ARTIFICIAL
INTELLIGENCE
FOR
DEVELOPMENT
AFRICA

Key action areas for more gender-responsive AI

Learning Brief No. 3 • Alex Berryhill and Lucía Mesa Vélez (Ladysmith) • December 2022

Objectives of this resource: To outline key action areas for advancing more gender-responsive AI, in order to ensure conversations around GEI move from theoretical to operational. *This brief's synthesis of the available literature is far from comprehensive, and is meant to serve as a starting point for discussions and actions around GEI excellence for AI4D labs and hubs.*

This is the third resource in the **AI4D GEI Support Team's learning brief series**:

- **Learning Brief No. 1** summarizes priority GEI considerations for the design of AI4D calls for proposals.
- **Learning Brief No. 2** provides an operational definition of gender-responsive projects in the context of AI-driven research and innovation, and synthesizes available research that illustrates why GEI considerations matter for the design and use of algorithmic decision-making.

Gender-responsiveness appears to be a particular challenge for AI research and innovation projects, as illustrated by a growing evidence base of the many ways in which AI applications have contributed to and exacerbated gender, racial, economic and global inequalities (see Learning Brief No. 2).[1] **There are no fully conclusive 'checklists' on how to 'fix' the AI sector's 'gender problem'**—although certain methodological and organizational considerations can help 'move the needle' towards more gender-responsive AI research and innovation, some of which are outlined below. For more transformative and sustainable impact, teams leveraging AI must continuously reflect upon *their* roles and responsibilities in broader organizational, cultural, and political systems change.

Indeed, it is important to note that *GEI will mean something distinct for each specific team, project and context.* With this in mind, these key action areas can be viewed as critical, evidence-driven starting points.

Key GEI considerations for AI-driven research and innovation

1. **Create representative teams with equitable decision-making structures:** Gender-responsiveness begins with how teams are formed, and how decisions are

[1] Mandal, A. (2021). "Algorithmic Origins of Bias." Women at the Table.

IDRC · CRDI   Sida   WOMEN AT THE TABLE   Ladysmith   gender at work

made. The AI sector is well known for its lack of gender and racial diversity, particularly in leadership and decision-making positions—as are the consequences of this underrepresentation in the design and use of AI. When considering diversity, it is important to consider identity, experiences, and also professional training or disciplines (i.e., a diversity of interdisciplinary perspectives also contribute to a team's gender-responsiveness).

Important first steps to ensure representativeness in a project team includes leading an assessment and analysis of team composition, along with an analysis of the barriers women and other marginalized groups encounter to enter and advance in the sector, within the project team's specific context. This data can then be used to inform the design of context-appropriate interventions.

Relevant tools and resources:

- To Build Less-Biased AI, Hire a More-Diverse Team (Li 2020)
- Diverse Teams build better AI. Here's why (Shastri 2020)
- Women in Tech (here)
- 2021 Women in Tech Report (TrustRadius 2021)
- How to keep human bias out of AI. (TED Talk by Kriti Sharma).

2. **Use qualitative and quantitative evidence to develop fit-for-purpose theories of change:** There are many ways in which AI can be used for positive social impact, including advancing gender equality. However, one of the core challenges of the AI sector has been the failure to assume new technologies will alone 'fix' inequality and injustice.

   Instead, gender-responsive projects are informed by evidence on the root drivers of identified challenges, which most often (if not, always) require a variety of interventions to break cycles of inequality and injustice (rather than technological interventions, alone). Importantly, this evidence should include both qualitative and quantitative data, and the perspectives from those who are most likely to understand the identified challenge, such as impacted communities and civil society organizations.[2] For challenges related to gender inequality, this should specifically include local feminist movements and women's rights organizations, as they are most likely to understand key entry points for achieving positive gender outcomes, along with potential risks for proposed activities.

Relevant tools and resources:

- Stanford's Gendered Innovations in Science and Engineering Project (here)
- Women at the Table's <AI & Equality> Human Rights Toolbox (here)
- AI4COVID Technical Brief 1: Designing gender-responsive data projects (Ladysmith 2021)
- Meeting the challenge of gender inequality through gender transformative research: lessons from research in Africa, Asia, and Latin America (Njuki et al 2022)
- Transforming gender relations: Insights from IDRC research (IDRC 2019).

