xCoAx 2019 Conference on Computation, Communication, Aesthetics & X Milan, Italy

Sarah Ciston
ciston@usc.edu

University of Southern California

Imagining Intersectional AI

Keywords
artificial intelligence
intersectionality
gender
race
sexuality
feminism
algorithms
ethics
data
methodologies
experimental practice
computational media
creativity
community

Artificial intelligence should be created and used with more critical thought to the biases and ideologies baked in. This paper surveys current research contextualizing the stakes of AI discrimination and looks to intersectionality as a set of overlapping frameworks applicable to AI for both analysis and tactics. It argues that intersectional approaches need to be implemented widely, in community, and throughout the entire AI pipeline—from development and implementation to cultural absorption and material impacts. Although that vision is far from being fully realized, this paper points to examples suggesting how experimental engagements with AI can help imagine its intersectional futures.

INTRODUCTION

Artificial intelligence is quietly shaping social structures and private lives. Although it promises parity and efficiency, its computational processes mirror biases of existing power structures even as often-proprietary data practices and cultural perceptions of computational magic obscure those influences. However, intersectionality—which foregrounds an analysis of institutional power and incorporates queer, feminist, and critical race theories—can help to rethink artificial intelligence. An intersectional framework can be used to analyze the biases and problems built into existing artificial intelligence, as well as to uncover alternative ethics from its counter-histories.

While offering tools for critical analysis of existing technologies, intersectionality can also shift approaches to creating new technologies. This paper, after contextualizing current arguments around AI bias and intersectionality, examines strategies from Black feminist, mixed race, and queer communities to show how these might be applied to algorithm design and implementation in culture. It argues that AI should be created and critiqued with an awareness of power – and reframed using intersectionality – to value multiple epistemologies, methodologies, and perspectives in order to address the social inequalities it reinforces. Finally, it uses case studies to sketch out a preliminary proposal for imagining intersectional artificial intelligence that can disrupt hegemonic structures and uncover its subversive procedural potentials.

AI AS AUTHORITY, AI AS MIRROR?

Bias manifested through the use of algorithms and artificial intelligence carries high stakes, in part because it is has the potential to put some populations at much higher risk. Safiya Noble (2018) argues "algorithmic oppression" is hard-coded into the algorithms that support AI and the very systems that determine much of day-to-day subsistence, "creating and normalizing structural and systemic isolation, [...reinforcing] oppressive social and economic relations" (10). Miriam E. Sweeney (2016) points out that in the design of artificial agents, "the normative subject is usually constructed as White, male, and presumptively heterosexual, and therefore unproblematic and uncomplicated as a design option. Female and non-White identities are seen as potentially problematic" (222). What gets framed as a matter of preference is linked to a system in which whiteness holds more value (Niesen 2016, 171). Despite the democratizing promise of digital technologies, identity markers are reinforced and even extracted as capital through (in)voluntary participation in algorithmic systems. In one of the more troubling cases of algorithmic bias, *ProPublica* analyzed the proprietary algorithms used to create criminal risk assessments – predictions which affect the harshness of sentencing:

> The score proved remarkably unreliable in forecasting violent crime: Only 20 percent of the people predicted to commit violent crimes actually went on to do so. [...] The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants. (Angwin et al 2016)

This example is among many “well-documented biases that should not have been news” and that precede digital technologies (Chun 2018, 64) but whose biases are further disguised by algorithmic systems.

These systems have incredible impact not only because they operate en mass, outside individual control, but also because AI’s mythology allows them to assume the status of impartial fact even as they operate by human interpretation and intervention at every level. The embodied human—while missing from the marketing of AI as superior for its detached, emotionless decision making—remains key to AI’s operation: “a human interlocutor is needed to keep artificial systems functioning effectively. All AI is HCI [(human–computer interaction)]” (Wilson 2010, 103). This labor is horrifyingly evident in the field of commercial content moderation (CCM), which Sarah T. Roberts (2016) argues lends a dangerous sense of naturalization to racist and biased content because of the perceived computational role of AI rather than human choice: “Companies’ desire to keep CCM work in the shadows therefore gives the impression that such content is just what is out there in the culture in some kind of natural, organic way and hides the human decision-making processes and curation work from the view of their user-participants” (157). Here AI “autonomy” is carefully curated for corporate profit, but contingent on human systems—who are acting as technology and who are ignored and exploited on its behalf.

