Capstone Project Concept Note and Implementation Plan

Project Title: Offline AI-Powered Community Support and Vulnerability Prediction System for Conflict-Affected Areas in Myanmar

Team Members

1. Thant Zin Moe
2. Zin Mi Mi Ko
3. Aye Yu Yu San
4. Yoon Poe Phyu
5. Min Khant Zaw

Concept Note

1. Project Overview

This project aims to empower communities in war-affected, rural, and politically oppressed areas of Myanmar by combining predictive modeling with an offline-accessible chatbot system. The predictive model identifies communities most vulnerable to poverty, food insecurity, health crises, and natural disasters. The offline chatbot provides residents with guidance on accessing NGO and UN support services related to health, education, livelihood, and disaster response, even in low-connectivity environments since Internet availability is very limited in those areas. Together, these components enable data-driven interventions and direct community support.

2. Objectives

The primary aim of this research is to design, build and evaluate a novel, hybrid AI system which can predict community vulnerabilities and deliver targeted, offline support and educational resources in the conflict-affected regions of Myanmar. To achieve this aim, the project has the following objectives:

1) **To Develop a Robust Vulnerability Prediction Model** – Hybrid machine learning (Random Forest) and deep learning (1D-CNN) satisfy this objective by training on a multi-source dataset including conflict (ACLED), socio-economic (UNDP), and satellite (NASA) data for accurately identifying the specific townships or regions at the highest risk of poverty and educational disruption.
2) **To Design and Implement a Lightweight, Offline Chatbot** – In this objective, creation of an efficient, keyword-based Natural Language Processing (NLP) chatbot which can run on low-resource devices without internet connection. This chatbot will also be populated with custom-curated knowledge based on UN and NGO sources to provide users with clear, actionable guidance on health, education, livelihood and emergency support.

3) **To Create a Cohesive Framework that Links Prediction to Action** – A novel approach is approached in this objective by creating a system where the output of the vulnerability prediction model directly informs the targeted deployment of the offline chatbot, closing the gap between simply identifying at-risk communities and proactively providing them with the tools they need.
4) **To Evaluate the System's Effectiveness and Feasibility** – This is the final objective in which measurement of the project's success is done by accessing both predictive accuracy of the vulnerability model and the usability and relevance of the offline chatbot's information, ensuring that the solution is both technically sound and practically useful for the target community.

3. Background

The problem of this project addresses is the crisis faced by communities in Myanmar's conflict-affected, rural, and politically oppressed regions. These populations experience severe primary vulnerabilities such as poverty, food insecurity, and health crisis which are the direct cause of secondary order emergencies, like the widespread disruption of schooling and high child dropout rates, which is critically exacerbated by a "digital divide", where limited or non-existent internet connectivity isolates these communities, cutting them off from the essential aid services and digital learning resources that could offer a lifetime.

After conducting some literature reviews, studies like Hall et al., (2023) and Alturif et al., (2024) have established a strong precedent for using satellite and socio-economic data for the identification of at-risk populations. Research like Sahu et al., (2023) and Bholesale et al., (2025) have focused on creating offline chatbots to deliver information. The gap here is that those two fields are not connected in such a way that prediction is not being used to guide the proactive deployment of support.

A machine learning approach is therefore beneficial and necessary for bridging this gap because the drivers of vulnerability such as patterns of conflict, economic decline, and climate events are too complex and dynamic to be tracked manually. And thus, a machine learning model is necessary to synthesize diverse, multi-source datasets (e.g., ACLED, UNDP, and NASA) and identify non-obvious patterns and critically impacted regions. This data-driven approach is the only feasible way for moving beyond reactive aid and enable the proactive, targeted deployment of the offline educational and support tools in these research proposals.