3. **Identify and appropriately address statistical and social data biases:** One must assume that all data has some degree of bias. This may include various types of "statistical bias (i.e., concerns about nonrepresentative sampling and measurement error)" as well as societal biases "(i.e., concerns about objectionable social structures and past injustice that are represented in the data)".[3] While there is no conclusive definition of how to define or measure a 'fair' or 'inclusive' dataset, recognizing the biases interwoven in the data used to inform AI applications, is a critical first step towards greater gender-responsiveness. By identifying potential biases, teams can then identify the most appropriate methods for further diagnosing biases and then appropriately addressing these biases.

   Note that here, once again, a variety of disciplinary perspectives is key to identifying potential biases that might otherwise go overlooked. For example, sociologists and

[2] Fuentes, L. and Cookson, T.P. (2020). "Counting gender (in)equality? a feminist geographical critique of the 'gender data revolution'." *Gender, Place & Culture* 27(6): 881-902.
[3] Mitchell, S., et al. (2021). "Algorithmic Fairness: Choices, Assumptions, and Definitions." *Annual Review of Statistics and Its Application* 8:141-163.

mathematicians may be best equipped to identify and respond to different types of biases, thus the need for cross-disciplinary collaboration.

Relevant tools and resources:

- Algorithmic Origins of Bias (Mandal 2021)
- A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle (here)
- Implicit Association Test (IAT) by Harvard University (here)
- IBM's open-source Library AI Fairness 360 to detect and mitigate biases (here)

4. **Monitor model performance and impacts, including on social norms and inequalities:** A second critical step towards more inclusive datasets, and thus more gender-responsive AI research and innovation, is identifying an appropriate methodology for closely monitoring the performance of prediction models and any potential unintended consequences, including potential negative social impacts like the exacerbation of discriminatory gender norms. Along with a strong and regular monitoring methodology, gender-responsive teams ensure there are effective means for connecting monitoring data with programmatic decision-making (i.e., do we need to iterate upon the design or use of our model, or is there a need to terminate the project altogether?).

Relevant tools and resources:

- IBM's Watson OpenScale performs bias checking and mitigation in real time when AI is making its decisions (here)
- Google's 'What-If tool' to test algorithms performance with different datasets (here)
- Local Interpretable Model-Agnostic Explanations (LIME) (here)
- FairML: a Python open-source toolbox that is used to audit machine learning predictive models (here)

5. **Facilitate impacted communities' participation in project design, implementation, monitoring, and evaluation:** Gender-responsive AI research and innovation meaningfully involves impacted communities throughout the project's design, implementation, monitoring and evaluation. This is key from a human rights perspective (i.e., to respect individuals and communities' rights to participate in the decisions that impact their lives), but also from an impact and sustainability perspective: Participation from those most impacted by new interventions can help identify and mitigate potential risks, and ensure project's relevance for local contexts.

   Formalizing partnerships is an important first step to successfully engage communities in research and innovation projects. This might include a formal contract, or a memorandum of understanding to ensure a shared understanding of the details of the relationship. Teams should also formalize time and space for community participation throughout project plans and processes, in order to ensure these activities are prioritized.

   Facilitating impacted communities' participation throughout a project's life cycle requires building trust and goodwill with representative organizations, which can be a challenge in contexts where AI-driven technologies have created harm, or where there is a lower understanding of AI tools. To strengthen partnerships and create meaningful opportunities for engagement and feedback, teams must invest in ongoing communication, be willing to adapt project plans based on feedback, and consider investing in mutual capacity building.

Relevant tools and resources:

- AI4COVID Technical Brief 3: Stakeholder engagement for gender-responsive health research (Ladysmith 2021)
- AI4COVID Technical Brief 4: Connecting Gender Data to Action (Ladysmith 2022)
- Responsible AI and its stakeholders (Lima et al 2020)

Annex A: Key terms

- **Gender**: Gender refers to the socially-constructed roles, responsibilities and relationships that society considers appropriate for women and men. It also has implications therefore for individuals and groups who identify as gender non-conforming. Gender is upheld by political, economic, social, and cultural institutions. Gender is context and time-specific, and thus changeable as well.
- **Sex:** The sum of biological and physiological characteristics that typically define men and women, such as reproductive organs, hormonal make-up, chromosomal patterns, hair-growth patterns, distribution of muscle and fat, body shape and skeletal structure.
- **Intersectionality:** The cumulative way in which the effects of multiple forms of oppression, discrimination and exclusion (including but not limited to racism, sexism, and classism) combine, overlap, or intersect.
- **Inclusion:** The aim of inclusion is to embrace all people irrespective of race, gender, disability, medical or other need. It is about giving equal access and opportunities and getting rid of discrimination and intolerance (removal of barriers). It affects all aspects of public life.