That curated cultural understanding of AI overlaps with the big data it utilizes and with specific types of AI like deep learning and neural networks, and so it is important to distinguish that this discussion focuses on “narrow” AI, which is various in kind and is implemented toward specific objectives such as purchase recommendations, news feeds, etc.—not “general” or “strong” AI, which remains speculative and conjures dystopian fears of the singularity. The conflation of the various types of artificial intelligence contributes to its mystique, allowing it to operate with an aura of unquestioned truthiness. This is compounded, claims Luciana Parisi (2017), by the structures of binary problem-solving, “which values making a clear decision quickly more than it does making the correct one” (1). These fuzzy understandings do little to undo the infrastructural inequalities embedded in the design and implementation of many AI.

INTERSECTIONALITY IS NOT JUST MORE REPRESENTATION, NOT JUST MORE DATA

Reading artificial intelligence through an intersectional lens can help decode and critique these power structures, and using intersectional approaches to design and implement artificial intelligent systems creates the possibility for restructuring them. Intersectionality is not merely shorthand for discussing individual identity representation by accumulating strings of hyphenates; rather, it examines and critiques systems of power and how those systems structure themselves to impact groups and individuals unequally. Brittney Cooper (2016) re-articulates the definition of intersectionality to distinguish it from how it has been oversimplified and misused:

> Intersectionality's most powerful argument is not that the articulation of new identities in and of itself disrupts power arrangements. Rather, the argument is that institutional power arrangements, rooted as they are in relations of domination and subordination, confound and constrict the life possibilities of those who already live at the intersection of certain identity categories, even as they elevate the possibilities of those living at more legible (and privileged) points of intersection. (10)

It is not enough to add more data to the neural network or to represent additional identities—these too will be opportunities for marketing: "the fracturing of users based on identity categories is, in fact, a key mechanism of capital to provide such data to advertisers" (Niesen 2016, 168). As Wendy Hui Kyong Chun (2018) points out, even feminist intersectional theory, when misread as a means of sifting data for sameness through identity difference, can be misappropriated toward racist ends (65). But when used instead to consider institutional structures that feed those data, Noble (2018) argues intersectional readings of technology are essential: "a feminist lens, coupled with racial awareness about the intersectional aspects of identity, offers new ground and interpretations for understanding the implications of such problematic positions about the benign instrumentality of technologies" (31).

Because intersectional theory owes its roots to Black feminist thought, the epistemologies and strategies employed by women of color are at the core of an intersectional critical praxis. Noble, Brendesha M. Tynes, and Joshua Schuschke (2016) argue that the queer women of color who founded Black Lives Matter offer a model for coalition-building through skills honed in community: "the movement's reflexivity, the ability to counter hegemonic narratives, and self-care are key components of digital intersectionality. By modeling the standard of reflexivity, the movement is able to critique and correct its own narrative and practices" (28). Reflexivity, self-care, counter narratives, coalition building, and other Black feminist methods could be incorporated into intersectional AI at the development, implementation, analysis, or data-gathering stages—and these methods could work to destabilize existing standards and biases.

Strategies from mixed race, trans, bisexual, and femme communities—whose identities are not easily categorized, who sometimes maneuver by passing within systems—may also be used to engage and subvert normative algorithmic practices in order to operate on multiple valences of infrastructural power and intersectional disenfranchisement. As Myra S. Washington (2017) argues, "in cutting across categories, transracialness is about the ‘potential mutual transformation’ of those categories [...,] how people position themselves and move within this spectrum of power and is not so much about identity” (14–15). Because queer theory troubles “assumptions about the natural unity of the category ‘women’” (Cipolla et al. 2017, 7), reading artificial intelligence through intersectional queer theory can also push back on assumptions in AI about gender, while using queer strategies to disorient those categories can push back on assumptions about technologies themselves. Geographers Daniel G. Cockayne and Lizzie Richardson (2017) read queer theory through software studies because “queer approaches are invested in conceptualising and (therefore) challenging both social and digital code(s)—or the norm—to show how they constrain normativity but also how forms of intimate life can transgress, disrupt, and distribute what is normal” (1643). Queer-of-color activists reclaim less-visible identities as sites of strength. They redefine femme to make it legible and instrumental for their communities: “not just being about blonde girls wearing pink, but about the big deal about being fierce women of color or down white girls who are hot strong girls who are political who see the connection between everything in our lives” (Mahmood 2008, 4). Such viewpoints can inspire intersectional connections and possibilities for AI that challenge how technologies are both connecting and othering individuals. They help frame how intersectional AI might instrumentalize its precarious orientations, reworking stereotypes of passing and instability to reprogram technologies of gender and agency.

