ARTIFICIAL INTELLIGENCE FOR OPERATIONAL SUPPORT TO THE FORCIBLY DISPLACED

Challenge poser: UNHCR, the United Nations (UN)
Location: Remote

I. ABOUT ORGANIZATION

UNHCR, the UN Refugee Agency, is a global organization dedicated to saving lives, protecting rights and building a better future for refugees, forcibly displaced communities and stateless people. We work to ensure that everybody has the right to seek asylum and find safe refuge, having fled violence, persecution, war or disaster at home. Since 1950, we have faced multiple crises on multiple continents, and provided vital assistance to refugees, asylum-seekers, internally displaced and stateless people, many of whom have nobody left to turn to. We help to save lives and build better futures for millions forced from home.

1.1. About challenge poser team:

UNHCR's Innovation Service is committed to creating an enabling environment for innovation to flourish in UNHCR by equipping staff with the knowledge, resources, and skills needed to ensure that they can increasingly draw on structured innovation to solve the most pressing challenges. We create safe spaces for experimentation to take place in field operations, as well as at Headquarters, whilst instilling a future-oriented approach into how we solve problems.

II. BACKGROUND RESEARCH

2.1 UNHCR Research

Project Jetson (2017)

Project Jetson, the first AI-based predictive analytics project in UNHCR was developed back in 2017 to estimate population flow movement, particularly IDPs and refugees movement cross-borders. The collaboration with the computer vision department, at the University of Essex, specifically Human RIghts and Big Data and Technology Initiative (HRBDT) focused on the nexus between climate change/weather anomalies, population flow movement, and conflict data. More specifically the second part of the research was to produce more accurate climate historic indicators for the region of Somalia of drought/flood conditions and correlate them with different types of violent conflict. The repository was developed to execute a specific collection of satellite images, process the necessary bands, and formulas to calculate the indicators, based on coordinates. Omdena Foundation also worked on exploratory data analysis on Project Jetson, via its local chapters: Documentation

Brazil Predictive Analytics (2020)

UNHCR in Brazil, in collaboration with UN Global Pulse New York Lab and the UNHCR Innovation Service, developed this tool so that humanitarian workers can understand population flow movements - in this case, from Venezuelans to Brazil. The initiative began during the COVID-19 pandemic, when a series of public policy changes (e.g., border closures) affected the flow of refugees and migrants at the northern border. The project consists of three elements that support UNHCR in Brazil - and potentially other UNHCR operations in different countries - to: 1) understand the factors affecting real-time population flow (nowcasting), push and pull factors, and other elements that could be indicators of changes in population flow; 2) estimate future flows based on these factors (forecasting, predictive analysis); and 3) support contingency planning scenarios, providing numbers that can help operations plan their services or activities (simulation), in the case of UNHCR in Brazil, this is the concrete operationalization for emergency shelter construction planning. For more information about the project, see: Modeling Population Movements under Uncertainty at the Border in Humanitarian Crises: A Situational Analysis Tool.

2.2 UN Global Pulse Research

Pulse Satellite Project (2020)

UN Global Pulse is Secretary General's Innovation Lab, UN Global Pulse supports responsible and inclusive innovation across the UN system. Their mandate is to explore the spectrum of innovations using data, digital, behavioral science, and strategic foresight methods – from broadening digital inclusion for women in Indonesia to predicting the impact of a disaster in the Philippines or building out a national data strategy in Uganda. UN Global Pulse has built PulseSatellite to reduce the time needed for UN users to acquire, process, and load satellite imagery into machine learning models, making it easier for users to take advantage of these automated tools. PulseSatellite also offers human-in-the-loop capabilities for improving these models, allowing users to correct them and increase their performance over time. We are working to extend PulseSatellite's capabilities to include damage detection as part of the Data Insights for Social & Humanitarian Action (Disha) initiative. More documentation here: paper

III. ABOUT THE CHALLENGE

3.1 Streams

The following are research streams suggested for analytics, to focus the scope of their research questions and the datasets to be used. These are just suggestions but the researchers can:

- Expand on adding different geographies and datasets within the same stream and/or
- Combine streams to find interesting trends (e.g. EDA, correlations, etc) among datasets in each stream. Some are shared, some are different.

