary, with a mix of policy and research-focused people in each meeting. A sort of "language" translation resulted from our organization shifting its focus towards sharing why sociotechnical approaches should be considered in NIST AISIC work. The coalition pointed to the large body of research on the positive effect of sociotechnical approaches in design, deployment, and evaluation of AI models (36; 61). The coalition meetings also inspired our organization's work on a policy report.

B.4 Policy Report

Due to the lack of influence we were able to exert through established channels like RFIs, our organization decided to draft a policy report to reach and inform lawmakers sympathetic to our cause. We highlight that the decision to publish a policy report involved many participatory-minded frameworks throughout. Ideally, we would want direct community involvement in policy creation; a core complication arises in how to capture differences along varying geopolitical and social contexts.

We released the first version of the policy report in August 2024. The policy positions presented in the report were inspired by our 2023 and 2024 briefings to the National AI Advisory Council (NAIAC), recent responses by our organization to federal information requests, and research on participatory governance, fairness, ethics, and privacy (58; 59). A common frustration by our organization and civil society organizations at the time was that sociotechnical and civil concerns were very narrowly considered in the NIST Risk Management Framework (RMF). We detailed the importance of sociotechnical approaches, providing concrete examples of queer harms, and sharing existing sociotechnical considerations in research and policy. Our policy report outlined the risks that algorithms, data, and AI systems pose to queer people and other marginalized groups, highlighting both historical and ongoing harms such as discrimination, violence, and erasure, advocating for a rights-based approach to AI development and emphasizing the need for diverse participation and oversight throughout the AI lifecycle. In addition, we called for greater transparency, consent, and pri-

vacy protections, as well as the restriction of pseudo-scientific uses like gender recognition, stressing the importance of including queer perspectives in AI design, limiting AI's role in law enforcement, and ensuring strict audits for existing AI systems. The document sought to protect marginalized communities by promoting autonomy, safety, and inclusive policy interventions. Specifically, our policy stances were the following:

Support algorithmic and dataset transparency. Our organization calls for some form of algorithmic and dataset transparency in all application, noting its ability to advance fairness towards marginalized groups through open collaboration between the public and AI developers.

Support consent, privacy, and data rights over data collection. Privacy needs can vary within marginalized communities. A core concern in the queer community is the "doxing" or outing of queer individuals. Policy provisions that allow for informed consent over automatic opt-in encourages autonomy. Transparent design and deployment practices can create trust.

Restrict use of AI for pseudo-scientific purposes. Queer people have been historically harmed by ill-designed and pseudo-scientific AI models (53; 67; 63; 15). Queer people are also disproportionately surveilled in efforts to ostensibly increase efficiency or empirical objectivity. As an example, Facebook implemented a real name policy in 2014 which adversely affected trans individuals and drag queens to be blocked. We note that these applications threaten the civil rights of queer people. Biometrics may also be used for ill-defined purposes. Existing regulation prohibits discriminatory biometric systems in the EU (i.e., Article 5 of the EU AI Act) but not in all jurisdictions in the U.S. (26).

Limit AI use within law enforcement. Require strict third-party audits where AI is already in use. Our organization takes a stance against AI use by law enforcement and immigration authorities. The use of AI in these spaces vary greatly and can lead to discriminatory outcomes, further opaqueness in policing, and weakened public trust (48).

Include queer perspectives when designing and deploying AI systems, protect the identities of these groups when applicable. Increased engagement from queer and marginalized groups at all stages of the AI lifecycle allow existing concerns to be addressed before deployment (5). They allow for more robust evaluation of AI systems by developers and can temper expectations of end-users.

The report was critiqued internally by our organization's policy team ($n = 77$), then made available to the entire organization's community ($n = 1518$). Finally, it was released publicly on our website (57).

B.5 NeurIPS Workshop

To engage more queer people in the development of our policy report, a public information request period was advertised in December 2024 and focus groups were held during a NeurIPS[7] workshop the same month. First, a panel discussion was held with members of the AI governance and research community to discuss what practical participatory governance efforts should be considered. The panel was a mix of policy and AI researchers, practitioners, and lawyers, with the goal of giving the audience an understanding of active policy participation efforts at the corporate and government levels in the U.S., Canada, Italy, and Brazil. This panel discussion also served as a primer to describe current AI governance efforts internationally and the approaches being considered.