4. Methodology

The methodology for this project is a novel hybrid design integrating two distinct machine learning components as follows:

Component 1: Vulnerability Prediction Model

For vulnerability prediction, a hybrid machine learning and deep learning methodology will be used, being crucial to effectively synthesize diverse and multi-source datasets of the project. Specific models and algorithms include:

- **1D-Convolutional Neural Network (1D-CNN)** – This is the deep learning model and this framework will be used to analyze our time-series data such as historical conflict events from ACLED, economic trends from UNDP. This is chosen because of its proven ability in efficient extraction of temporal correlations and accuracy in poverty-related metrics forecasts.

- **Random Forest** – This algorithm is a machine learning one and this will be used for the analysis of "hard" indicators from our structured data like satellite data including nighttime lights and static socio-economic indicators. For us to accurately identify which factors are the most significant predictors, Random Forest is critical because of its high accuracy, efficiency and interpretable nature.

Component 2: Offline Chatbot System

The methodology that will be used for this support tool is Lightweight Natural Language Processing (NLP), which is specifically designed for a "low resource" project.

Keyword Recognition-Predicated Model – We will implement a chatbot based on this model, but not by using a high-resource Large Language (LLM) like LLaMA3, since such models are too large and computationally resource intensive which is not suitable for our target environment. Instead, our methodology is simple, fast and efficient, parsing a user's query for specific keywords such as "food", "school" or "clinic", and then match these keywords to the pre-defined, custom-curated knowledge base of NGO or UN information. This approach makes sure that the chatbot is "lightweight".

