Applying Reflexivity to Artificial Intelligence for Researching Marginalized Communities and Real-World Problems

Keywords: Social Work, Artificial Intelligence, Reflexivity, Privacy, Accountability, Fairness

Nathan Aguilar, Columbia University, New York City, USA, nja2141@columbia.edu

Aviv Y. Landau, University of Pennsylvania, Philadelphia, USA, landauau@upenn.edu

Siva Mathiyazhagan, University of Pennsylvania, Philadelphia, USA, sivam@upenn.edu

Alex Auyeung, Columbia University, New York City, USA, aa4502@columbia.edu

Sarah Dillard, Columbia University, New York City, USA, sd2716@columbia.edu

Desmond U. Patton, University of Pennsylvania, Philadelphia, USA, dupatton@upenn.edu

Abstract

Despite advances in artificial intelligence (AI), ethical principles have been overlooked, harming marginalized communities. These flaws are due to a lack of critical insight into the complex positionality of the researcher, power dynamics between scholars and the communities being studied, and the structural impact on real-world problems when AI systems appear to be accurate but ethically fail. Reflexivity is a process that yields a better understanding of community-specific nuances, areas requiring local expertise, and the potential consequences of scholastic interventions for real-world problems (i.e., social, environmental, or socioeconomic). The paper builds on the five stages of social work reflexivity that can be applied to AI researchers and provided questions that can be asked in order to increase privacy, accountability, and fairness. We discuss the effective implementation of reflexivity in research, detail the stages of social work reflexivity and highlight key questions for AI researchers to ask throughout the research process.

1. Introduction

Artificial intelligence (AI) algorithms widely influence stakeholders' decisions concerning education, employment, citizenship, national security, and health care (Buolamwini & Gebru 2018; Citron & Pasquale 2014; O'neil, 2016; Rodrigues, 2020; Whittlestone et al., 2021). These stakeholders include government entities, social workers, and law enforcement personnel--all of whom hold great influence and power over the lives of the individuals they interact with. This transformation in decision-making requires ongoing critical, ethical, and methodological considerations and the development of ethical principles regarding AI research (Lazer et al. 2020).

Unfortunately, the principles that do seek to promote privacy, accountability, and fairness in AI face serious limitations (Hagendorff, 2020). The lack of transparency and fairness in the development and implementation of AI has disproportionately impacted marginalized communities (Noble, 2018). For example, Black populations have been harmed by risk assessment algorithms in the criminal justice systems that are racially and socioeconomically biased (Angwin et al. 2016), contributing to the disproportionate incarceration rates of young Black men (Patton et al. 2017). In addition, biometric technology's inability to recognize people of color due to a lack of diverse faces in the facial recognition training model lead to discrimination and the reproduction of unequal power relations (Browne, 2015). Furthermore, algorithms that assess healthcare insurance can lead to adverse health outcomes for low-income people of color by preventing them from receiving adequate medical attention (Benjamin 2019b; Obermeyer, Power & Mullainathan, 2019). The examples of AI attempting to address real-world problems such as violence and health show a lack of cultural context regarding systematic and inherent bias in AI development (Gebru, 2019; Noble, 2018; O'neil, 2016). Additional efforts by researchers to address such injustices should include social work ethical principles such as privacy, accountability, and fairness.

Social work is a practice-based profession with transdisciplinary theories and practice to promote human rights and eco-social justice in the real-world. Social work thinking in AI brings reflexivity in the social and environmental context of the real-world in order to practically implement ethical considerations and reduce possible bias and harm to vulnerable communities. The social workers' primary goal is to work with stakeholders to address social problems, challenge social injustice, respect privacy, hold themselves and systems accountable and promote fairness within macro and micro level systems. These ethical values are practiced with people in need and marginalized communities on an

URI: https://hdl.handle.net/10125/102719
978-0-9981331-6-4

everyday basis. These practiced based ethical values and constant connections with marginalized communities would practically address ethical issues in AI research. Moreover, a central framework when social workers engage in direct practice with clients or conducting research is reflexivity. Reflexivity requires the questioning of one's knowledge, attitudes, behavioral practice, thought processes, assumptions, biases, and habits to strive to understand complex roles concerning others (Bolton, 2010). Our research implements computational and social work approaches to study the fields of community violence prevention (Blandfort et al, 2019; Patton et al, 2018), and health disparities (Landau et al, 2022; Landau et al, 2022b) with future projects utilizing a strength based approach to understand online expression of Black joy. As public interest technologists, we engage in reflexivity and collaborate with computer scientists and data scientists in order to utilize qualitative, natural language processing, and computer vision methods to better understand real world problems and support marginalized communities. Reflexivity offers an opportunity for AI researchers to apply a social work approach to get contextual clarity on real-world consequences of the product they are working on and prevent possible harms to vulnerable communities. We propose a reflexivity framework as a new lens through which researchers may process and implement these ethical values. The following section outlines the stages of a social work reflexive framework, provides a case study where reflexivity was needed and proposes foundational questions to be applied throughout each stage of AI research for analysis of complicated, human-centered, real world problems.

