

Session: Strengthening the local AI ecosystem through an open approach to data set creation

A Kenya Innovation Week Side Event

With advancements in technology, governments and other stakeholders are increasingly turning towards AI to support decision-making in various sectors. Natural language processing (NLP) techniques have enabled critical applications to achieve the ability to communicate and be understood in one's own language. This has been recognized as a prerequisite to digital and societal inclusion in education, finance, healthcare, agriculture, communication, and disaster response. However, machine learning datasets are heavily geographically and linguistically skewed towards the United States and Europe, leading to AI models that perform poorly in developing countries. There's a need therefore to expand and maintain open-sourced training and evaluation datasets to enable robust and more equitable application of machine learning tools of high social value.

To this end, GIZ's FAIR Forward project under the umbrella of the Digital Transformation Centre Kenya, through the Lacuna Fund, is supporting the creation of a corpus of Kenyan languages (also KenCorpus) – including Swahili, Luo and Luhya – in close cooperation with Makerere University, Maseno University, University of Nairobi and Africa Nazarene University. Further, in the agriculture sector, there are initiatives targeting similar approaches that will result/promote open, quality, labelled training data that can be used to support climate-smart agricultural practices amongst small holder farmers.

However, open innovation is not without complexities and risks. The risks associated with evolving applications of AI and machine learning, including data standards, privacy considerations, ethical issues, criminal and civil liability, are all properly the subject of regulation. FAIR Forward advocates for value-based AI that is rooted in human rights, international norms such as accountability, transparency of decision-making and privacy.

Objectives: In this session we want to share the work that has been undertaken in the fields of Language technology as well as planned work in Agriculture. The aim is to: 1) highlight the need for open approaches when it comes to sharing lessons, methods up until accruing datasets, 2) advocate for responsible and ethical AI. Both tenets are foundational for the successful building of local AI solutions and applications in the future.

Target audience: Academia – mostly universities teaching and undertaking research on AI topics (language data, images, etc.); AI practitioners from the private and CSO's sectors who are also working on practical AI issues, data protection and ethical AI policies in Kenya.

Format: A Panel discussion/Q&A (first hour) with a breakout session (next hour) where students will also be allowed to come in and show/tell their research work/projects in the field of language technologies and agriculture data.

Where: Physical - University of Nairobi, C4D Lab, Chiromo Campus

When: 9th December, 2021 from 14.30 to 16.30 EAT

Speakers

1. Dr. Lilian Wanzare – Maseno University (Ken Corpus)
2. Dr. Edward Ombui – Africa Nazarene University (Ken Corpus)
3. Britone Mwasaru – Mozilla Foundation (Swahili Commonvoice platform)
4. Dr. Boniface Akuku – Kenya Agricultural and Livestock Research Organization
5. Victor Ndede – Amnesty International, Kenya (Data Rights/Access to Information)
6. Moderator: Muchiri Nyaggah – Local Development Research Institute

Agenda:

- **2:30pm – 2:05pm EAT: Welcome & Introduction**
- **2:05pm – 2:35pm EAT: Opening remarks by each of the panelists (5 min) presentation plus 10 min. Q&A)**
- **2:35pm – 3:15pm EAT: Guided questions by panel moderator & Q/A**
- **3:15pm – 3:30pm EAT: Conclusions and closing remarks by each of the panelists (2 min)**
- **Break out session:**

Contact: For more details, contact mark.gachara@giz.de