5. Architecture Design Diagram

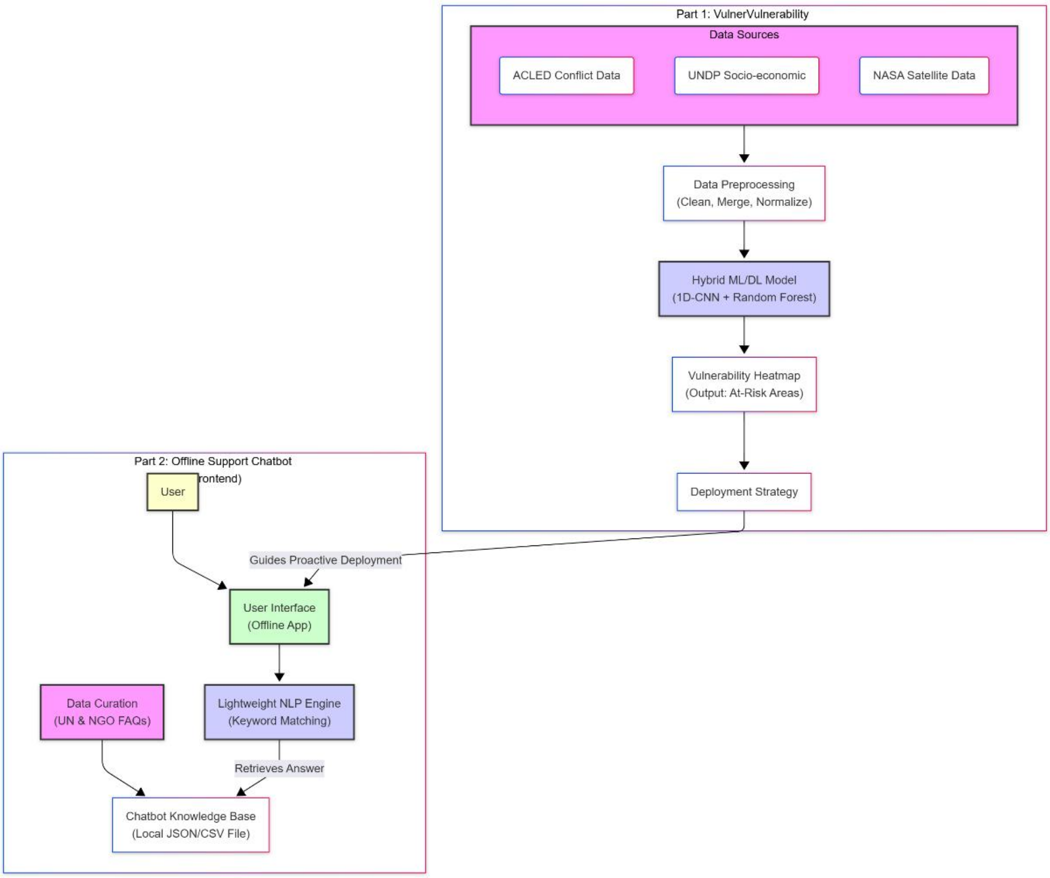


Figure 1: High Level System Architecture

The above diagram shows the hybrid nature of the project including two parts:

- **Part 1 (Backend)** – This is the “Vulnerability Prediction Model” which shows the internal processes such as taking various Data Sources from ACLED, UNDP or NASA, processing them through the Hybrid ML or DL Model and producing a Vulnerability Heatmap, creating a Deployment Strategy by the output.
- **Part 2 (Frontend)** – This is the “Offline Support Chatbot” which is the user-facing application, in which a user interacts with the User Interface, which used a Lightweight NLP Engine to pull answers from the local Chatbot Knowledge Base.

6. Data Sources

This project will leverage a multi-source data strategy. The vulnerability prediction model will be trained using open-source CSV/GeoJSON datasets, including ACLED for conflict data, World Bank & UNDP for socio-economic indicators, and NASA for climate/satellite data. This data is highly relevant as it provides the quantitative inputs including conflict rates, income index, necessary for the machine learning model to identify communities vulnerable to poverty, health crises, and natural disasters. The offline chatbot will use a custom dataset curated from NGO and UN FAQs such as health, education, livelihood guidance. This data is relevant as it forms the core knowledge base for providing direct community support in low-connectivity environments. Required preprocessing steps for the prediction data include cleaning, normalization, and feature engineering including creating a conflict rate or climate risk index to prepare the data for the machine learning algorithms.

7. Literature Review

The chosen methodology is strongly supported by existing research. For the prediction component, literature confirms that a hybrid Machine Learning (ML) and Deep Learning (DL) approach is state-of-the-art for vulnerability analysis, with studies like Hall et al. (2023) validating the use of satellite imagery and Alturif et al. (2024) demonstrating the power of 1D-CNNs for time-series socio-economic data. A key finding from Alturif et al. (2024) that underpins this project is the strong 0.84 correlation between SDG 1 (No Poverty) and SDG 4 (Quality Education), which validates our model's focus. For the support component, research by Bhosale et al. (2025) and Sahu et al. (2023) confirms the feasibility of offline chatbots, with Sahu et al. (2023) specifically highlighting the use of lightweight, keyword-based models for educational use cases in low-resource environments. This project builds upon and extends this work by addressing a critical gap: existing research treats prediction and aid delivery as separate fields. Our project will be the first to create a single, cohesive framework that uses the output of the vulnerability model to proactively guide the deployment of the offline educational and support chatbot, moving from passive analysis to active, data-driven intervention.

Implementation Plan

1. Technology Stack

Programming Language: Python 3.14.0 will be the primary language for all components, including data preprocessing, model training (ML/DL), and the chatbot’s simple NLP logic.

Data Processing and Analysis Libraries:

- **Pandas** - This library is for loading, cleaning and managing the structured CSV data such as ACLED, Worlds Bank.
- **GeoPandas** – This is for handling and merging the GeoJSON data and preparing the geospatial visualizations.
- **Numpy** – For high-performance numerical computations required by the ML/DL models.

Machine Learning and Deep Learning Frameworks:

- **Scikit-learn**: This will be used to implement the Random Forest model, as well as for data preprocessing tasks like normalization and train-test splits.
- **Keras** (with TensorFlow backend): This framework is crucial for building, training, and evaluating the 1D-Convolutional Neural Network (1D-CNN), which is ideal for our time-series data.
- **NLTK** (Natural Language Toolkit): This will be used for the chatbot's lightweight NLP, specifically for "keyword recognition" (e.g., tokenization, stemming) to match user queries to the local knowledge base.