Workshop participants came from many backgrounds and varying levels of AI expertise and participated in two focus groups aimed at measuring alignment with positions in the document and areas for improvement. The first working group session focused on the existing policy report by our organization. The second working group session was more open-ended, inviting participants to envision what rights, rules or regulations in regards to data and AI would meet their needs. Each working group session followed Chatham House Rules stating that information shared is free to share publicly but may not be attributed to a particular person. This was done to protect sensitive information around sexual orientation and gender identity, and allow workshop members to share independent of any respective organization they might belong to. Workshop organizers described

[7] `https://neurips.cc/`

the sections of the policy report and motivating reasons behind each policy stance. The audience members were then divided into groups with policy report authors to read and discuss each stance .

An early critique during these discussions was the heavy U.S. focus of this document in a geographically diverse community. The interplay between technology governance and civil protections for queer people varied greatly across participants' home countries, where queer people could face legal, safety, and societal challenges. At the same time, participants emphasized elevating the voices of queer folks directly harmed by AI technologies, as well as the need for community building in AI development.

C Participatory Design (PD) and Governance

The idea of community participation in governance is fundamental to the democratic process. Participation in governance does not apply exclusively to governmental organizations but rather to any system requiring governance (e.g., online forums, companies, municipalities, etc.). In this work, AI governance refers generally to policy creation at any level, private corporations and governments alike. An organization's structural characteristics also influence practical participation. For instance, our members identify as artists, practitioners, researchers, lawyers, and ethicists. Our involvement, along with that of other diverse groups, signals a shifting policy approach to AI governance (6; 12). These conversations are numerous and break norms of policy advocacy in a central geographic location or by traditional advocacy groups.

History and Definitions. PD can be traced back to the 1970s workplace democracy movement in Scandinavian countries, with the fundamental *political* ideal of emancipation by shifting and reclaiming power by those who lack it (11, §2) — a constant refrain as PD enters into its fourth era now with special emphasis on feminist, queer, indigenous and/or decolonial frame, among others. Reflecting on the history of PD, (39) proposed six guiding principles: *equalising power relations*, *democratic practices*, *situation-based actions*, *mutual learning*, *tools and techniques*, and *alternative visions*. Similarly, according to (10), PD has the following four strong commitments (§2.1, p. 7): (a) *democracy* at the workplace and beyond; (b) *empowerment* of people through the *processes of design*; (c) *emancipatory practices* rooted in mutual learning between designers and people; and (d) seeing humans as *skillful* and *resourceful* in the development of their future practices.

Scaling. While PD seems to have a reputation of mostly being conducted in small-scale projects, since its inception, PD has been envisaged across a multiplicity of scales, up to multi-year, multinational settings. In this context, the concept of "scaling" PD, although often seen as a desideratum, is not always explicitly and precisely defined. Recently, (7) identified three aspects of scaling in PD: socio-spatial, temporal, and – most notably – onto-epistemological. While the first two aspects are no doubt important for PD projects in information technology (IT), it is the onto-epistemological, inspired by Arturo Escobar's conception of our shared world as a *pluriverse* — "a world where many worlds fit" (25, p. xvi) — that has proven most relevant and fruitful to our policy efforts. In this context, onto-epistemological aims at incorporating diverse lived experiences from global queer communities and empowering individuals and communities that have historically been oppressed. Meanwhile, (80) represents the most recent comprehensive survey on scaling PD, revealing the tensions of maintaining PD's democratic and emancipatory ideals vis-à-vis increased scale, especially for projects involving distributed IT systems.

Infrastructure The scaling(-up) of PD crucially depends both on the technological infrastructure, such as software and online platforms, and on social infrastructure such as communities, collectives, and relationality (35). Scholars have spoken of "infrastructuring" as the ongoing process of building and adapting infrastructures for sustainable large-scale PD projects. An important account was given in (9) where the notion of a spectrum between relatively stable *networks* and more fluid and transient *knotworks* was mobilized to illustrate the importance of backstage activities in participatory infrastructuring.