2. Positionality

We are a diverse group of social work scholars with years of practice and research experience in different global domains. Our dedication to supporting marginalized communities are guided by our identities that encompass different religious beliefs, racial and ethnic backgrounds, and life experiences. We utilize reflexivity to better understand our own biases and how they may impact our work with clients and research participants. In this paper, we advocate for reflexivity that extends current AI ethical principles. We acknowledge that reflexivity is an ongoing process and one that is challenging especially when researching marginalized communities and real-world problems.

3. Reflexivity

A significant body of literature urges social workers to be reflexive in their work as practitioners and researchers (D'cruz, 2007; Longhofer & Floersch 2012; Probst, 2015; Probst & Berenson 2014). Scholars on reflexivity encourage social workers to be rigorous in their analysis regarding power dynamics and privilege, and also to consult both colleagues and participants to work through their own biases (Finn, 2020; Probst & Berenson 2014; Watts, 2019)

Though a reflexive process is needed to reduce bias, it's not without challenges. For instance, Finlay (2002) warns about social workers and researchers thinking about themselves and engaging too much reflexivity when they should be prioritizing the needs of their clients. Moreover, scholars have described reflexivity as a muddy process and cautioned against assuming that by engaging in reflexivity research will automatically be better, more truthful or more valuable (Brown, 2006; Pillow, 2003; Valandra, 2012). These authors suggest that reflexivity is not a magic bullet and that we do not escape the ramifications of our positionality by discussing it.

Although these challenges have merit, scholars have found that utilizing a reflexive process is essential within the practice of social engagement and promoting client's self-determination (Furlong, 2003). Self-determination is an ethical principle in social work that recognizes the rights and needs of clients to make their own decisions and to identify their own treatment goals (Furlong, 2003). Without reflexivity, social workers may implement their own personal identity (Race, Gender, Caste, Class, Religion, Ethnicity, Ability, Sexuality, Legal Status, Power, Privileges) biases into the treatment plan and thus compromise their client's autonomy, right to self-determination and can cause possible harm. Social work researchers have found reflexivity to be a valuable tool for enhancing the ethics, quality, and results of their studies (Probst & Berenson 2014). Reflexivity can be beneficial within research, where the 'experimental effect' of a cognitive bias can cause scholars to unconsciously influence the interpretation of qualitative and quantitative data (Probst, 2015). A growing number of researchers (e.g., Dodgson 2019; Gilgun, 2008; Palaganas et al. 2017; Reid et al. 2018) stress the incorporation of reflexivity into all stages of research (Schon, 1992). Through this process, researchers may develop an understanding of the bounds and limits of their expertise which can lead to local collaboration, and thus, an in-depth understanding of the most pressing community concerns (Cizek & Uricchio 2019;

Nkonde, 2019). When conducting AI research, it's important to implement reflexivity throughout each stage of the process (i.e. design, data collection, development, deployment, and data usage). Embracing the meaningful participation of key community stakeholders, analyzing outputs as they come through, and constantly questioning where knowledge is coming from are all critical parts of reflexivity. It is an ongoing and never-ending learning process where the researcher is constantly expanding their awareness of Power, Race, Oppression and Privilege (PROP) issues and how those differences can be perpetuated through AI. For example, in AI deployment, algorithms can unintentionally produce biased results if the data being used skews towards one population compared to another (Raji & Buolamwini, 2019).

It is imperative that researchers utilize a reflexivity framework in order to facilitate more reliable communication between social scientists and individuals working in AI. These efforts can enhance transparency and accountability in the development and implementation of AI systems (Gebru et al. 2018). The following section will discuss implementation of reflexivity and how it can prioritize the privacy of research participants' identifying information, increase accountability by engaging with community residents about local issues to reduce unintended consequences, and promote fairness by including community experts in intervention efforts.