Visualization Libraries:

- **Matplotlib and Seaborn**: To generate the correlation matrix and other statistical plots.
- **Folium**: To create the interactive "Vulnerability Heatmap" for the model's final output.

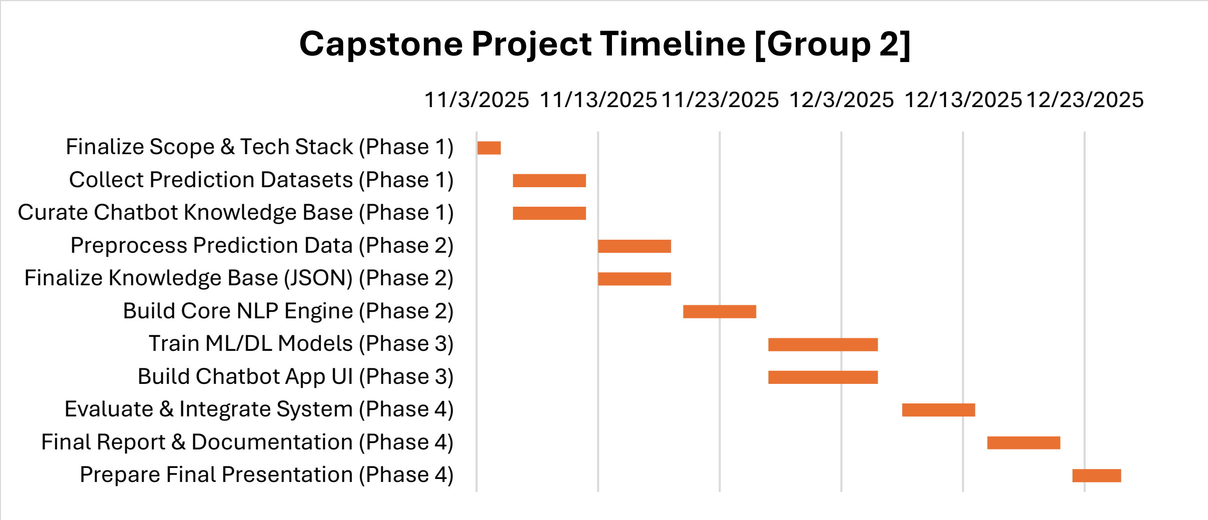
Application Framework (Chatbot):

- **Flutter (or a similar lightweight framework like Kivy/React Native)**: To build the cross-platform (iOS/Android) offline mobile application. This framework is essential for creating the user interface and ensuring the app is "lightweight" and can run efficiently on low-resource mobile devices.

Other Software and Hardware:

- **Jupyter Notebooks / VS Code**: For model development, prototyping, and analysis.
- **Git / GitHub**: For version control.
- **Hardware**: The models will be trained on a standard development machine. The target deployment of hardware is low-resource mobile devices (e.g., basic smartphones) in the target communities.

2. Timeline



3. Milestones

A list of the key project milestones include:

- **Milestone 1**: Data Acquisition and Processing: All data sources (ACLED, UNDP, NASA, and UN/NGO FAQs) are successfully collected, cleaned, normalized, and formatted for use in the model and chatbot.
- **Milestone 2**: Prediction Model Trained: The hybrid ML/DL vulnerability model (1D-CNN + Random Forest) is successfully trained, evaluated, and achieves a target accuracy metric.
- **Milestone 3**: Chatbot Functionality (Key Feature): The lightweight, keyword-based NLP engine is fully implemented and successfully retrieves accurate answers from the local JSON knowledge base.
- **Milestone 4**: System Integration and Deployment Package: The final vulnerability heatmap is generated, and the offline chatbot mobile app is fully integrated and packaged for deployment.
- **Milestone 5**: Project Completion: The final capstone report and all project documentation are submitted.

