A FRAMEWORK FOR GRASSROOTS RESEARCH COLLABORATION IN MACHINE LEARNING AND GLOBAL HEALTH

∀*, Christopher Brian Currin 2,3,4, Mercy Nyamewaa Asiedu 5,6, Chris Fourie 7,10, Benjamin Rosman 7, Houcemeddine Turki 8,9, Atnafu Lambebo Tonja 11,12, Jade Abbott 12,1, Marvellous Ajala 13,16, Sadiq Adewale Aseyado 14, Chris Chinenyi Emezue 12,15, Daphne Machangara 17

* SisonkeBiotik, 2 Deep Learning IndabaX South Africa, South Africa, 3 Computational Neuroscience Imbizo, South Africa, 4 Institute of Science and Technology Austria, Austria, 5 Google Research, USA, 6 Massachusetts Institute of Technology, USA, 7 University of the Witwatersrand, South Africa, 8 University of Sfax, Tunisia, 9 University of the People, USA, 10 LifeQ, USA, 11 Instituto Politécnico Nacional, Mexico, 12 Masakhane, 13 University of Lagos, Nigeria, 14 University of Vienna, Austria, 15 Technical University of Munich, Germany, 16 Cohere For AI, 17 IndabaX Zimbabwe.

ABSTRACT

Traditional top-down approaches for global health have historically failed to achieve social progress [Hoffman et al., 2015; Hoffman & Røttingen, 2015]. Recently, however, a more holistic, multi-level approach termed One Health (OH) [Osterhaus et al., 2020] is being adopted. Several sets of challenges have been identified for the implementation of OH [dos S. Ribeiro et al., 2019], including policy and funding, education and training, and multi-actor, multi-domain, and multi-level collaborations. These exist despite the increasing accessibility to knowledge and digital collaborative research tools through the internet. To address some of these challenges, we propose a general framework for grassroots community-based means of participatory research. Additionally, we present a specific roadmap to create a Machine Learning for Global Health community in Africa. The proposed framework aims to enable any small group of individuals with scarce resources to build and sustain an online community within approximately two years. We provide a discussion on the potential impact of the proposed framework for global health research collaborations.

1 INTRODUCTION

Global health seeks to understand and accommodate the complex systems of our planet-wide society as they relate to health [Salm et al., 2021]. Traditionally, global health institutions (e.g. WHO) have, in a top-down manner, negotiated with the highest levels of government across our world’s nations toward aligned health policy [Salm et al., 2021]. Unfortunately, top-down approaches have produced mixed results. While they succeed in shaping economic matters, they consistently fail to achieve social progress [Hoffman & Røttingen, 2015; Hoffman et al., 2015].

Recently, such institutions have taken steps towards a more holistic, multi-level approach under the banner of the One Health (OH) approach [Osterhaus et al., 2020]. The OH approach strives to mobilize multiple sectors, disciplines, and communities at varying levels of society to work together to foster well-being and tackle threats to health and ecosystems [Adisasmito et al., 2022], with the inclusion of digital health [Benis et al., 2021; Ho, 2022].

However, there remain challenges regarding the implementation of OH and other institution-led communities. Key sets of challenges include policy and funding, education and training, as well as multi-actor, multi-domain, and multi-level collaborations [dos S. Ribeiro et al., 2019]. These exist to represent the whole SisonkeBiotik community. Community as first author. The author list is otherwise randomised and does not represent a level of contribution.
despite the increasing accessibility to knowledge and digital collaborative research tools through the internet. As these technologies are relatively new, their potential for impact is still being explored. For instance, rapid changes in knowledge sharing and curation through community-driven efforts such as the Wikimedia Foundation have led to new economic theories such as Commons-based peer production (Benkler & Nissenbaum, 2006; Bauwens & Pantazis, 2018; Dobusch & Kapeller, 2018). However, even these kinds of initiatives are yet to effectively spread across the African continent. Barriers to entry for those not from the global north still remains a challenge (Graham & Hogan, 2014). Notably for research related to heath, Africa shows higher rates of collaboration than the rest of the world, as measured by academic co-authorship patterns. However, the vast majority of these collaborations are for research outputs with at least one co-author from outside Africa, where exceedingly few collaborations are exclusively African (Pouris, 2017; Pouris & Ho, 2014).

To tackle these key challenges in Global Health (and more specifically in Global Health and Machine Learning), we propose a bottom-up or grassroots community-based means of participatory research to bring together researchers from varying parts of society. Participatory research, unlike conventional research, emphasizes the value of research partners in the knowledge-production process where the research process itself is defined collaboratively and iteratively (English et al., 2018).

