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

Anonymous authors
Paper under double-blind review

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 shaping in 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. Key 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). This is despite the increasing accessibility to knowledge and digital research tools through the internet.

To tackle those 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 Grassroots Participatory Community 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 2 or so years. Under this framework, we provide a recommended roadmap to create a machine learning for global health community in Africa (ML4GHA) GPC, as a means to alleviate some of the problems highlighted in implementing OH.

1 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 De-

sign for Social Good, ML Tokyo, ML Collective, Active Inference Institute, Ro'ya, MLOps Africa (Delta-Analytics., 2022; MD4SG, 2023; ML-Tokyo., 2022; ML-Collective, 2023; Active-Inference-Institute., 2023; RoyaCV4Africa, 2022; MLOps-Africa, 2021). 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; Eleuther-AI, 2023; Ghana-NLP, 2023; in NLP., 2022; Turkic-InterLingua, 2023; Lanfrica, 2023; AmericasNLP, 2023). In Healthcare, we note SisonkeBiotik, and OpenBioML (SBiotik, 2023; OpenBioML, 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, Deep Learning Indaba and Khipu (Africa., 2023; Imbizo, 2023; Neuromatch, 2023; in Africa., 2020; DLI, 2023; KHIPU., 2023).

2 Grassroots Participatory Community Framework for Research

The goal of the following framework is to help enable small group of individuals, with scarce resources, to build and sustain a topic-based online community within the space of 2 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 is 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 authorshipBrodrick (17AD). 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.

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); SBiotik (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); SBiotik (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 has 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; SBiotik, 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 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; SBiotik, 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 (Community, 2023). We highlight these as they differ from traditional research structures.

Building: These are the core and initially the founding members of the GPC. They will spearhead efforts to lay the foundation and grow the community. Initially, as the GPC is still developing 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 experiential evidence from the Masakhane and SisonkeBiotik 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 having a partner GPC incubate their GPC. By 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, meeting platform, 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 a MLGHA-GPC(MLGH-workshop, 2023). We propose identifying interested individuals from workshop participants to initiate regular weekly meetings toward this purpose. Incubating members and guidance to be sourced from the *Grassroots Participatory Community Collective Africa* (grassroots parti, 2023). As an initial task, we suggest a bibliometric or survey study, with the goal to investigate 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 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 workshopsn(A-NLP), 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 recommended 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).

3 Conclusion and discussion

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) propose a number of solutions to increase funding but forgo suggestions to make better use of existing funds. GPCs can be extremely lean with regards 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: Iteratively developing a culture of peer-to-peer knowledge sharing for members from arbitrary contexts is a key feature of GPCs. Members are not required to have any institutional affiliations.

Multi-actor and multi-domain collaborations: Considering *Disciplinary and cultural silo thinking*, community first and other positive social values of GPCs, by moving the main collaborative focus off academic productivity, help to foster spaces for more fluid and interdisciplinary collaboration. On *Difficulties to promote and sustain OH collaborations*, in all GPCs considered we observed ongoing sustainable collaborations and 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 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.

REFERENCES

A-NLP. Africanlp 2023. URL https://sites.google.com/view/africanlp2023.

Active-Inference-Institute. Active inference institute., 2023. URL https://www.activeinference.org/.

Wiku Adisasmito, Salama S. Almuhairi, Casey Barton Behravesh, Pepe Bilivogui, Salome A. Bukachi, Natalia Casas, Natalia Margarita Cediel Becerra, Dominique F. Charron, Abhishek Chaudhary, Janice R Ciacci Zanella, Andrew A. Cunningham, Osman Dar, Nitish Debnath, Baptiste Dungu, Elmoubasher Farag, George F. Gao, David T. S. Hayman, Margaret L. Khaitsa, Marion P. G. Koopmans, Catherine Machalaba, John S. Mackenzie, Wanda Markotter, Thomas C. Mettenleiter, Serge Morand, Vyacheslav Smolenskiy, and Lei Zhou. One health: A new definition for a sustainable and healthy future. *PLoS Pathogens*, 18, 2022.

DataScience Africa. Datascience africa., 2023. URL https://www.datascienceafrica.org/.

AmericasNLP. Second workshop on nlp for indigenous languages of the americas (americasnlp), 2023. URL http://turing.iimas.unam.mx/americasnlp/index.html.

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

Arriel Benis, Oscar Tamburis, Catherine E. Chronaki, and Anne Moen. One digital health: A unified framework for future health ecosystems. *Journal of Medical Internet Research*, 23, 2021.

Melissa Brodrick. Harvard medical school authorship guidelines, Dec 17AD.

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

Delta-Analytics. Delta analytics., 2022. URL http://www.deltanalytics.org/.

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

DLI. Deep learning indaba, 2023. URL https://deeplearningindaba.com/about/our-mission/.

Carolina dos S. Ribeiro, Linda H.M. van de Burgwal, and Barbara J. Regeer. Overcoming challenges for designing and implementing the one health approach: A systematic review of the literature. *One Health*, 7, 2019.