3.1 Social Work Ethical Principles: The Implementation of Reflexivity

Social work ethical principles of human rights, anti-oppression, respect for diversity, privacy, and safety (NASW, 2017) should be integrated as a framing guide for the development of AI. Reflexivity, which is rooted in the aforementioned social work principles, is an appropriate heuristic tool that can be applied in AI research to shift the focus from efficiency to participants' mental and physical wellbeing. We have identified three central ethical principles critical in the field of both data science research and social work practice: privacy, accountability and fairness (NASW, 2017; Floridi et al. 2020; Jobin, Ienca & Vayena, 2019; Wiens et al. 2019). Contemporary researchers who are advocating for equity in the creation and implementation of AI echo these same principles (Benjamin, 2019b; Brock, 2015; Costanza-Chock, 2018; Eubanks, 2018; Noble, 2018; Patton, 2020). For example, In Automating Inequality, Virginia Eubanks (2018) emphasizes that researchers need to have respect for people's privacy, identities and histories as well as prioritize equity in the creation of predictive AI models. This is because there is no standard legal/professional guiding process and models for AI researchers to follow. In addition, technologies are not trained in socio-economic, political, and environmental contextualization. There is a significant social knowledge gap in the AI product development process. Reflexivity offers an opportunity for AI researchers to apply a social work approach to get contextual clarity on real-world consequences of the product they are working on and prevent possible harms to vulnerable communities. We propose a reflexivity framework as a new lens through which researchers may process and implement these ethical values. The following section outlines the necessity of a reflexive framework and provides foundational questions to be applied throughout each stage of data science research for analysis of complicated, human-centered, real world problems.

3.1.1 Privacy

Privacy and confidentiality are one of the main ethical principles social workers must employ when working with clients (NASW, 2017). Social workers are expected to protect the privacy and confidentiality of their clients during and after their professional relationship has concluded. As a result, to employ the social work value of privacy and confidentiality ethical AI research must protect human subjects' privacy and dignity by ensuring autonomy, freedom, confidentiality, informed consent, and nonmaleficence (Floridi et al. 2020; Franzke et al. 2020). Similarly, ethical AI prioritizes privacy as both a value to uphold and as a right to be protected (Jobin, Ienca & Vayena, 2019).

Researchers have the power to choose what details and methods of dissemination they will use even in situations where participants can determine what to share (Ben-Ari & Enosh 2013; Reid et al. 2018). Similar power dynamics parallel those between social workers and their clients, with social workers required to obtain consent by speaking and reflecting with clients about the confidential information they want to share with third parties such as lawyers. Although AI researchers may not be able to obtain consent from those whose data they have, decisions about what information to make public should involve similar reflexive conversations. AI scholars engage with communities that are impacted by their research in order to respect and address their privacy concerns and to uphold human dignity.

For example, according to the New York Police Department (NYPD), individuals can be added to the city's gang database through “self-admitting” social media posts (Groups Urge NYPD Inspector General to Audit the NYPD “Gang Database.”, 2020). Social media indicators that law enforcement use to corroborate gang membership include being in a known gang location, in possession of gang-related documents or is shown associating with known gang members. “Admitting” to gang membership through social media relies on officers' judgment which overwhelmingly lacks the cultural understanding and neighborhood context necessary to accurately discern the local traditions of Black and Latinx youth (Stuart, 2020). This is especially concerning as inclusion in a gang database can adversely impact an individual's access to housing, schooling, and justice involvement. Moreover, Lageson (2020) highlights how individuals and companies produce online criminal record content such as personal names and arrest photos. She showed how these instances of digital punishment restrict a person's ability to obtain employment or housing even if their charge never leads to adjudication. These examples display the grave importance of protecting participant privacy. As a result, researchers should consider the real-life impact of their decisions regarding participants' privacy.

When researching complex social problems such as gender, racial bias, and community violence, engaging in a reflexive approach should foster participation from community experts. AI researchers utilizing social work practices, which promote participant expertise, can bolster community defined privacy guidelines that ensure individual and group security throughout the research process. For instance, through a reflexive approach to community engagement, social media AI researchers can increase privacy standards that result in the protection of users’ identities by omitting username, and other identifying content (e.g. location, race, gender) in presentations and publications. Using unsearchable altered text, similar creative open-source images, anonymous examples, and password-protected URLs can increase the privacy of social media users. All measures and methods must be vetted with the community through continuous dialogue (i.e. focus groups, research meetings, local advisory meetings) to ensure that privacy and protection remain at the forefront, regardless of research objectives.

3.1.2 Accountability

The ethical principle of accountability and commitment to clients is a focus of social work practice. Social workers’ primary responsibility is to promote the well-being of clients and respect their self-determination. By respecting the self determination of clients, social workers promote the rights of clients and assist clients in their efforts to identify and clarify their goals. The social work value of accountability and thus respect for the self determination of clients can be extended to data science research in order to emphasize the importance of returning benefits to the communities under study (franzke, 2020). However, AI researchers like social workers have their own biases, experiences, and objectives which may embed harmful biased interpretations and values in the algorithms they develop if they fail to communicate with communities being most impacted. For example, data collected to develop facial recognition algorithms tends to historically be mostly White, Euro-centric, and thus discriminatory in nature (Benjamin, 2019; Browne, 2015; Nkonde, 2019). Facial recognition applications have been shown to produce false positives for Black faces which have led to disproportionate arrests. In Detroit, facial recognition led to a wrongful arrest of a Black resident as the software found false positives between his driver’s license photo and the granular surveillance footage taken at the crime scene (Garvie, 2020).

