moz://a

Mozilla Kiswahili Project Gender Action Plan

Led By Rebecca Ryakitimbo
Supported by: Chenai Chair, Britone Mwasaru, Kathleen Siminyu, Hillary Juma
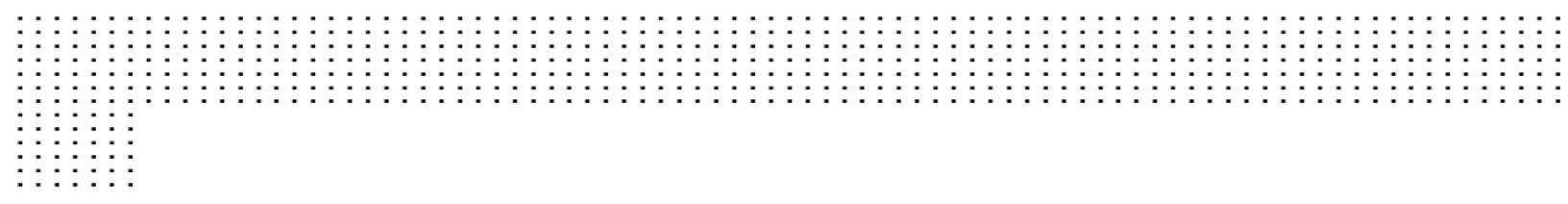*June 2021*

Introduction

Mozilla Foundation is embarking upon a three year funded project focused on building out language data corpora in East Africa. Kiswahili has generously been supported by the [Bill and Melinda Gates Foundation](), [GIZ]() and UK's Foreign Commonwealth & Development Office (FCDO). This core work will build skills and capabilities within the Mozilla Foundation, and within the region, to democratize voice tech through community engagement and contribution. It will also support use case development within agriculture and finance.

The project seeks to mitigate AI biases by developing strategies for engagement with women who are often missing within such innovations, and on equity in data collection and model creation. At Mozilla Gender integration is emphasized through [Mozilla's 10 principles and the accompanying addendum]() such as our commitment to "***An internet that includes all the peoples of the earth — where a person's demographic characteristics do not determine their online access, opportunities, or quality of experience".*** Also, Mozilla's work on trustworthy AI advocates for gender equality as our AI theory of change calls for diverse stakeholders including people historically underrepresented or shut out of tech to be engaged in the design of AI.[1] It is important to reflect how these principles will be applied through a Gender Plan.

The Gender Plan details how gender diversity will be ensured during data collection, model creation, and use case development. We will also assess the overall impact of the project on the gender digital divide. The action plan contains specific gender elements to be considered in project design and

---

[1] Mozilla WIKI foundation AI: https://wiki.mozilla.org/Foundation/AI

implementation, monitoring and evaluating progress. It will help build an effective strategy for gender integration in the project ensuring diversity and inclusion.

How are we thinking about gender?

The Kiswahili voice project wants to curate a diverse and inclusive community that will help build an inclusive and trustworthy AI. Gender inequality within technological development, skewed more so towards women and gender diverse communities, often means these groups are not part of the development and implementation of new technologies. According to World Bank Data, the literacy rate for women in sub-Saharan Africa is only 58%,[2] Voice technology can help to break down the digital gender divide by reducing the need for traditional literacy in using applications; this is dependent upon the use, and acceptance, of the technology by women. Women are a crucial part of the tech and language community the Kiswahili common voice project intends to curate. Engagement and trust-building, with partners sensitive to the need for gender equity, will be vital to this outcome. We want to ensure that our project has an equal reach to across genders; that women's participation and leadership is visible within the project and ensure that the products we develop integrate the needs of women as well as men and that the use cases consider gender dynamics.

The following definitions will guide and influence how we are approaching gender in this project:

Gender: refers to the roles and responsibilities of men and women that are created in our families, our societies and our cultures. The concept of gender also includes the expectations held about the characteristics, aptitudes and likely behaviors of both women and men (femininity and masculinity). Gender roles and expectations are learned. They can change over time, and they vary within and between cultures. Systems of social differentiation such as political status, class, ethnicity, physical and mental disability, age and more, modify gender roles. The concept of gender is vital because, applied to social analysis, it reveals how women's subordination (or men's domination) is socially constructed. As such, the subordination can be changed or ended. It is not biologically predetermined nor is it fixed forever.

Sex: Sex refers to the biological differences assigned at birth - usually taking on a binary of male and female. Sex often forms the basis of indicators of gender participation and engagement through sex disaggregated data. This often takes the approach of asking participants to identify as male/female or other as in the practice of the Common Voice platform.

---

[2] World bank data: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=ZG

Gender Equality: means that women and men have equal conditions for realizing their full human rights and for contributing to, and benefiting from, economic, social, cultural, and political development. Gender equality is therefore the equal valuing by society of the similarities and the differences of men and women, and the roles they play. It is based on women and men being full partners in their home, their community and their society.