The datasets are focused on a specific country (Somalia) that cover the different legal status of a person who is forced to flee:

1. Internally displaced people (IDPs) people who are forced to flee within the borders of their country of origin (CoO).
2. And once crossing the border of the neighboring countries (countries of asylum, CoA) Somalis are considered in some of the neighboring countries as, prima facie refugees;
3. As well as people with voluntary return (returnees) and/or pendular movement.

3.2 Ethical Considerations for Humanitarian Action

The streams should focus also on the humanitarian ethics - a specific sub-section of ethics - considerations of each of the modeling approaches as it has both operational implications and most importantly, on the agency of the people it is intended to target. In addition to that the overarching data science research/modeling questions, researchers should consider the following policy/ethical questions:

- What is the policy problem that is being tackled?
- What could be a suggested humanitarian action and what are the implications for the suggested action
    - Suggestion to read part 3 of Hugo Slim book on humanitarian decision-making as well as ALNAP webinar, this paper about modeling in humanitarian emergencies and this ICRC blog on AI) and
- What are the consequences of inaction?
- Is Anonymisation Enough To Protect The Data Privacy Of People Affected By Crises?
- Are machine learning technologies ready to be used for humanitarian work and development?
- Are there any sampling bias? for example this research on climate and conflict?
- Take into consideration also the Principles for the Ethical Use of Artificial Intelligence in the United Nations System
- Understanding the climate-conflict nexus from a humanitarian perspective: a new quantitative approach

Finally, it is suggested to team up with multi-disciplinary profiles and backgrounds: data scientist, computational social scientists, computer developers, anthropologists, sociologists, design, political scientists and other profiles that can contribute to the development of the research. We encourage research to use the language of any natural/programming language of their preference for presenting their preliminary results, ideally with well-documented notebooks.

STREAM 1: Linear ML modeling, predictive analytics

This stream of work will be focused on the quantitative multivariate analysis of different factors that exacerbate historical movement (forced displacement) vis-à-vis voluntary migration movements in Somalia. Some of the factors are not push and pull factors (e.g. market prices/commodities) but they are extremely sensitive to changes in weather/climate that contributes to the ongoing/underlying larger movements that might not be captured due to lack of humanitarian access and/or capacity to collect data on the ground.

Potential data science research/modeling questions

- How many people are moving from one region/country to another and to a final destination? (e.g. point A, settlement/region/country of origin → point B transit → point C, final destination/CoA)
- What are the regions/countries they are moving towards?
- Approximately, when are they expected to arrive?
- How long does it take for them to arrive safely to a place?
- What are some of the drivers for their displacement? (push and pull factors, sensitivity analysis)
- Prediction of specific vulnerable groups fleeing/numbers or regions (e.g. children)?
- Correlation among the above-mentioned questions/variables

The research teams would need to build their own parsers for some of the following datasets.

Datasets and resources: triple nexus

Humanitarian sector datasets: population datasets

- UNHCR PRMN (IDPs). Read methodology notes: "*The network does not capture all population movements in all areas of Somalia at all times – reports of displacement figures can be seen as indicators of potentially larger movements and their underlying causes*"
- IDMC (IDPs)
- UNHCR operational data portal (refugees)
- OCHA HDX (many)
- OCHA catalog of predictive analytics projects

Development sector datasets: food security, natural resources, climate data

- FSNAU: commodities affected by changes in climate: local goat prices, water drum price
- IPC (food security)
- SWALIM: water resources, different datasets, including this research
- FEWS
- Climate in Somalia, *special attention to the references of Gu and Dayr. If a season fails, people are at risk of lacking food security*

Peacebuilding sector datasets: violent conflict

- Conflict in Somalia: historical highlights
    - ACLED API user manual
    - ACLED API code
- Uppsala University: VIEWS project and API access

Other datasets: private sector

- Google Distance Matrix API and Google Earth (research paper)
- Meta marketing API (following this methodology, and this repo)
- Meta Data for good: population during crises
- Meta social connectedness index
- Meta commuting zones

Ethical and Human rights considerations: We recommend reading out the following reading for researchers to take into account when building regression/ML quantitative models.