This example highlights how AI systems with the objectives of promoting safety can exacerbate social diseases such as inequality and systematic racism. In this instance AI researchers who practiced reflexivity may have understood the errors that are prominent when facial recognition analyzes Black faces (Buolamwini & Gebru, 2018) and thus hold themselves accountable and begin collaborating with more diverse populations and experiences to improve its accuracy. These reflexive practices that promote diversity and inclusion and self-determination of the community can reduce bias (Probst & Berenson, 2014) and thus some of the unexpected harms toward the communities with whom AI researchers and social work researchers are hoping to support. It is important that scholars using AI to address real-world problems such as safety and health engage in reflexivity and accountability to constantly question their research process, how their models are used and if it respects the self-determination of vulnerable communities.

For example, within the field of violence prevention, it is vital that AI researchers consider the delicate balance between developing tools that increase surveillance and respecting the agency and self-determination of communities impacted by violence. It can be assumed that communities, and specifically Black and Latinx communities, feel averse to outsiders, such as police officers, social workers, or academics, patrolling their “digital streets”. It is no secret that AI tools can be--and are--used against youth through state-sanctioned violence, enacted by surveilling the hyperlocal language and pictures on social media that can be used as evidence or negative character testimony within the criminal justice system (Patton et al, 2019). While families with children who are vulnerable to gun violence, simply seek safety and protection. They are often willing to listen and adhere to any possible computational tools that may reduce victimization. Others may feel uneasy with institutions or agencies surveilling their social media accounts even if it’s meant to increase their safety from acts of community violence. While it is difficult to find a balance between a community’s desire for safety and protection and the potential for harm, reflexivity offers tools and questions to carry this discussion forward.

3.1.3 Fairness

In a broad sense, two central ideas permeate the social justice literature, namely equality and fairness. Social work literature, social work history and social work ethical codes all reflect and, in many instances, actively draw on the idea of social justice as a central component in social work practice. Through this ethical principle, social workers are expected to engage in reflexivity in order to achieve fairness, and equality of outcomes and treatment with their clients; recognizing their dignity and equal worth; work to the meet their basic needs; reduce the inequalities in wealth, income and life chances; and the participation of all, including the most disadvantaged.

Within AI research a similar reflexive approach towards fairness requires incorporating social work principles of social justice that involves people and ethics in research--emphasizing that technology complements people, does not replace them, but involves them in the research (Grosz, 2019). The current development of AI systems concerning marginalized communities often lacks much of their input, and therefore requires the ability to precisely understand the community's online and offline worlds. An example of AI technology that is fair, and inclusively developed, is Being 1.5, a virtual therapist that is powered by AI (Small, 2020). The artist, Rashaad Newsome, is developing this service in the wake of George Floyd’s death in order to address the lack of mental healthcare for Black Americans and create a technology that will feel like a therapeutic exchange. Being 1.5 will address collective trauma that Black Americans experienced by pulling data from previous research and writings authored by Black scholars, activists, and psychotherapists (Rapid Response Fellow, 2020). Fairness is not a hierarchy of relevance but a web of connections; this example shows that with an eye towards social justice and inclusivity. Fair tech development comes with inclusive data sets that are representative of multiple populations. If social justice principles of fairness and equality are excluded from the development of the AI system, then that system may fail to take them into consideration and its impact on vulnerable populations and communities throughout its use. Social justice and Inclusive efforts should follow reflexive conversations where researchers utilize the expertise of community members to understand local and cultural norms

Tech development without user-based and community-based research is a simply myopic, potentially dangerous way of operationalizing new ideas. Ethical tech development does not only involve the participation of participants, but also their input in order to develop an effective technology that addresses community desires and needs. Ruha Benjamin (2019b) discusses “healthcare hot-spotting,” which is the use of data to redistribute resources to high-needs, high-cost patients. Although this technology has the potential to help people, hot-spotting uses geographic information systems, as well as racial profiling, in order to identify the highest-need populations, which inherently leads to stigmatization. As a result, those classified as high needs are seen as dependent and incapable of self-care, another reinstitution of racialized discrimination. This is a great example of a good-intentioned technology potentially harming those in the community because it does not necessarily involve working *with* the people the data represents. On the other hand, the Camden Coalition of Healthcare Providers is made up of an interdisciplinary team of nurses, social workers, and community health workers, instituting healthcare hot-spotting with social and racial justice in mind (Camden Core Model, n.d; Finkelstein et al. 2020). With this focus on fairness, inclusivity, and community collaboration, they are dedicated to understanding the “non-medical needs that affect health: housing, mental health, substance abuse, and emotional support” (Benjamin, 2019b).