Eleuther-AI. Eleuther ai., 2023. URL https://www.eleuther.ai/.

Paul B English, Maxwell J Richardson, and Catalina Garzón-Galvis. From crowdsourcing to extreme citizen science: participatory research for environmental health. *Annual review of public health*, 39:335–350, 2018.

Mica Estrada, Alegra Eroy-Reveles, and John Matsui. The influence of affirming kindness and community on broadening participation in stem career pathways. *Social issues and policy review*, 12(1):258–297, 2018.

Ghana-NLP. Ghana natural language processing (nlp) - ghana nlp, 2023. URL https://ghananlp.org/.

grassroots parti. Grassroots participatory community collective, Feb 2023. URL https://www.grassroots-parti.africa.

Calvin Wai-Loon Ho. Operationalizing "one health" as "one digital health" through a global framework that emphasizes fair and equitable sharing of benefits from the use of artificial intelligence and related digital technologies. *Frontiers in Public Health*, 10, 2022.

Steven J. Hoffman and John-Arne Røttingen. Assessing the expected impact of global health treaties: Evidence from 90 quantitative evaluations. *American journal of public health*, 105 1:26–40, 2015.

Steven J. Hoffman, John-Arne Røttingen, and Julio Frenk. Assessing proposals for new global health treaties: An analytic framework. *American journal of public health*, 105 8:1523–30, 2015.

Susan Horwitz, Susan H Rodger, Maureen Biggers, David Binkley, C Kolin Frantz, Dawn Gundermann, Susanne Hambrusch, Steven Huss-Lederman, Ethan Munson, Barbara Ryder, et al. Using peer-led team learning to increase participation and success of under-represented groups in introductory computer science. *ACM SIGCSE Bulletin*, 41(1):163–167, 2009.

Imbizo. Computational neuroscience imbizo, 2023. URL https://imbizo.africa/.

TReND in Africa. Courses — trend in africa., 2020. URL https://trendinafrica.org/courses/.

North-Africans in NLP. North africans in nlp., 2022. URL https://sites.google.com/view/NorthAfricansInNLP.

Deep Learning Indaba. Indabax - deep learning indaba, 2022. URL https://deeplearningindaba.com/2022/indabax/.

indigenous ai. Indigenous ai., 2019. URL https://www.indigenous-ai.net/.

KHIPU. Khipu., Feb 2023. URL https://khipu.ai/.

Vasilis Kostakis. Peer governance and wikipedia: Identifying and understanding the problems of wikipedia's governance. *First Monday*, 2010.

Lanfrica. Lanfrica., 2023. URL https://lanfrica.com/records.

LXAI. Latinx in ai, 2023. URL https://www.latinxinai.org.

Masakhane. Authorship. URL https://www.masakhane.io/authorship.

Masakhane. Masakhane., 2023. URL https://www.masakhane.io/.

MD4SG. Mechanism design for social good., 2023. URL https://www.md4sg.com/.

ML-Collective. Ml collective., 2023. URL https://mlcollective.org/.

ML-Tokyo. Machine learning tokyo., 2022. URL https://www.mlt.ai.

MLGH-workshop. Machine learning global health workshop, Feb 2023. URL https://mlgh-2023.netlify.app/.

MLOps-Africa. Mlops africa., 2021. URL https://www.mlops.africa.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, et al. Participatory research for low-resourced machine translation: A case study in african languages. *arXiv preprint arXiv:2010.02353*, 2020.

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/.

Albert D M E Osterhaus, Chris Vanlangendonck, Maurizio Barbeschi, Christianne J. M. Bruschke, René Flamsholt Christensen, Peter Daszak, Frouke de Groot, Peter Doherty, Patrick Drury, Sabri Gmacz, Keith Hamilton, John Hart, Rebecca Katz, Christophe Longuet, J Mcleay, Gaetano Morelli, Joergen Schlundt, Trevor Frise Smith, Sameera Suri, Khristeen Umali, Jan van Aken, and Jaap A. Wagenaar. Make science evolve into a one health approach to improve health and security: a white paper. *One Health Outlook*, 2, 2020.

QAI. Queer in ai., 2023. URL https://www.queerinai.com.

RoyaCV4Africa. Roya-cv4africa-community., 2022. URL https://github.com/Ro-ya-cv4Africa/Roya-CV4Africa-Community.

Melissa Salm, Mahima Ali, Mairead Minihane, and Patricia A. Conrad. Defining global health: findings from a systematic review and thematic analysis of the literature. *BMJ Global Health*, 6, 2021.

SBiotik. Sisonkebiotik., 2023. URL https://www.sisonkebiotik.africa/home.

Kathleen Siminyu, Godson Kalipe, Davor Orlic, Jade Abbott, Vukosi Marivate, Sackey Freshia, Prateek Sibal, Bhanu Neupane, David I Adelani, Amelia Taylor, et al. Ai4d–african language program. *arXiv preprint arXiv:2104.02516*, 2021.

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/.