In this work, we review some existing Grassroots Participatory Communities (GPCs) and propose a GPC Framework summarising the participatory approaches by Machine Learning communities. We intend this framework to enable any small group of individuals, with scarce resources, to build and sustain an online community within the space of approximately two years. Under this framework, we provide an example roadmap to create a machine learning for the global health community in Africa (ML4GHA) GPC, as a means to alleviate some of the problems highlighted in implementing OH.

2 The Rise of Grassroots Participatory Communities in Machine Learning

We define three types of GPCs, according to the scope and focus of what GPCs have organised around, and categorize examples of existing GPCs into those groups. One should note that this list is geographical-biased towards African communities.

i) Affinity As early as 2016, communities have organised along their affinity to historically underrepresented people in the ML field, hosting workshops, socials, and providing support to individuals in the groups. In this category, we include Women in ML (WiML), Women in ML and Data Science (WiMLDS), Black in AI, LatinX AI, Queer in AI, Indigenous AI, DisAbility in AI and Data Science Nigeria to name a few (WiML, 2023; WiMLDS., 2023; BAI, 2023; LXAI, 2023; QAI., 2023; Disability-EthicalAI, 2023; indigenous ai, 2019; Nigeria)., 2022).

ii) Topic Communities arranged around research topics share similarities to traditional research groups, but the focus is on participation from researchers who may not have had access to traditional research groups. This category includes groups such as Delta Analytics, Mechanism Design for Social Good, ML Tokyo, ML Collective, Active Inference Institute, Ro'ya (Delta-Analitics, 2022; MD4SG, 2023; ML-Tokyo., 2022; ML-Collective, 2023; ActInf, 2023; RoyaCV4Africa, 2022). In particular, there was rapid rise of NLP-focused GPCs, namely Masakhane, EleutherAI, GhanaNLP, North Africans in NLP, Turkic InterLingua, Lanfrica, and Americas NLP (Masakhane, 2023; EleutherAI, 2023; Ghana-NLP, 2023; NLP, 2022; Turkic-InterLingua, 2023; Lanfrica, 2023; AmericasNLP, 2023). Related to healthcare, we note OpenBioML, SisonkeBiotik, RISE-MICCAI(AFRICA), African Society for Bioinformatics and Computational Biology, MedARC (OpenBioML, 2023; SisonkeBiotik, 2023; Lekadir et al., 2022; ASCB, 2023; MedARC, 2023).

iii) Event We define event-focused GPCs as communities arising from volunteer-led events with a community-first grassroots approach. This includes events such as Data Science Africa, The IBRO-Simons Neuroscience Imbizo, NeuroMatch Academy, TReND in Africa, Deep Learning Indaba and Khipu (DSA, 2023; Imbizo, 2023; Neuromatch, 2023; TReNDinAfrica, 2020; DLI, 2023; KHIPU., 2023).
The goal of the following framework is to help enable a small group of individuals, with scarce resources, to build and sustain a topic-based online community within the space of two or so years. The framework consists of sets of values which can underpin community development and maintenance. The framework was developed by analysing common values and priorities of a few GPCs, specifically Masakhane, SisonkeBiotik, and Ro’ya.

**Community First:** We define "community first" as the prioritization of the community above all other objectives (such as publishing research). In practice, this might mean slowing down the pace of research to ensure that the community does it together - as per the African proverb: "If you want to go fast, go alone, if you want to go far, go together". Evidence for this has been observed in Masakhane, where the community-first approach led to the first machine translation systems built for some African languages. This was achieved by enabling individuals who spoke the languages to build their first machine translation models (Nekoto et al., 2020).

**Inclusivity, Collaboration, and Transparency:** A goal of many GPCs are to achieve capacity-building, knowledge sharing, and multi-disciplinarity (Masakhane, 2023; Delta-Analytics, 2022; MD4SG, 2023). To do so, this requires lowering barriers to entry, helping guide newcomers, and helping grow each others’ expertise. Developing peer and group mentorship culture, creating opportunities for idea sharing and feedback and endorsing open science have been shown to be effective at achieving these values (Nekoto et al., 2020). Additionally, being intentional about special invites to under-represented groups has been shown to increase representation from them (Horwitz et al., 2009). Another example of practising inclusion is reconsidering authorship criteria. Many institutions use the Harvard Medical Journal criteria for authorship (Brodrick, 1999). Alternatively, the Masakhane Authorship Model acknowledges each participant’s contribution whether they wrote code, annotated data, did community development, provided analysis or guided the writing journey (Masakhane, 2021).

**Trust, Respect, Kindness, Kinship, and Celebration:** Estrada et al. (2018) describe how particularly in STEM fields policies that affirm social inclusion to all members of the academic population can be helpful in broadening participation. Masakhane (2023); SisonkeBiotik (2023) found that fostering a culture of trust for any GPC members to lead in some capacity can help toward creating a more distributed leadership structure within a GPC.

