Inclusive Datasets for Artificial Intelligence

Sexual, Reproductive and Maternal Health and Rights

Lacuna Fund RFP Webinar

23 March 2023

Objectives

- Provide potential applicants with a detailed understanding of RFP requirements, including budget and allowed costs.

- Introduce the mentorship opportunity with The Deep Learning Indaba

- Technical Advisory Panel (TAP) members, the Lacuna Fund Secretariat, and The Deep Learning Indaba will provide answers to applicant questions and document them for the written Q&A.

A lacuna is a gap, or a missing portion, of a manuscript or piece of music.

Lacuna Fund was formed to provide resources to data scientists in low- and middle-income contexts to fill the gaps in ML datasets needed to solve urgent problems in their communities.

Funders

Structure

- **Steering Committee**: Strategic direction

- **Technical Advisory Panels**: RFP design and proposal selection

- **Secretariat**: collaboration, convening, facilitation, communication, and logistics

Principles

- Accessibility
- Equity
- Ethics

- Participatory Approach
- Quality
- Transformational Impact

Funding

Open solicitation process (Requests for Proposals) in different domains:
- Agriculture
- Language
- **Health**
- Climate
- Other areas where the Steering Committee identifies a need

Lacuna Fund provides resources for dataset creation, expansion, or maintenance.

Requirements for Sexual, Reproductive and Maternal Health and Rights RFP

What We Are Looking For

ML Datasets!

- Collecting and/or annotating new data

- Annotating, augmenting, or releasing existing data; quality validation of existing datasets

- Linking, harmonizing, and adapting existing datasets

Proposal Components

- Applicant information
- Proposal narrative
- Project timeline
- Budget

Key Dates

24 March: Question and Answer deadline

5 April: Answers posted

Round 1:
19 April: Full Proposals due
June: Technical Advisory Panel makes decisions on grant awards
July: Grant Awardees are notified

Round 2:
20 June: Full Proposals due
August: Technical Advisory Panel makes decisions on grant awards
October: Grant Awardees are notified

Elements of Strong Proposals

Elements of Strong Proposals

- **Multi-disciplinary team** with experience in data science and SRMHR

- **Use cases and community engagement**

- Clear **problem statement** and how the proposed dataset or aggregation will help solve the problem

- Specificity about dataset size – should be of **sufficient size and quality** to be useful in future applications

- **Budget** is appropriate for the size of the dataset produced

Elements of Strong Proposals

- **Equity** considerations (gender, socio-economic status, ethnicity, etc.)

- Consideration and plan for potential **privacy and ethics** issues, including compliance with national laws

- Plan for **data management and licensing**

- Plan for **dataset hosting and documentation**

- Working across areas where possible and relevant: time, diseases, regions, climate events

Community Engagement

- Describe previous consultation and/or proposed collaboration with intended beneficiaries.
  - When and where you have met/ will meet with partners
  - How partners will be involved in:
    - Identifying data needs
    - Data collection, labeling
    - Data governance
    - Dataset use
  - How partners will benefit from the new/expanded dataset

Privacy and Ethics

Identify issues and potential risks, and explain how your team will address:

- Communities and identities
- Informed consent
- Ensure equity in project labor (e.g., fair compensation in labeling)
- Diversity in team composition
- Plan for anonymization of PII and compliance to laws
- Risk of adverse impacts and downstream misuse

Describe the process you will use to screen for potential ethical issues (e.g., an institutional review board, etc.).

Data Standards and Sharing

- **F**indable - easy to find on a public, widely used platform
- **A**ccessible - open format (CCBY 4.0)
- **I**nteroperable - data format
- **R**eusable – metadata and data well described

- Sustainable - plan for maintenance
- Shared – engaged community of users & plan for sharing dataset once complete

https://www.go-fair.org/fair-principles/

What relevant platforms are you seeing for storing and sharing data in your domain?

Please answer in the chat.

Data Standards and Sharing

Health care electronic data exchange standard:

- **F**ast
- **H**ealthcare
- **I**nteroperability
- **R**esources

https://fhir.org

Budget and Allowed Costs

- Must be submitted in the Excel template.

- When in doubt, explain!

- Allowed costs include:
  - Training/capacity building
  - Data collection, labeling, QA/QC or verification, and post-processing
  - Licensing and dataset hosting
  - Open access publication of results
  - Metadata or data statement
  - Crowd-sourcing efforts, such as label-a-thons
  - Baseline model
  - Competition or hackathon

- In-kind storage and computing power might be available from partners Google Cloud and Amazon Web Services

Sexual, Reproductive and Maternal Health & Rights Request for Proposals

Caitlin Augustin, DataKind

Why SRMRH Datasets?