TOWARD INTERSECTIONAL AI STRATEGIES

Speculating a more intersectional techno-ideological imaginary, Kara Keeling (2014) proposes the Queer OS: “to make queer into the logic of ‘an operating system of a larger order’ that unsettles the common senses that secure those presently hegemonic social relations [...but it] acknowledges its own imbrication with and reliance on those logics while still striving to forge new relationships and connections” (154). Keeling calls it a “malfunction with a capacity to reorder things” (157), which moves away from the urge to read neutrality and rationality into algorithms. Wilson (2010) also asks how malfunction might contribute to more advanced artificial agents: “Is error (and its affective corollaries: shame and anger and contempt) the limit of an artificial system, or might error be part of its internal coherence? Might there be artificial systems that can tolerate their own inadequacies?” (57). Parisi (2017) suggests that machine learning can be read as a new form of knowing: “reasoning through and with uncertainty” (8). Her critique rejects techno-utopias that privilege Western empiricism; instead, she strives to reshape reason itself through experimentation with artificial intelligence (9).

Applying intersectional tactics to artificial intelligence could offer material impacts, but those may be difficult to trace without approaches like Critical Technocultural Discourse Analysis, an intersectional research method designed by André Brock that “recommends the analytical integration of the technological artifact and user discourse, framed by cultural theory, to unpack semiotic and material connections between form, function, belief, and meaning" (Sweeney and Brock 2014, 3). This two-pronged approach is designed to "jointly interrogate culture and technology" (1). The combination is essential to understand the imbricated impacts of artificial intelligence, intersectional or otherwise. Kate Crawford (2016) echoes this need, saying: "We would go further than simply analyzing the design of the algorithm and pay close attention to shifts in power, from programmers to the algorithms themselves to the wider network of social and material relations" (82–83).

Research methods to trace impacts of artificial intelligence can analyze the structures themselves, not only their inputs. Catherine Griffiths (2018) proposes "computational visualization" that close reads source code as well as using visualization as a critical inquiry into AI: "A key component of this method is its focus on process, both temporally and spatially, in which data is parsed, forked, and on which decisions are executed" (220). This method uses synthetic data to isolate the structure of the algorithm from the original data "to understand whether the data structure or algorithmic process can also reveal discrimination, either alone or by means of augmenting a latent bias [...]. Does bias lie solely in the data, as frequently stated, or can it also lie in the structure of the classifier, and perhaps in the process that couples those together" (224). Griffiths argues that computational visualization can help to determine whether biases exists in a dataset, in the algorithm, or the processes that combine them.

In another example of how to employ visualization techniques toward intersectional ends, Catherine D'Ignazio and Lauren F. Klein argue that "data, design, and community of use, are inextricably intertwined" (2). They propose principles for feminist data visualization—"rethink binaries," "embrace pluralism," "examine power and aspire to empowerment," "consider context," "legitimize embodiment and affect," and "make labor visible" (2–3)—that could be adopted to design intersectional artificial intelligence. Crawford (2016) argues for designing AI with a logic of agonistic pluralism, which would emphasize how "algorithmic decision making is always a contest, one that is choosing from often counterposed perspectives, within a wider sociotechnical field where irrationality, passion, and emotion are expected" (87). She uses Carl DiSalvo's concept of "adversarial design" to understand algorithms by beginning "with the premise of ongoing struggle between different groups and structures—recognizing that complex, shifting negotiations are occurring between people, algorithms, and institutions, always acting in relation to each other" (82–83). DiSalvo (2012) identifies three primary tactics of adversarial design: "revealing hegemony, reconfiguring the remainder, and articulating collectives" (26). He sees these tactics as both research and practice: "Through the process of making contestational objects, adversarial design is a kind of inquiry into the political condition" (116). But also: "[They] do more than raise awareness and critique. They instantiate a possibility for another ordering of sociotechnical structures that allows us to act in the world in a different way" (119).

Still, he cautions that design is not automatically revolutionary, instead calling for adversarial design to be a participatory practice, collective and collaborative (124), which aligns with intersectional strategies. In addressing AI bias, it is important to remain mindful of both the empty, shiny promises of design thinking and the cheerful calls for collectivity that ignore intersectional inequality. Lauren Berlant (2016) argues that "the commons" can threaten "to cover over the very complexity of social jockeying and interdependence it responds to" (395), but should instead point to what is broken and "the difficulty of convening a world conjointly" (395). Imagining intersectional artificial intelligence cannot be done from a single subject position. In order to address entrenched and pervasive power structures, this work must happen in multiple communities and take intersecting forms, morphing and subverting, with no singular ethic or aesthetic, rather a meta-ethics of multiplicity and intersubjective relation.