- Explicability of humanitarian AI: a matter of principles
- Cognitive biases in humanitarian sensemaking and decision-making lessons from field research
- Predictive modeling of movements of refugees and internally displaced people: towards a computational framework
- Harnessing Data Innovation for Migration Policy: A Handbook for Practitioners (Chapter 5)
- Cambridge University - Communicating Uncertainty Project
- Developments and Lessons Learned in Humanitarian Innovation for Forced Displacement

STREAM 2: Computer Vision, satellite imagery and Agent-based modeling (ABM)

This stream of work will be focused on the geographical information systems (GIS) analysis and/or build computer vision models and/or build agent-based modeling to understand the patterns for forced displacement movement, as well as the influence of different factors that exacerbate historical movement (forced displacement) vis-à-vis voluntary migration movements in Somalia. Some of the factors are not push and pull factors (e.g. market prices/commodities) but they are extremely sensitive to changes in weather/climate that contributes to the ongoing/underlying larger movements that might not be captured due to lack of humanitarian access and/or capacity to collect data on the ground.

Potential data science/modeling research questions

- How many people are moving from one region/country to another and to a final destination? (e.g. point A, settlement/region/country of origin ➔ point B transit ➔ point C, final destination/CoA)
- What are the regions/countries they are moving towards?
- Approximately, when are they expected to arrive?
- How long does it take for them to arrive safely to a place?
- What are some of the drivers for their displacement? (push and pull factors, sensitivity analysis)
- What human rights (protection-related) incidents can be described using computer vision?
- Prediction of specific vulnerable groups fleeing/numbers or regions (e.g. children)?
- Correlation among the above-mentioned questions/variables

Datasets and resources: triple nexus

Humanitarian sector datasets:

- UNHCR GIS maps: population, official administrative boundaries, UNHCR presence
  - UNHCR Somalia admin boundaries
  - ArcGIS Somalia admin boundaries
  - ArcGIS Somalia district boundaries
- UNOSAT
- Humanitarian Open Street Maps (HOTSOM) export tool
  - Shape files HOTSOM, and the old export tool HOTSOM
- OpenStreetMap Tool

Other non-related to Somalia but relevant modeling approaches

- Drought Times Series Classification Using Normalized Difference Vegetation Index (NDVI) 2000-2011 **(Kenya, CoA)**
- An Analysis of the Impacts of Ongoing Drought across the Eastern Horn of Africa
- UNICEF Arm 2030 Vision #1: Flood Prediction in Malawi
- UNICEF Arm 2030 Vision #2: Malawi Floods Data Visualization and Reporting Challenge

Development sector datasets: food security, natural resources, climate data

- UNHCR Project Jetson NDVI calculation
- Somalia NDVI conditions
- FEWS
- Group of Earth Observations (GEO)
    - Knowledge resources
- LANDSAT including Mogadishu urban extent
- Drought monitoring tool: FAO and FAO data in emergencies
- Research on AI water and conflict
- DINA Somalia map

Peacebuilding datasets: violent conflict

- Conflict in Somalia: historical highlights
    - ACLED API user manual, particular focus to the columns with lat/lot (GIS) coordinates
    - ACLED API code
- Uppsala University: VIEWS project and API access, particular focus to the columns with lat/lot (GIS)

Academic research toolkits/datasets:

- FLEE toolkit (Brunel University)
- Climate Data Library
- Vegetation Stress Monitor—Assessment of Drought and Temperature-Related Effects on Vegetation in Germany Analyzing MODIS Time Series over 23 Years
- Measures of Vegetation: tool produces maps of estimated vegetation

Call Detail Records (CDRs)

- CDR datasets
    - Flowkit (Flowminder anonymization kit)
    - Assessing Refugees' Integration via Spatio-Temporal Similarities of Mobility and Calling Behaviors
    - Towards an Understanding of Refugee Segregation, Isolation, Homophily and Ultimately Integration in Turkey Using Call Detail Records
- Mobile Lab University Tartu
    - Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones

Other GIS related datasets and toolkits

- LinkedIn Economic graph, case study for mobility in Europe
- Google Distance Matrix API
- Meta marketing API (following this methodology, and this repo)
- Meta Data for good: population during crises
- Meta social connectedness index
- Meta commuting zones

- UC Merced Land Use Dataset
- Open Images V6
- DODS: tool allows one to give a application a url and have it directly download data
- Tutorial: NDVI from Landsat 8 in SNAP
- espa-bulk-downloader - Retrieves all completed scenes for the user/order and places them into the target directory
- Rasterio - reads and writes GeoTIFF and provides a Python API based on Numpy N-dimensional arrays and GeoJSON
- scikit-image - a.k.a. skimage is a collection of algorithms for image processing and computer vision
- matplotlib - object-oriented plotting library
- earthpy - EarthPy is a python package devoted to working with spatial and remote sensing data
- geopandas - GeoPandas is an open source project to make working with geospatial data in python easier

Ethical and Human rights considerations: We recommend reading out the following reading for researchers to take into account when building GIS/computer vision models, ABM and other computer simulation modeling.

- Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping
- A generalized simulation development approach for predicting refugee destinations
- Sensitivity-driven simulation development: a case study in forced migration
- How Policy Decisions Affect Refugee Journeys in South Sudan: A Study Using Automated Ensemble Simulations
- Explicability of humanitarian AI: a matter of principles
- GSMA guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak
- Epidemic Modeling in Refugee Settlements To Detect Disease Spread (video)
    - Operational response simulation tool for epidemics within refugee and IDP settlements: A scenario-based case study of the Cox's Bazar settlement

STREAM 3: Generative AI (GAI) and NLP, text-based data

With the evolution in research of natural language processing (NLP) and the increased capacity of computers to be able to correctly apply techniques for multi-class, multi-label categorization of text-based information, we can help humanitarian organizations respond in a timely and effective manner to humanitarian crises. Humanitarian organizations require quick and accurate analysis of large amounts of text data, a process that can highly benefit from expert-assisted NLP and nowadays Generative AI systems.

Potential data science/modeling research questions

- IOM social media case studies : quantifying mobility/movement patterns using social media or mapping geographies, vulnerable populations groups
- Quantification/summarization of incidents of misinformation, disinformation and/or hate speech (MDH) or xenophobia towards people who are forced to flee both in a) country of origin - that causes them to flee or country of asylum, that prevents them to be fully integrated?
    - Where is it more prevalent?
    - Prediction of hate-related crimes
- Quantification/summarization human rights (protection-related incidents) reported either by external parties or refugees themselves around the following topics?
    - **SGBV**: (sexual/gender-based violence, including LGBTIQ+ persons)
    - **SH:** sexual harassment and **PSEA**: sexual exploitation and abuse
    - **CP**: child protection. From CP, special category for UASC = Unaccompanied and Separated Children.

    - **Physical protection:** menace to be killed or persecuted, assistance needed to hide or move out of a certain place (e.g. camp, house, settlement, city) due to persecution,
    - **Detention:** cases where refugees have been deprived from physical freedom (e.g. jail)
    - **Legal protection**: Urgent need for legal assistance to support on different issues (detention, refoulement (violation of the no return principle), asylum access, registration, status determination). Extended definition here: Website
    - **Disabilities:** any person with a mental or physical disability who is forced to flee.

Datasets and resources: triple nexus

Humanitarian datasets:

- HumSet
- Situational reports (SitReps) UNHCR (left menu at the bottom "Situation Reports/Updates")
- OCHA Reliefweb reports
- UNHCR needs assessments
    - Other agencies needs assessments
- The DEEP and documentation

Development datasets:

- FEWS reports/message updates
- FSNAU reports/updates
- World Bank reports: Somalia

Peacebuilding sector datasets: violent conflict

- Conflict in Somalia: historical highlights
    - ACLED API user manual, focus on the text-based columns (e.g. incident report, actors perpetuating conflict)
    - ACLED API code, focus on the text-based columns (e.g. incident report, actors perpetuating conflict)
- Uppsala University: VIEWS project and API access (focus on the text-based columns)

Other datasets and/or research

- Twitter API
- World News API

Ethical and Human rights considerations: We recommend reading out the following reading for researchers to take into account when using GAI:

- Generative AI for Humanitarians
- Misinformation and Disinformation and Deep Fakes
- Digital protection and combating hate speech against refugees
- Social Media and Forced Displacement: Big Data Analytics & Machine-Learning
- Improving Humanitarian Needs Assessments through Natural Language Processing
- Using Social Media in Community Based Protection: A Guide
- Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP
- Using large language models to help train machine learning SDG classifiers
- 'Refugee voices', new social media and politics of representation: Young Congolese in the diaspora and beyond
- ChatGPT and Large Language Models: Challenges and Opportunities (webinar)