**Robustness, Sustainability, and Flexibility:** Across the considered GPCs, participation is typically of a voluntary and part-time nature. Masakhane (2023); SisonkeBiotik (2023) found that creating redundancy in leadership can help toward developing robustness to changes in availability and avoid leadership bottlenecks. An example includes developing a culture of embracing emergent practices such that any members have the freedom to experiment with their own ideas on how to run the GPC or parts of the GPC. Other examples include allowing for dynamic role self-assignment and for the space to redefine common practice within the GPC.

**Scalability, Mutual support, Ownership:** Scaling collaborations can be challenging (dos S. Ribeiro et al., 2019). Planning for the impacts of growth and fostering positive relationships with new and existing entities towards mutual support can help grow the research collaboration ecosystem of GPCs for related topics (Masakhane 2023; SisonkeBiotik 2023). Incubation or stewardship practices for new GPCs is an example of this mutual support. One form of incubation is the reciprocal participation of members in the activities of partner communities. Through this reciprocal community mixing, opportunities for providing guidance and support can arise. Another example is the splitting, forking or budding off of a GPC when a subset of members wish to expand beyond the scope of the original GPC. This can help foster the growth of the shared ecosystem while providing a sense of ownership and identity for the new GPC. Another means of stimulating growth for a GPC is to organize large impactful events such as workshops, seminars, hackathons and ideathons (Masakhane, 2023; SisonkeBiotik, 2023; Indaba, 2022; DLI, 2023).

**Participation Roles**

We describe the various roles of participation a GPC can expect and provide examples of engagement with these member groups, using a popular tiered model of community involvement (Orbital-Community, 2023). We highlight these as they differ from traditional research structures.
Building: These are the core members of the GPC that spearhead efforts to lay the foundation and grow the community. Initially, these will likely also be the founding members as the GPC develops its direction. This tier can take on ad-hoc roles according to demand.

Contributing: These are members that will commit to higher levels of participation for a bounded amount of time. This is usually to contribute to a project or an event such as an academic paper or a workshop.

Participating: These are members who engage sporadically for brief and low to moderate levels of participation. Despite their lower levels of participation, this tier is important for extending the reach of the GPC and providing a large network that can help members, for example, find academic collaborators or job opportunities.

Exploring: These are visitors and newcomers to the GPC. Those interested in finding out what the GPC could be for them. It is important to provide a simple and welcoming on-boarding for this tier. It is important to make joining the GPC and getting access to initial points of participation as simple as possible. Avoid requesting lots of details. Provide them with an overview of where the GPC is at and give anchor points for new members to start participating.

ML AND GLOBAL HEALTH FOR AFRICA COMMUNITY ROADMAP

Here we suggest phases a prospective GPC might progress through and activities that have shown to be important through those phases. As a Machine Learning and Global Health for Africa (MLGHA) GPC Roadmap, we provide practical suggestions for each phase, based on empirical evidence from the Masakhane, SisonkeBiotik and Ro’ya communities.

Setup (i) Gather and align a few core participants to build the GPC, and establish a regular online meeting. The Masakhane community found having a simple but standard agenda for this meeting while rotating the meeting facilitator helped toward developing a safe digital space (Nekoto et al., 2020). (ii) SisonkeBiotik found it helpful being incubated by a partner GPC. Embedding members of another GPC in their own GPC helped share knowledge and develop practices. (iii) Establish an identity and brand by defining the scope, values, code-of-conduct and long-term vision of the community, as well as by finding a name and creating a logo. In particular, it can be useful to re-use the values and code of conduct of partner communities. (iv) Set up a community website, chat and meeting platforms, shared calendar, social media and mailing list. When considering trade-offs for software tooling and services, prioritising accessibility helps avoid technical onboarding bottlenecks. (v) Find an initial collaborative task that enables lowering barriers to participation. For Masakhane, this was a machine translation task, for SisonkeBiotik, this was a bibliometric survey. These tasks were well-defined, had preexisting data, and required no additional funding.

The MLGH workshop could be a good opportunity to find core members to seed an MLGHA-GPC and grow existing ones (MLGH-workshop 2023). We propose identifying interested individuals from workshop participants to initiate regular weekly meetings for this purpose. The incubation of members as well as mentorship and guidance can be sourced from the Grassroots Participatory Community Collective Africa (grassroots part, 2023). We suggest starting with a survey study on the GPC’s focus. For example, SisonkeBiotik started early with a bibliometric study of ML and Global Health in Africa. These help the community understand the landscape of their chosen topic and identify future problems to work on, and fit the above criteria.

Growth: A combination of regular smaller activities and less regular large activities have been successful at growing GPCs: For example, ensuring that at least one project is running at any time provides an opportunity for newcomers to contribute. Ro’ya started with a regular reading club, while Lanfrica and SisonkeBiotik invited individuals to do online seminars. Larger less regular events might include running workshops (A-NLP 2023), seminars, competitions (Siminyu et al., 2021), or hackathons, and these provide large-scale visibility for the GPC.