3.2 Stages and Questions for Reflexivity:

In the previous sections we have detailed reflexivity within the field of social work and its importance in promoting the social work ethics of privacy, accountability and fairness. In this last section we draw on Houston's (2015) five states of engaging in reflexivity for social work practice. The framework includes five stages of reflexivity that we've modified for AI researchers. The rationale for adopting these stages of reflexivity is that they highlight how power circulates as a permissive and constraining aspect within our individual and social lives (Houston, 2015). As a result, we advocate that Data Scientists consider the stages of reflexivity at each stage of AI development to assess various implications that may impact marginalized communities. Nevertheless, not all AI projects can follow the steps in sequence. Therefore, these stages are flexible, and AI researchers should engage in the stages of reflexivity, regardless of the order outlined in the paper.

We also modified questions produced by Houston that AI researchers may utilize when promoting reflexivity within their research. When embedding the ideas of reflexivity into AI models, AI researchers would benefit from utilizing reflexive questions in all stages of AI development. For example, while collecting data to train the AI model, researchers should discuss how their life experiences and cultural backgrounds may impact the interpretation of the data and its use in training the AI model. These conversations can provide meaningful discussions around potential blind spots and biases that may impact the AI model. A summary of the stages and questions can be found in table 1 below.

3.2.1 Stage One: The AI research team applies the framework of reflexivity to their own life experience.

In this initial stage, the AI research team should consider how personal experiences, organizations, culture and the politico-economy have molded their perspectives and outlook regarding privacy, accountability and fairness. In doing so researchers might ask themselves: How have my personal experiences or identities impacted my outlook on life? How have my personal experiences or culture impacted my outlook regarding privacy, accountability and fairness? This initial stage may present a challenge to the AI research team, especially if they work in a setting where a bureaucratic and procedural culture don't leave room for reflexivity. Not having the time to engage in such introspection, because of unremitting practice demands, pose a formidable barrier to this kind of activity.

3.2.2 Stage Two: The AI research team considers how their personal experiences and cultural background impacts their interactions with each other.

In this second stage, the research team explores together how their respective personal and social attributes impact their own interactions with one another. There could be a gender and class difference between the research team. Moreover, they could look at the world through different cultural lenses which impact how they view privacy, accountability and fairness. Researchers might ask themselves: how do these differences affect the way the AI research team relates to one another? What issues may it create? What potential misunderstandings may it evoke?

3.2.3 Stage Three – The AI research team applies the reflexive framework to 'tune-in' to the needs of the community being studied.

Building on the preceding stages, the AI research team attempts to understand how the domains of privacy, accountability and fairness and the power dynamics operating within them, have shaped the meanings, perspectives, needs, experience and the risks that the communities being studied face. This stage of reflexivity involves a process of tuning-in to the life of the community being studied in order to deepen accurate empathy, compassion and sensitivity and also to gain greater insight into how assessment, planning, intervention and evaluation should be structured. This process of tuning-in may be done in collaboration with communities where researchers learn directly from populations being impacted regarding their needs and how they think about privacy, accountability and fairness. How have wider economic and systemic realities impacted the well-being of the population being studied?

3.2.4 Stage Four – The AI research team apply the reflexive framework to reflect on their interaction with populations being studied.

In this stage, the researchers examine how they interact with the communities that will be impacted by their research. A critical issue here is how the researcher's gender, age, cultural background, race, religion and social class interface with communities with different (or perhaps similar) profiles in relation to these personal and social categories. Importantly, how does the community think about and define privacy, accountability and fairness? How are differences in power and cultural capital expressed? How will the differences in background between researchers and participants influence how the research will be conducted?

3.2.5 Stage Five – Towards more meta and abstract reflexivity.

In this final, cumulative stage in the reflexive process, the insights that were illuminated from the preceding stages are brought together, examined, processed and synthesized. By reviewing the overall process, the AI research team is able to identify recurrent themes around the use of power when conducting research. Moreover, this process can highlight how their thoughts of privacy, accountability and fairness have changed. Important questions within stage five may be: How does who we are, because of our background and range of social experience, shape how we carry out research? What population or community experiences may differ radically from our own? Responses to this question come as a result of a process of meta- reflection, a process which integrates the insights from stages one to four.

By engaging with each stage of the social work reflexive process AI scholars can better understand how their personality experiences and identities impact their relationships with fellow researchers as well as the communities and populations being studied. By having a deeper understanding of these potential differences AI researchers can be more effective in promoting privacy, accountability and fairness within each stage of research.