Gender Equity : is the process of being fair to men and women. To ensure fairness, measures must often be put in place to compensate for the historical and social disadvantages that prevent women and men from operating on a level playing field. Equity is a means. Equality is the result.[3]

Gender diversity: is an umbrella term that is used to describe gender identities that demonstrate a diversity of expression beyond the binary framework and a gender that may not be tied to the sex assigned at birth. [4]

Overall, we understand gender as diverse and heterogeneous; located in societal constructs about roles, expectations and practices that would impact and shape participation of gender groups — in particular women differently from men. We aim to build a diverse and inclusive voice community that is accessible to all considering these issues together with other structural issues impacting gender inequality such as location, sexual orientation, education, region, and accent.

Challenges around gender within Voice systems

 There is a proliferation of voice assistants often being designed to replicate female gendered characteristics. However, there is much to be desired in voice tech for gender inclusivity for actual participation and meaningful engagement of women. Recent work has found how gendered biases against women are embedded into these systems[5] We highlight some of the challenges that we take into consideration for our gender plan:
   a. **Speech recognition is more accurate for men than it is for women:** Research by Dr. Tatman published by the North American Chapter of the Association for Computational Linguistics

---

[3] UNESCO's Gender Mainstreaming Implementation Framework:
http://portal.unesco.org/en/ev.php-URL_ID=11481&URL_DO=DO_TOPIC&URL_SECTION=201.html

[4] Gender rights, what is Gender Diversity: https://genderrights.org.au/information-hub/what-is-gender-diversity/

[5]I'd blush if I could: closing gender divides in digital skills through education
https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=85

(NAACL) indicates that Google's speech recognition is 13% more accurate for men than it is for women.[6]

b. **Less female voice data sets and less inclusive voice tech**: Majority of voice programs are geared towards profit generation and often created to cater for specific markets often using male white voices. It is difficult to find sex disaggregation of data sets to at least indicate the extent of participation for purposes of gender analysis. In addition, even when designed to be open and inclusive there is a challenge of female voice contributions. Mozilla's common voice platform sex disaggregated data sets show majority of data sets being heavily male, for example English voice donations shows male voice data sets dominate at 47% while women fall behind at only 15%.[7] With lack of representation of female voice data sets to train the models it is difficult to find the patterns within data that are representative of the female population, leading to biases in algorithms created.[8]

c. **Access and affordability gender gaps**: Women especially in developing countries have less access to technology than men, in terms of access to devices and the internet. The gaps in access and affordability are experienced more so by women in comparison to men in lower income segments, lower levels of education, limited access to resources as well as urban and rural disparities. These gaps result in less contribution to and making use of technologies for women in marginalized groups.

d. **Lack of diversity in voice tech communities**: Majority of developers in the field of voice tech are men, leading to biases in the voice curation, modeling and designing of speech based tech. Princeton professor Ruha Benjamin notes that "Any analysis of AI bots should consider the diversity and associated biases of the teams that design them."[9]

e. **Gender-based harassment in AI bots:** An analysis conducted by Leah Fessler of Quartz in 2017 titled "*We tested bots like Siri and Alexa to see who would stand up to sexual harassment*"[10] found that most AI bots responded either with gratitude or denial to sexually

---

[6] www.ethicsinnlp.org/workshop/pdf/EthNLP06.pdf

[7] Common Voice English datasets: https://commonvoice.mozilla.org/en/datasets

[8] Voice Recognition Still Has Significant Gender Biases:
https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases

[9] **How AI bots and voice assistants reinforce gender bias;**
https://www.brookings.edu/research/how-ai-bots-and-voice-assistants-reinforce-gender-bias/

[10] We tested bots like Siri and Alexa to see who would stand up to sexual harassment:
https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/

assaulting words this has changed in 2020 but still leaves a lot of room for building gender stereotypes.

f. **Lack of open data relating to gender and diversity in voice products and team:** Majority of voice tech innovations do not openly publish data on the diversity of their datasets or in their design/developers' teams. This will perpetuate transparency among tech companies on the gender sensitivity of their products and teams.

g. **Digital literacy -** Marginalized women - in particular those with limited income and lower levels of education also face disproportionate digital literacy and skills gaps which impact their access and use of technology. Research by the Web Foundation focused on the Global South, found that 50% of women in rural areas said they didn't use the internet because they don't know how. 45% of women in urban areas said the same.[11]

h. **Trust -** Issues around privacy and data protection also impact how women interact with voice technologies. There are concerns around increased surveillance, bias, and discrimination as well as data governance that shape the levels of trust for women and gender diverse people. UNESCO recommends that AI should address issues of consent and confirmation of ethical use of data, privacy and security for women and girls.[12]Web Foundation in their research found that 54% of female respondents said they would not allow companies to use any of their data, compared with 47% of men.

---

[11] Women's Rights Online: closing the digital gender gap for a more equal world:
https://webfoundation.org/research/womens-rights-online-2020/

[12] UNESCO Artificial Intelligence and Gender Equality:
https://en.unesco.org/system/files/artificial_intelligence_and_gender_equality.pdf

CV Gender Action Plan

Having identified the issues at hand we have developed a plan of gender integration at each stage of the project. The integration includes measures, actions, and indicators that consider heterogeneity and diversity of gender within the Kiswahili community. We will also account for other factors to be considered for gender integration such as age, rural/urban gaps, and educational background.