4. Challenges and Mitigations

Challenge	Mitigation Strategy
Data Quality	**Challenge:** Data from diverse sources (ACLED, UNDP, NASA) will have missing values, different formats, and mismatched geographic levels. **Mitigation:** We will implement rigorous data preprocessing, including aggregation to a uniform township level, imputation of missing data, and normalization of all features.
Model Performance	**Challenge:** The vulnerability prediction model may have low accuracy or overfit the training data. **Mitigation:** We will use a hybrid model (Random Forest + 1D-CNN) as supported by the literature. We will perform extensive feature engineering,

	cross-validation, and hyperparameter tuning to ensure the model generalizes well.
Technical Constraints	**Challenge:** The offline chatbot must run on "low-resource" mobile devices in "low-connectivity environments". **Mitigation:** We have already mitigated this by rejecting a high-resource LLM (like LLaMA 3). Our chosen architecture uses a lightweight, keyword-based NLP model and a small JSON knowledge base, ensuring the app is fast and small.

5. Ethical Considerations

Consideration	Ethical Challenge	Mitigation Strategy
Data Privacy	**Prediction Model:** The aggregated data on vulnerability (conflict, poverty), if leaked, could be misused. **Chatbot:** User queries (e.g., "Where is food aid?") are highly sensitive and reveal personal needs.	The **prediction model** will use aggregated, anonymized data at the township level, not data on individuals. The **Chatbot** is designed to be 100% offline**.** User queries are processed locally and never transmitted, ensuring that "all information remains within the user's local environment," which provides maximum data privacy and security.
Bias	**Model Bias:** The training data (e.g., ACLED, World Bank) may under-report data from hard-to-reach or conflict-affected regions. This could cause the model to be biased against the most vulnerable populations, leading to misallocation of the chatbot. **Data Bias:** The curated NGO/UN knowledge base may have implicit biases (e.g., focusing on services in specific areas, not having info in the correct local languages).	We will audit the model for bias by cross validating its predictions with qualitative reports from the ground. The model's limitations will be clearly stated. The chatbot's knowledge base will be curated from multiple diverse sources (not just one NGO) and will be reviewed for neutrality. We will prioritize translation into the primary local languages.
Potential Impact	**Risk of Harm:** The primary ethical risk is providing inaccurate or outdated information in a dynamic conflict zone. Directing a user to a clinic or food distribution point that has been moved or destroyed could cause real-world harm.	Information in the chatbot's knowledge base will be clearly dated and include **disclaimers** (e.g., "Verify this information locally before traveling"). The app will be designed to be discreet, will not collect any personal user data (e.g., name,

	Security Risk: The app itself could put a user at risk if discovered by authorities.	location), and will be a standalone tool, mitigating user security risks.

6. References

Alturif, G., Saleh, W., El-Bary, A. A., & Osman, R. A. (2024). Using artificial intelligence tools to predict and alleviate poverty. *Entrepreneurship and Sustainability Issues*, *12*(2), 400–413. https://ideas.repec.org/a/ssi/jouesi/v12y2024i2p400-413.html

Bhosale, V. N., Ghorpade, A. S., & Khamkar, K. M. (2025). Echo AI Platform: Revolutionizing Offline Chatbots with LLAMA 3 for Data Retrieval. *Journal Publication of International Research for Engineering and Management (JOIREM)*, *05*(04). https://joirem.com/wp-content/uploads/journal/published_paper/volume-05/issue-4/J_bVznkGgb.pdf

Hall, O., Dompae, F., Wahab, I., & Dzanku, F. M. (2023). A review of machine learning and satellite imagery for poverty prediction: Implications for development research and applications. *Journal of International Development*, *35*(7), 1753–1768. https://doi.org/10.1002/jid.3751

Sahu, A. K., Pandey, S. K., Agarwal, M., & Chauhand, S. (2023). *Offline Virtual Chat Bot by Using Natural Language Processing*. SSRN. https://ssrn.com/abstract=4386503