Table 1: Stages of Reflexivity

Stages of Reflexivity	***Reflexive Questions***
Stage 1: The AI research team applies the framework of reflexivity to their own life experience.	How have my personal experiences or identities impacted my outlook on life?
	How have my personal experiences or culture impacted my outlook regarding privacy, accountability and fairness?
Stage 2: The AI research team considers how their personal experiences and cultural background impacts their interactions with each other.	How do these differences affect the way the AI research team relates to one another?
	What issues might these differences create?
	What potential misunderstandings may these differences evoke?
Stage 3: The AI research team applies the reflexive framework to 'tune-in' to the needs of the community being studied.	How have wider economic and systemic realities impacted the well-being of the population being studied?
Stage 4: The AI research team applies the reflexive framework to reflect on their interaction wi populations being studied	How does the community think about and define privacy, accountability and fairness?
	How are differences in power and cultural capital expressed?
	How will the differences in background between researchers and participants affect how the research will be conducted?
Stage 5: Towards more meta and abstract reflexivity.	How does who we are, because of. our background and range of social experience, shape how we carry out research?
	What population or community experiences may differ radically from our own?

4. Conclusion

Within the practice of social work, practitioners often have to contend with everyday social complexities and real-world problems. These range from interactions with individuals, families, communities and institutions. Through these experiences many social workers understand the messiness of the human experience and that prescriptive/rigid interventions do not always resolve the challenges clients may be facing. As a result, social workers use reflexivity to delicately balance their professional expertise and the goals of the client that may not coalesce (Furlong, 2003). The goals of the client guide the social worker's professional relationship, even if they are at odds with his or her expertise. In these situations, social workers offer information about potential short- and long-term consequences. Similarly, AI developers and researchers could engage in reflexivity and utilize their expert knowledge to determine whether to continue the project as planned or modify goals to reduce potential harm.

As the focus of AI shifts to more practical, real-world interventions, it begins to enter the space that social work has inhabited. Thus, we advocate for AI researchers to be flexible by engaging in a more reflexive and ethical approach with the understanding that their AI model may produce unintended consequences that affect lives and livelihoods of vulnerable/marginalized communities.

In this paper we highlight the shortcomings of AI application in research and elucidate how practicing reflexivity while conducting research can help center the experiences of participants and determine if AI is the correct tool for intervention. We extend social work principles of respecting privacy, accountability, and fairness as a tool for improving annotations, data labeling, data recruitment, validation and implementation of AI models. One of the inevitable complications in applying reflexivity is the large degree of interpretation involved. Researchers may consider the questions presented in each stage of reflexivity while also expanding upon them based on their individual projects.

We strongly recommend that researchers open the channels of communication between themselves and the populations being impacted by their research. It is imperative that the research team collaborates with the people who will potentially use this technology. By integrating the user into the tech development and creating roles for researchers to be community liaisons can lead to a greater amount of community advocacy. Regarding accountability, although it might not be possible to guarantee that no harm has been done, it again goes back to opening the conduits of communication, and ideally, ensuring that there is not an alarming amount of separation between researcher and participants. Ideally, these two would be one in the same. By embedding the stages of reflexivity and reflexive questions into scholarship, AI researchers can better tackle real-world problems in the 21st century given their unique insight and skill set. We believe that the application of social work ethics and approaches to data science can possibly prevent future mistakes in our research with communities.

References:

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine Bias: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Benjamin, R. (2019). *Assessing Risk, Automating Racism.* Science, 366(6464), 421-422. Doi: 10.1126/science.aaz3873.

Benjamin, R. (2019b). *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity.

Ben-Ari, A., and Enosh, G. (2013). Power Relations and Reciprocity: Dialectics of Knowledge Construction. Qualitative Health Research, 23(3), 422-429. doi:10.1177/1049732312470030.

Blandfort, P., Patton, D. U., Frey W. R., Karaman, S., Bhargava, S., Lee, F. T., Varia, S., Kedzie, C., Gaskell, M. B., Schifanella, R., McKeown, K. (2019). Multimodal social media analysis for gang violence prevention. *In Proceedings of the International AAAI conference on web and social media*, vol 13(1). pp. 114–124.

Bolton, G. (2010). *Reflective practice: Writing and professional development*. Sage publications.

Brock, A. (2015). *Deeper data: A Response to boyd and Crawford.* Media, Culture Society, 37(7), 1084-1088. doi:10.1177/0163443715594105.

Brown, J. (2006). Reflexivity in the research process: Psychoanalytic observations. *International Journal of Social Research Methodology*, *9*(3), 181–197. https://doi.org/10.1080/13645570600652776

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness.* Duke University Press.

Buolamwini, J., and Gebru, T. (2018). *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Conference on fairness, accountability and transparency, in PMLR. 81:77-91.

Camden Core Model (n.d.). Https://camdenhealth.org/care-interventions/camden-core-model/.