SN	Objective	Measures and Actions	Indicators/Targets	Responsible
1	Community: Language and Tech	Ensure a gender-responsive and inclusive awareness raising campaign on Voice tech and Kiswahili cutting across by location, accent, age, income, education background, and gender diversity groups	Information and communication material (posters, newsletters) as well as avenues to events online or onsite are designed and offered in a barrier-free way and equally accessed by women and men.	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Curate a gender balance in community building and engagement cutting across by location, accent, age, income, and education background	# Of gender diverse groups i.e., women forming the tech and language community in equal proportion to men	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Host Gender exclusive events, meetings and roundtables i.e., women	#Of gender exclusive events hosted for specific gender groups # Of gender specific groups that attended each specific event	CV team i.e., community engagement fellows, Special advisor Africa

				Mradi
		Training and capacity building of different gender groups on Common Voice, AI, and Voice technology.	# Of women equal to the number of men trained on usage of the common voice platform, AI, and voice technology. #Take into account age and location to have diverse participation	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Identifying and recruiting voice tech community champions from across the gender spectrum.	# Of gender diverse community champions recruited.	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Development and utilization of existing Mozilla participatory guidelines adhering to gender concerns such as harassment	A set of community guidelines addressing gender concerns such as harassment.	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Engaging stakeholders through round table discussions and meetings that deliberately have equal representation.	# Of people represented at the meetings by gender groups.	CV team i.e., community engagement fellows, Special advisor Africa Mradi

2.	Data collection	Engagement of gender groups, NGOs and media working in sentence curation through writeathorns and available phrases that they can share	# Of gender sensitive phrases/sentences collected # Gender based groups and NGOs involved in the process.	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Ensuring availability of access and devices through partners spaces hubs for gender groups who have no access to contribute	# Of gender diverse groups from rural areas without devices who have donated through our community champions or partners spaces # Of partners willing to offer their space and devices for contributors without access to donate # Of gender diverse groups from rural areas without devices who have don	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Development of a localized Kiswahili consent form that explains usage, data protection and privacy for voice donors.	Users of the CV platform can consent to donation, are made aware of why the data is collected and protect their privacy. This is to address consent and data ownership.	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Aggregation and labeling of data by gender to ensure data can be disaggregated by gender.	Disaggregated data by gender of voice donors showing gender parity.	CV team i.e., community engagement fellows, Special advisor Africa Mradi

3	Model creation	Consultation with Gender diverse groups working on AI and/or Voice technology on gender biases and needs in AI	# Of gender diverse groups consulted on gender biases and needs in AI	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Creation of a gender bias AI checklist to identify bias along axes of gender.	A gender bias AI checklist established to guide AI programmers to audit gender biases in their algorithms and data sets	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Creation and application of models by men and women tech communities that integrate voice services.	# Of digital service integrations achieved # Of individuals reached by those voice services, disaggregated by gender	CV team i.e., community engagement fellows, Special advisor Africa Mradi
4	Use case Development	Wide consultation with feminist and gender based CSO's working on agriculture and finance from a gender perspective	# Of gender based groups and CSOs working on agriculture and finance consulted in use case development.	CV team i.e., community engagement fellows, Special advisor Africa Mradi
		Development of gender specific use cases addressing needs identified by	Gender specific use cases in agriculture and finance created and	CV team i.e., community

		gender stakeholders in the field of agriculture and finance. These use prototypes will subsequently be tested by the respective gender groups.	tested by the respective gender groups.	engagement fellows, Special advisor Africa Mradi
		Adoption and development of Use cases addressing gender focused needs i.e., nutrition	Impact of voice technology on the adoption, use and impact of those services - including improved nutritional outcomes for women and children, through the identified use cases.	CV team i.e., community engagement fellows, Special advisor Africa Mradi

Resources

Iyer, N. Chair, C. and Achieng, G. (2021). Afrofeminist Data Futures. Policy. https://pollicy.org/wp-content/uploads/2021/03/Afrofeminist-Data-Futures-Report-ENGLISH.pdf

Equal. *EqualAI Checklist to Identify Bias in AI.* https://www.equalai.org/resources/

West. M, Kraut, R. and Chew. H,E. (2019) I'd blush if I could: closing gender divides in digital skills through education. *UNESCO and Equals Coalition.* https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1

Mozilla Community Participation Guidelines. Updated January 2020 https://www.mozilla.org/en-US/about/governance/policies/participation/

The Design Justice Network Principles: https://designjustice.org/

UNESCO (2021). AI-enabled Voice Assistants: No longer female by default.

https://en.unesco.org/news/tackling-gender-prejudice-artificial-intelligence-are-we-right-track

------------- (2020). Artificial Intelligence and Gender Equality : Key findings of UNESCO's Global Dialogue.

https://en.unesco.org/system/files/artificial_intelligence_and_gender_equality.pdf