COMMUNITY METHODOLOGIES, ARTISTIC EXPERIMENTS, FUTURE RESEARCH

Several case studies offer inspiration for imagining intersectional AI to come. MIT Media Lab's Joy Buolamwini calls algorithmic bias the "coded gaze," and she founded the Algorithmic Justice League to create interventions from poems to corporate pledges to code modules. Her poem "AI, Ain't I a Woman" asks why AI systems are not trained to see Black historical figures as female: "Can machines ever see my queens as I view them? Can machines ever see our grandmothers as we knew them?" (Buolamwini 2018). Other projects like "Aspire Mirror" and "Safe Face Pledge" also address computer vision bias, and Buolamwini (2017) uses each to emphasize that "who codes matters, how we code matters, why we code matters."

Also out of MIT Media Lab, the Data Nutrition Label addresses bias before data enters the algorithms that run artificial-intelligent systems. Holland et al. (2018) have designed a visual aid for assessing problems in potential training datasets: "issues such as surprising variable correlations, missing data, anomalous data distributions, or other factors that could reinforce or perpetuate bias in the dataset. Addressing these factors in the model creation and training phase saves costs, time, and effort, and also could prevent bad outcomes early on, rather than addressing them after the fact" (13). They hope to help those who work with data build better habits by "questioning datasets through analysis and interrogation techniques, even if a particular dataset does not include a Label" (15). While not AI systems themselves, these examples make interventions that draw on intersectional theory and suggest an ethics of relation and care.

But imagining intersectional AI means intervening with practical and speculative approaches simultaneously. One arts-based example, developed by this author, is a suite of absurd "pataphysical" designs—including ladymouth, a chatbot that tries to explain feminism to online misogynists (Ciston 2019a). After the initial prototype that posted quotations from feminist scholars to men's rights subreddits, future versions of ladymouth will use natural language processing and sentiment analysis for more nuanced conversations as well as an interface for contributing feminist responses. I designed the tool to be adaptable to other intersectional issues, and it has inspired a collaborative project that uses its technology to help address diversity labor in STEM workplaces (Billard, Ciston, and Loop 2019). These examples show how a speculative project can spawn additional practical solutions for other audiences by opening a space to ideate and iterate on intersectional possibilities.

Such spaces thrive in community, and the organization Feminist.AI takes a community-driven approach to rethinking what they call "hegemonic AI. Rather than simply criticize the lack of diversity in AI design and development, we propose an alternative by co-designing intelligent products, experiences and futures from a feminist posthumanist (inclusive) approach" (Meinders). The Feminist.AI philosophy cites many intersectional values, including that the group "must be invited to participate," offer "multiple entry points for involvement, so we can pull from different knowledge systems," "revisit every step of our process with every new project," "work to contribute to and community source our own data," and "attribute everything (people who have come before us, original parallel research)."

As another community-based practice, I used intersectional strategies to develop a student organization for media artists to learn programming. USC's Creative Code Collective offers project-based co-learning across arts and computer science disciplines, with workshops on computer vision, text generation, etc. Our ethos emphasizes "scrappy artistic strategies not perfect code; growth not mastery; all practice is theory–practice; and we all have skills to teach each other" (Ciston 2019b). In the collective, I model my approach to working with existing technologies, like OpenAI's GPT-2, through a poetic practice that can draw out critical considerations and aesthetic oddities alike. I argue that artistic experiments with AI are one approach to develop possibilities for more fair and inclusive technologies. Fostering communities in which multiple voices feel valued and free to experiment is essential.

Whether designing new AI tools, examining and experimenting with existing tools in unexpected ways, or supporting the practice of others in community—utilizing intersectional aesthetics, ethics, and tactics to imagine artificial intelligence can potentially reveal and remake the structures that have reinforced heteronormative patriarchal white supremacy and rendered its power invisible and rational.

REFERENCES

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner.
2016. "Machine Bias." *ProPublica*, May. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Billard, T.J., Sarah Ciston, and Anna Loup.
2019. "chats with *diversibot:* Automating the Emotional Labor of Diversifying STEM." Annenberg Graduate Research Symposium. Los Angeles. Apr 18, 2019.

Berlant, Lauren.
2016. "The Commons: Infrastructures for Troubling Times*." *Environment and Planning D: Society and Space* 34 (3): 393–419.

Buolamwini, Joy.
2017. *How I'm Fighting Bias in Algorithms | Joy Buolamwini.* TEDx BeaconStreet. https://www.youtube.com/watch?v=UG_X_7g63rY.