Citron, D. K., and Pasquale, F. (2014). *The Scored Society: Due Process for Automated Predictions*. Washington. Law Review, 89(1), 1-34.

Cizek, K., and Uricchio, W. (2019). Collective Wisdom: Co-creating Media Within Communities, across Disciplines, and with Algorithms. *Works in Progress,* https://wip.mitpress.mit.edu/collectivewisdom.

Costanza-Chock, S. (2018). Design justice: Towards an Intersectional Feminist Framework for Design Theory and Practice. *Proceedings of the Design Research Society*.

D'cruz, H., Gillingham, P., and Melendez, S. (2007). Reflexivity, its Meanings and Relevance for Social Work: A Critical Review of the Literature. *The British Journal of Social Work*, 37(1), 73-90. Doi:10.1093/bjsw/bcl001.

Dodgson, J. E. (2019). *Reflexivity in Qualitative Research*. Journal of Human Lactation, 35(2), 220–222. Doi:10.1177/0890334419830990.

Eubanks, V. (2019). *Automating inequality: How high-tech tools profile, police, and punish the poor*. Picador, St. Martin's Press.

Floridi, L., Cowls, J., King, T. C., and Taddeo, M. (2020). *How to Design AI for Social Good: Seven Essential Factors.* Science and Engineering Ethics, 26(3), 1771–1796. Doi;10.1007/s11948-020-00213-5.

Finkelstein, A., Zhou, A., Taubman, S., and Doyle, J. (2020). Health Care Hotspotting—A Randomized, Controlled Trial. *New England Journal of Medicine*, *382*(2), 152-162. Doi:10.1056/NEJMsa1906848.

Finlay, L. (2002). Negotiating the swamp: the opportunity and challenge of reflexivity in research practice. *Qualitative Research*, *2*(2), 209–230. https://doi.org/10.1177/146879410200200205.

Finn, J. L. (2020). *Just Practice: A Social Justice Approach to Social Work*. Oxford University Press.

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C. and The Association of Internet Researchers (2020). *Internet Research: Ethical Guidelines 3.0.* https://aoir.org/reports/ethics3.pdf.

Furlong, M. A. (2003). Self-determination and a critical perspective in casework. *Qualitative Social Work*, *2*(2), 177–196. https://doi.org/10.1177/1473325003002002004.

Garvie, C. (2020). *The Untold Number of People Implicated in Crimes They Didn't Commit Because of Face Recognition*. American Civil Liberties Union. https://www.aclu.org/news/privacy-technology/the-untold-number-of-people-implicated-in-crimes-they-didnt-commit-because-of-face-recognition.

Gebru, T. (2019). *Oxford Handbook on AI Ethics Book Chapter on Race and Gender.* arXiv preprint arXiv:1908.06165.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., and Crawford, K. (2018). *Datasheets for datasets*. arXiv preprint arXiv:1803.09010.

Gilgun, J. F. (2008). *Lived Experience, Reflexivity, and Research on Perpetrators of Interpersonal Violence.* Qualitative Social Work, 7(2), 181-197. Doi:10.1177/1473325008089629.

Grosz, B. J. (2019). *The AI Revolution Needs Expertise in People*, Publics and Societies. 1.1, 1(1). Doi:10.1162/99608f92.97b9554.

Groups Urge NYPD Inspector General to Audit the NYPD "Gang Database.". (2020). Human Rights Watch. https://www.hrw.org/news/2020/09/22/groups-urge-nypd-inspector-general-audit-nypd-gang-database#.

Hagendorff, T. (2020). *The Ethics of AI Ethics: An Evaluation of Guidelines.* Minds and Machines, 30, 99–120. Doi:0.1007/s11023-020-09517-8.

Houston, S. (2015). Enabling others in Social Work: Reflexivity and the theory of social domains. *Critical and Radical Social Work*, *3*(2), 245–260. https://doi.org/10.1332/204986015x14302240420229.

Jobin, A., Ienca, M., and Vayena, E. (2019). *The Global Landscape of AI Ethics Guidelines.* Nature Machine Intelligence, 1(9), 389-399. doi:10.1038/s42256-019-0088-2.

Lageson, S. E. (2020). *Digital punishment. privacy, stigma, and the harms of data-driven criminal justice*. Oxford University Press.

Landau, A.Y., Blanchard, A., Cato, K., Atkins, N., Salazar, S., Patton, D.U., & Topaz, M. (2022). Considerations for Development of Child Abuse and Neglect Phenotype with Prioritizing Reduction of Racial Bias: A Qualitative Study. *Journal of the American Medical Informatics Association*. 29:3, 512-519.

Landau, A.Y., Ferrarello, S., Blanchard, A., Cato, K.,Atkins, N., Salazar, S., Patton, D.U., & Topaz, M. (2022b). Developing machine learning-based models to help identify child abuse and neglect: key ethical challenges and recommended solutions. *Journal of the American Medical Informatics Association*. 29:3, 576-580.

