Course: POWER LEARN PROJECT.

Student: Luke Mbogo

Personal Project

Saturday, July 19, 2025.

Week 8 Final Project:

Gender Based Violence Monitoring AI system.

Contents

Project Title

GBVTrackAI: An AI-Powered Dashboard for Monitoring Gender-Based Violence (SDG 5)

SDG Focus

Goal:

- **SDG 5: Gender Equality**

Problem:

Gender-Based Violence (GBV) remains a widespread and underreported issue in many regions, including Kenya. NGOs and local authorities lack real-time, data-driven tools to identify emerging GBV hotspots, analyze trends, and intervene effectively. Manual reporting systems often miss out on early warning signs found in media, public forums, and social data.

AI Approach

Software Engineering Skills Applied:

- **Automation:**
  Automated text extraction and classification using NLP to identify GBV-related information from structured datasets and unstructured online sources (e.g., news).
- **Testing:**
  Unit testing of the NLP classification pipeline, integration tests for dashboard functionality, and model validation with precision, recall, and F1-score metrics.
- **Scalability:**
  Modular architecture using Python classes and functions for data ingestion, model training, prediction, and visualization. The codebase is container-ready and deployable via Streamlit.

Technical Solution:

We use NLP techniques to detect GBV-related text from public datasets and scraped news articles. A classification model identifies GBV-related posts, and analytical tools detect patterns (e.g., time, location, and frequency of incidents). Results are visualized in a user-friendly dashboard with charts and reports.

Tools & Frameworks

AI/ML:

- scikit-learn for model training (Logistic Regression, SVM)
- spaCy for Named Entity Recognition
- TextBlob or VADER for sentiment analysis
- NLTK for tokenization and preprocessing

Software Engineering:

- Git for version control
- pytest for unit and integration testing
- Streamlit for dashboard deployment
- Docker (optional) for modular container deployment

Data Sources:

- Kaggle: Gender-Based Violence Dataset
- Kaggle: Women Safety Reports (India)
- News scraping using newspaper3k Python library
- Optional Twitter scraping using Tweepy + Twitter API (filtered by GBV keywords)

Deliverables

- **Code:**
  Clean Python scripts with comments and reusable functions for:
  - Data cleaning
  - NLP-based classification
  - Sentiment analysis
  - Visualization pipelines
- **Deployment:**
  Interactive dashboard using Streamlit with the following features:
  - Line charts showing GBV trends over time

    - Sentiment breakdown of reports
    - Text classification outputs
    - Word cloud for key terms
- **Report:**
  A comprehensive technical report covering:
    - Problem context and SDG relevance
    - AI model architecture and metrics
    - Ethical review and impact assessment
    - Future work and deployment possibilities

Ethical & Sustainability Checks

- **Bias Mitigation:**
  Datasets will be reviewed for regional and linguistic diversity to avoid geographical or cultural bias. Model performance will be analyzed across multiple countries (where data allows).
- **Environmental Impact:**
  Lightweight ML models are used (e.g., Logistic Regression), avoiding high-energy deep learning unless necessary. Execution is done on local machines or Google Colab.
- **Scalability:**
  Code is optimized to work on low-resource devices. The dashboard can run locally without internet after setup and can be translated into a mobile-first interface in future updates.

Sample Project Outline

Phase	Tasks
Ideation	Research GBV trends in Kenya and globally; select datasets; finalize technical direction.
Development	Build data preprocessing scripts; train and evaluate GBV classifier; set up dashboard.

Phase	Tasks
Testing	Perform accuracy tests; run unit tests; analyze classifier bias and false positives.
Deployment	Deploy Streamlit dashboard locally and document usage. Optional: deploy on Hugging Face.
Monitoring	Gather test-user feedback; assess model drift and add data augmentation as needed.

How AI for Software Engineering Concepts Apply to SDGs

Key Concept	Application to SDG 5 (Gender Equality)
Automated Testing	Ensures the GBV classifier is robust and reliable for sensitive use cases like social policy monitoring.
CI/CD Pipelines	Future goal: automate scraping, classification, and dashboard updates using GitHub Actions or Jenkins.
Version Control	Git is used to manage collaborative changes in model scripts and dashboard UI.
Ethical AI Design	Minimizes false positives and ensures sensitive information is processed ethically and securely.
Modular Code	Enables rapid adaptation for related SDGs (e.g., SDG 16 – Peace and Justice) or other locations/regions.

Reflection Questions

- **SDG Alignment:**
  This solution directly contributes to SDG 5 by improving GBV data collection, monitoring, and response — especially in under-reported areas.

- **Ethical Risks:**
  Risk of false positives in GBV detection from text, potential data sensitivity. Mitigated through model evaluation, user disclaimers, and ethical review of datasets.
- **Sustainability via Software Engineering:**
  Modular code ensures future upgrades. Testing ensures consistent performance. Git version control enables collaboration for scaling.