Buolamwini, Joy.
2018. *AI, Ain't I A Woman? - Joy Buolamwini.* https://www.youtube.com/watch?v=QxuyfWoVV98.

Chun, Wendy Hui Kyong.
2018. "Queerying Homophily." *Pattern Discrimination.* Clemens Apprich, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl. Minneapolis: University of Minnesota Press.

Cipolla, Cyd, Kristina Gupta, and David A. Rubin.
2017. *Queer Feminist Science Studies: A Reader.* Edition. Seattle: University of Washington Press.

Ciston, Sarah.
2019a. "Ladymouth: Anti-Social-Media Art As Research." *Ada: A Journal of Gender, New Media & Technology* 15. https://adanewmedia.org/2019/02/issue15-ciston/.

Ciston, Sarah.
2019b. "Creative Code Collective Introduction." University of Southern California. Los Angeles. Jan 18, 2019.

Cockayne, Daniel G., and Lizzie Richardson.
2017. "Queering Code/Space: The Co-Production of Socio-Sexual Codes and Digital Technologies." *Gender, Place & Culture* 24 (11): 1642–58.

Cooper, Brittney.
2016. "Intersectionality." In *The Oxford Handbook of Feminist Theory*, edited by Lisa Disch and Mary Hawkesworth. Vol. 1. Oxford University Press.

Crawford, Kate.
2016. "Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics." *Science, Technology, & Human Values* 41 (1): 77–92.

D'Ignazio, Catherine, and Lauren Klein.
n.d. "Feminist Data Visualization." Accessed September 26, 2018. https://www.academia.edu/28173807/Feminist_Data_Visualization.

DiSalvo, Carl.
2012. *Adversarial Design*. Edited by Ken Friedman and Erik Stolterman. Cambridge: The MIT Press.

Griffiths, Catherine.
2018. "Visual Tactics Toward An Ethical Debugging." *Digital Culture & Society* 4 (1): 217–226.

Holland, Sarah, Ahmed Hosny, Sarah Newman, Joshua Joseph, and Kasia Chmielinski.
2018. "The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards." ArXiv:1805.03677 [Cs], May. http://arxiv.org/abs/1805.03677.

Keeling, Kara.
2014. "Queer OS."
Cinema Journal 53 (2): 152–57.

Mahmood, Zuleikha, and Leah Lakshmi Piepzna-Samarasinha.
2008. "All Our Holes Are Hungry: Hungry for Justice and Fucking." *Femme Shark Communique*, no. 1. http://archive.qzap.org/index.php/Detail/Object/Show/object_id/432.

Meinders, Christine.
Feminist.AI. n.d. Accessed Apr 27, 2019. https://www.feminist.ai/

Noble, Safiya Umoja.
2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York: NYU Press.

Niesen, Molly.
2016. "Love, Inc.: Toward Structural Intersectional Analysis of Online Dating Sites and Applications." In *The Intersectional Internet: Race, Sex, Class and Culture Online,* 161–78. Digital Formations, Vol. 105. New York: Peter Lang Publishing, Inc.

Parisi, Luciana.
2017. "Reprogramming Decisionism." *E-Flux Journal* #85 (October).

Roberts, Sarah T.
2016. "Commercial Content Moderation: Digital Laborers' Dirty Work." In *The Intersectional Internet: Race, Sex, Class and Culture Online*, 147–60. Digital Formations, Vol. 105. New York: Peter Lang Publishing, Inc.

Sweeney, Miriam E.
2016. "The Intersectional Interface." In *The Intersectional Internet: Race, Sex, Class and Culture Online*, 215–28. Digital Formations, Vol. 105. New York: Peter Lang Publishing, Inc.

Sweeney, Miriam E , and André Brock.
2014. "Critical Informatics: New Methods and Practices." Proceedings of the American Society for Information Science and Technology 51 (1): 1–8.

Tynes, Brendesha M., Joshua Schuschke, and Safiya Umoja Noble.
2016. "Digital Intersectionality Theory and the #Blacklivesmatter Movement." In *The Intersectional Internet: Race, Sex, Class and Culture Online*, 21–40. Digital Formations, Vol. 105. New York: Peter Lang Publishing, Inc.

Washington, Myra S.
2017. Blasian Invasion: Racial Mixing in the *Celebrity Industrial Complex. Race, Rhetoric, and Media Series*. Jackson: University Press of Mississippi.

Wilson, Elizabeth A.
2010. *Affect and Artificial Intelligence*. Seattle: University of Washington Press.