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., and Wagner, C. (2020). *Computational Social Science: Obstacles and Opportunities*. Science, *369*(6507), 1060. Doi:10.1126/science.aaz8170.

Longhofer, J., and Floersch, J. (2012). The Coming Crisis in Social Work: Some Thoughts on Social Work and Science. *Research on Social Work Practice*, *22*(5), 499-519. Doi:10.1177/1049731512445509.

National Association of Social Work (NASW). (2017) *The NASW Code of Ethics*. https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English.

Nkonde, M. (2019). *Automated Anti-Blackness: Facial Recognition in Brooklyn, New York*. Kennedy School Review, 20, 30-36.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting Racial Bias in an Algorithm used to Manage the Health of Populations.* Science, 366(6464), 447-453. Doi: 10.1126/science.aax2342.

O'neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.

Palaganas, E. C., Sanchez, M. C., Molintas, V. P., and Caricativo, R. D. (2017). *Reflexivity in Qualitative Research: A Journey of Learning*. Qualitative Report, 22(2), 426-438.

Patton, D. U. (2020). Social work thinking for UX and Ai Design. *Interactions*, *27*(2), 86–89. https://doi.org/10.1145/3380535.

Patton, D. U., Brunton, D.-W., Dixon, A., Miller, R. J., Leonard, P., & Hackman, R. (2017). Stop and frisk online: Theorizing everyday racism in digital policing in the use of social media for identification of criminal conduct and Associations. *Social Media + Society*, *3*(3), 205630511773334. https://doi.org/10.1177/2056305117733344.

Patton, D. U., Frey, W. R., & Gaskell, M. (2019). Guns on social media: Complex interpretations of gun images posted by Chicago Youth. *Palgrave Communications*, *5*(1), 1–8. https://doi.org/10.1057/s41599-019-0330-x.

Patton, D. U., MacBeth, J., Schoenebeck, S., Shear, K., & McKeown, K. (2018). Accommodating grief on Twitter: An analysis of expressions of grief among gang involved youth on Twitter using qualitative analysis and natural language processing. *Biomedical Informatics Insights*, *10*, 117822261876315. https://doi.org/10.1177/1178222618763155.

Pillow, W. (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education*, *16*(2), 175–196. https://doi.org/10.1080/0951839032000060635.

Probst, B. (2015). *The eye regards itself: Benefits and Challenges of Reflexivity in Qualitative Social Work Research.* Social Work Research, *39*(1), 37-48. Doi:/10.1093/swr/svu028.

Probst, B., Berenson, L. (2014). *The double arrow: How qualitative social work researchers use reflexivity.* Qualitative Social Work, *13*(6), 813-827. doi:10.1177/1473325013506248.

Raji, I., Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3306618.3314244.

Rapid Response Fellow. (2020). https://www.eyebeam.org/residents/rashaad-newsome-rr/.

Reid, A. M., Brown, J. M., Smith, J. M., Cope, A. C., and Jamieson, S. (2018). *Ethical Dilemmas and Reflexivity in Qualitative Research*. Perspectives on Medical Education, *7*(2), 69-75. Doi:10.1007/s40037-018-0412-2.

Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*, 100005. https://doi.org/10.1016/j.jrt.2020.100005.

Schön, D. A. (1992). *Designing as reflective conversation with the materials of a design situation.* Research in Engineering. *3*(1), 131-147. Doi:10.1007/BF01580516.

Small, Z. (2020). *How Artists Are Trying to Solve the World's Problems.* New York Times. Https://www.nytimes.com/2020/07/15/arts/design/eyebeam-art-project.html.

Stuart, F. (2020). *Code of the Tweet: Urban Gang Violence in the Social Media Age.* Social Problems, 67(2), 191-207. Doi:/10.1093/socpro/spz010.

Valandra, V. (2012). *Reflexivity and professional use of self in research: A doctoral student's journey*. Journal of Ethnographic and Qualitative Research, 6, 204–220. ISSN: 1935-3308.

Watts, L. (2019). *Reflective Practice, Reflexivity, and Critical Reflection in Social Work Education in Australia*. Australian Social Work, *72*(1), 8-20. Doi:10.1080/0312407X.2018.1521856.

Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., and Goldenberg, A. (2019). *Do No Harm: A Roadmap for Responsible Machine Learning for Health Care.* Nature Medicine, *25*(10), 1627-1627. Doi:10.1038/s41591-019-0609-x.

Whittlestone, J., Arulkumaran, K., & Crosby, M. (2021). The Societal Implications of Deep Reinforcement Learning. *Journal of Artificial Intelligence Research*, *70*. https://doi.org/10.1613/jair.1.123.