AITIEWS Project – Development Progress Report

Reporting Period: Weeks 0–2
Date: 12th February 2026

1. Project Overview

The AITIEWS project aims to develop an AI-driven intelligence and early warning system capable of collecting, analyzing, and predicting security threats from structured and unstructured data sources. This report summarizes progress made during the **Preparation Phase (Phase 0)** and **initial implementation of Phase 1**.

2. Phase 0: Preparation – Status COMPLETED

Objectives Achieved

- Project requirements and high-level technical architecture finalized.
- Development and deployment environments successfully established.
- Core DevOps and cloud foundations put in place.

Key Activities Completed

- Defined project scope, objectives, KPIs, and success metrics.
- Established Git repositories with version control best practices.
- Set up CI/CD pipelines for automated testing and deployment.
- Configured cloud infrastructure and development environments.
- Designed initial system architecture and data flow diagrams.

Deliverables Completed

- Technical Specification Document
- Cloud & DevOps Environment Setup
- System Architecture Blueprint

3. Phase 1: Data Collection & Preprocessing – Status IN PROGRESS

Objectives

- Build scalable data ingestion pipelines.
- Prepare structured and unstructured data for downstream AI/ML modeling.

Progress to Date

- Initial data ingestion pipelines have been implemented.
- Web crawling mechanisms for public online sources are operational.
- API connectors for structured data ingestion have been configured.
- Text preprocessing workflows initiated, including:
    - Language detection
    - Tokenization
    - Stop word removal
    - Multilingual handling (English, Kiswahili, Somali, Sheng)

Data Storage & Indexing

- PostgreSQL database configured for structured data storage.
- Elastic search set up for indexing and fast search of unstructured text data.

Current Deliverables

- Initial data pipeline operational
- Partially cleaned and structured datasets
- Prototype dashboards displaying ingested raw data

4. Challenges & Mitigation

- **Multilingual data complexity:** Addressed by integrating language detection and multilingual NLP preprocessing.
- **Data heterogeneity:** Standardized schemas and preprocessing rules introduced to improve consistency.

5. Next Steps (Remaining Phase 1 Tasks)

- Complete full preprocessing and normalization of datasets.
- Expand data ingestion coverage and improve crawler robustness.
- Finalize dashboards for data visibility and monitoring.
- Prepare datasets for NLP and text analytics model development (Phase 2).

6. Overall Status Summary

- **Phase 0:** Completed
- **Phase 1:** Ongoing, on schedule
- **Git Activity:** Active development with regular commits reflecting pipeline setup, preprocessing logic, and infrastructure updates.

Conclusion:
The project is progressing as planned, with foundational infrastructure completed and core data pipelines actively under development. The system is well-positioned to transition into **NLP & Text Analytics Model Development (Phase 2)** upon completion of data preprocessing.