Goal: Open and accessible datasets for equitable, unbiased, and responsible machine learning applications to improve Sexual, Reproductive and Maternal Health & Rights (SRMHR) outcomes in Africa.

SRMHR Topics

- Family Planning
- Abortion
- Maternal and newborn health
- Menstrual health
- Non-communicable diseases of the reproductive system
- Sexual and Gender-Based Violence
- Sexually Transmitted Infections, including HIV

...and others

SRMHR Datasets Needs

- **Q&A datasets containing high quality, evidence-based medical information**

- **Image datasets**

- **Large population datasets** (including longitudinal)

- **Datasets representing the experience in the treatment process**

- Across all datasets: **gender-responsiveness and inclusion of key vulnerable and often excluded groups**

SRMHR Datasets Needs

... and many more!

These are just a few ideas and key points.

The TAP and Lacuna Fund welcome proposals outside these topics.

A few additional considerations

- **Work towards the harmonization and standardization of SRMHR data.** Adopting the Fast Healthcare Interoperability Resources (FHIR) is encouraged where relevant and possible

- **The ethics of accessing and handling SRMHR data**, including informed consent and privacy considerations at the community and individual health level

- **Political sensitivities** towards SRMHR topics in some contexts

- Rapidly evolving **data privacy laws** in Africa

The Indaba Mentorship Programme

- Mission: to strengthen Machine Learning and Artificial Intelligence in Africa and works towards the goal of Africans being not only observers and receivers of the ongoing advances in AI, but active shapers and owners of these technological advances.

- On a **one-time basis** for this Lacuna Fund call, offering applicants a special opportunity to **join the Deep Learning Indaba community** and match those who are interested with a **mentor who will review your draft** proposal and discuss avenues for strengthening it.
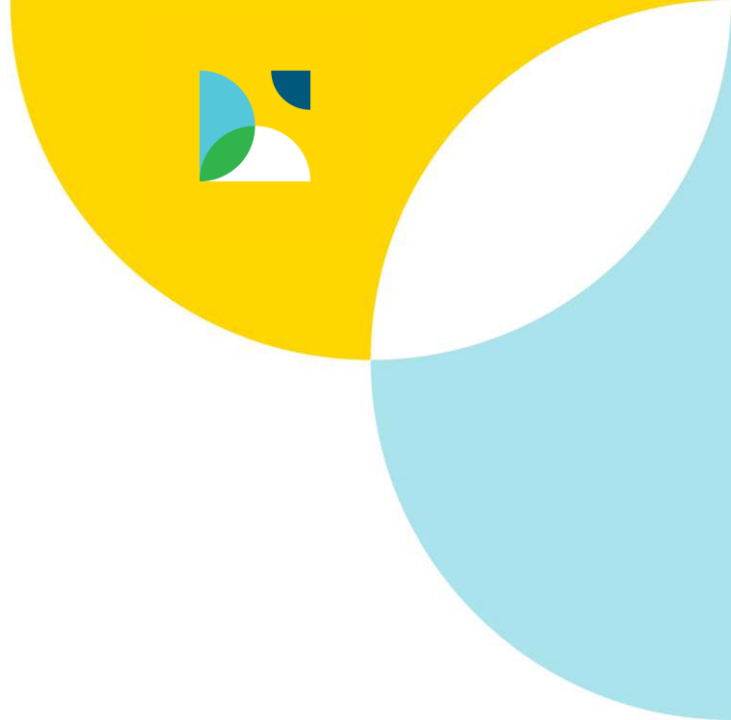
DEEP LEARNING
INDABA

The Indaba Mentorship Programme

- Apply on **Deep Learning Indaba website** https://deeplearningindaba.com/mentorship

- Select to participate **"As a mentee" > "Writing a research proposal"**
  - In the field "Please tell us more about your draft proposal" type **LACUNA FUND**, and a **detailed description of your proposed dataset**.

- We encourage you to prepare a detailed outline or draft proposal before applying for mentorship

- Apply **at least 4 weeks before the due date**

- Mentors will be assigned on a **first come, first-served** basis.

- All applicants are expected to **read and abide by the mentorship programme's code of ethics and conduct**.

DEEP LEARNING INDABA

Questions?

Next Steps

- **Submit additional questions to secretariat@lacunafund.org by 24 March**

- **Answers posted** publicly on Lacuna Fund Apply page on **5 April**

- Proposals due on:
  - **19 April** (Round 1)
  - **20 June** (Round 2)