ML4GHA-GPC: We recommend an online seminars series to bootstrap a MLGHA-GPC. Speakers could be sourced from the MLGH workshop, and using the existing network from incubating GPCs will help reach an initial audience. Evidence from the SisonkeBiotik community suggests that those with biomedical and healthcare backgrounds tend to be less well-represented in online activities,
so actively inviting members from backgrounds that are underrepresented is important. For larger engagements, we recommend applying to existing conferences to host workshops on the topic.

**Maintenance:** After community momentum has been gained, maintenance and longevity of the community will become important. For Masakhane, this meant the creation of the Masakhane Research Foundation. This allowed them to hold funds for the Masakhane community. To do so, specific governance structures needed to be set up, the creation of a board, succession plans for that board, as well as decision-making processes on the use of funds.

**ML4GHA-GPC:** We recommend only establishing a legal entity after the community has stabilized so that the constraints and needs are properly understood. Codifying parts of the community need to be done with utmost care and inclusivity, to avoid community harm (Kostakis 2010).

## 4 Discussion and Conclusion

Some challenges in complex social systems such as research collaborations may never be fully solved. The GPC framework addresses and could potentially help to alleviate various implementation challenges encountered in a One Health approach (dos S. Ribeiro et al., 2019). Considering these by challenge category:

**Policy and Funding:** To address the *Lack of resources and funding for OH initiatives*, dos S. Ribeiro et al. (2019) proposes a number of solutions to increase funding but forgo suggestions to make better use of existing funds. GPCs can be extremely lean with regard to funding and for the most part, do not require any funding to be initiated or to maintain function adequately. The conditions required to initiate a GPC can be as minimal as 3-5 people organising online around a shared topic of interest.

**Education and training:** A key feature of GPCs is iteratively developing a culture of peer-to-peer knowledge sharing, regardless of member backgrounds. Members are not required to have any institutional affiliations.

**Multi-actor and multi-domain collaborations:** By shifting the main focus of collaboration away from academic productivity and prioritizing positive social values such as community-building and breaking down *disciplinary and cultural silos*, GPCs create spaces that encourage more interdisciplinary and fluid collaboration. All observed GPCs are addressing *Difficulties to promote and sustain OH collaborations*, through ongoing collaborations and with active efforts toward fair representation. The distributed nature of GPC leadership and project coordination helps to alleviate the challenges related to *Lack of facilitated collaborative process and Unequal power/representation of actors*.

**Multi-level collaborations:** To help alleviate *Institutional and academic fragmentation* GPCs offer an independent space that values neutrality with regard to the institution and academic domain. Toward addressing *Geographic and cultural fragmentation* developing interpersonal rapport in an online setting is a fundamental requirement of GPCs and helps overcome geographic boundaries. The intersecting and non-mutually exclusive nature of community membership in a GPC ecosystem allows for GPCs to form along geographic and cultural dimensions and then interact and mix along other dimensions.

Optimising research collaborations will require an ongoing journey of experimentation and iteration. The One Health approach is a milestone in this journey for global health. Aligned with this approach, here we propose GPCs as a means to implement accessible and collaborative spaces for machine learning and global health research.

## Acknowledgments

Houcemeddine Turki’s contributions to this final output have been funded through the *Adapting Wikidata to support clinical practice using Data Science, Semantic Web and Machine Learning* project, which is part of the Wikimedia Research Fund maintained by the Wikimedia Foundation in San Francisco, California, United States of America.
REFERENCES


BAI. Black in AI, 2023. URL https://blackinai.github.io/#/about


Disability-EthicalAI. Disability ethicalai, Feb 2023. URL https://disabilityethicalai.org/about/


Neuromatch. Neuromatch Academy, 2023. URL [https://neuromatch.io]

Data Scientist Network (Data Science Nigeria). Data Science Nigeria is leading Nigeria’s Artificial intelligence, Big Data… 2022. URL [https://www.datasciencenigeria.org/]

OpenBioML. OpenBioML - Democratizing Scientific Research, 2023. URL [https://openbioml.org/]

Orbital-Community. Orbit Levels, Jan 2023. URL [https://orbit.love/model/love/orbit-levels]


QAI. Queer in AI, 2023. URL [https://www.queerinai.com]


TReNDinAfrica. Courses — TReND in Africa., 2020. URL [https://trendinafrica.org/courses/]

Turkic-InterLingua. Turkic InterLingua, 2023. URL [https://turkic-interlingua.org/]

WiML. Women in Machine Learning, 2023. URL [https://wimlworkshop.org/mission/]

WiMLDS. Women in Machine Learning and Data Science, 2023. URL [http://wimlds.org/].