GROUP 14 PROJECT PROPOSAL

Localized Swahili LLM for AI-Assisted Terrorism Threat Reporting in Underserved Kenyan Communities.

GROUP MEMBERS:
SOLOMON NJOGO
LEWIS MWANGI

PART 1: INNOVATIVE AI APPLICATION ADDRESSING THREATS

1.1 The Security Threat: Terrorism & Communication Breakdown

Primary Threat Context:

Kenya faces a persistent terrorism threat, particularly from Al-Shabaab in northern and coastal regions. However, the greater danger lies in the **communication gap** between communities who witness suspicious activities and security agencies who need timely intelligence.

Quantified Problem Impact:

Challenge	Data Point	Source
Access barrier	71% of rural Kenyans lack formal security reporting channels	Kenya Police Annual Report 2023
Language barrier	69% of Kenyans speak Swahili as primary language; existing systems are English-only	Kenya National Bureau of Statistics 2024
Fear factor	80%+ underreporting rate in border areas due to retaliation fears	UNDP Kenya Conflict Analysis 2024
Economic impact	KES 12B+ annual tourism losses in insecure regions	Kenya Tourism Board 2024
Human cost	150+ Al-Shabaab incidents in 2023; delayed reporting contributes to preventable casualties	Counter-Terrorism Centre Reports
Response delay	48-72 hours average from incident observation to security response	Security sector analysis

Multi-Domain Threat Intersections:

- **Security:** Direct terrorism casualties and property destruction
- **Economic:** Stifled growth in border regions; disrupted trade corridors
- **Social:** 300,000+ internally displaced persons; community disengagement
- **Governance:** Eroded trust between security forces and marginalized communities; state fragility

1.2 The Innovation: Swahili-Localized LLM with Agentic Intelligence

Core Innovation Statement:

We develop Kenya's **first Swahili-localized security LLM** that combines generative AI (natural language understanding/generation) with agentic AI (autonomous threat validation, categorization, and escalation) to create an accessible, anonymous, and intelligent threat reporting system via WhatsApp.

Technical Approach:

Component 1: Generative AI (Fine-tuned LLM)

- **Base Model:** LLaMA 2-7B
- **Fine-tuning Strategy:**
    - Domain adaptation on 50,000+ Swahili security texts (police reports, news, community narratives, synthetic data)
    - LoRA (Low-Rank Adaptation) for parameter-efficient training (70% cost reduction)
    - Custom tokenizer extension for Sheng, coastal Swahili, and 6 regional dialects
- **Unique Capabilities:**
    - Multi-dialect conversation (understands "mtaa huu iko na shida" and "ebu angalia hizi vitu")
    - Contextual geography understanding (recognizes "Boni Forest," "Ijara subcounty," "Garissa-Mandera road")
    - Bilingual intelligence: Swahili input → English threat summaries for security analysts

Component 2: Agentic AI (Multi-Agent System)

Three autonomous agents create intelligent threat processing pipeline:

1. **Validator Agent**
    - *Function:* Ensures report completeness (Who? What? Where? When?)
    - *Technique:* Named Entity Recognition + semantic similarity scoring
    - *Agentic Behavior:* Autonomously prompts users for missing critical details before escalation
2. **Escalation Agent**

- *Function:* Classifies threat severity (Low/Medium/High/Critical) and assigns priority
- *Technique:* Fine-tuned BERT classifier (4-class) + urgency keyword detection
- *Agentic Behavior:* Makes autonomous escalation decisions:
  - **Critical:** Immediate attack indicators → SMS alert to security dashboard (<2 min)
  - **High:** Weapons/explosives → 30-minute escalation window
  - **Medium/Low:** Suspicious activity → daily/weekly intelligence digest

3. **Learning Agent**
   - *Function:* Improves system accuracy from security analyst feedback
   - *Technique:* Reinforcement Learning from Human Feedback (RLHF)
   - *Agentic Behavior:* Continuously adapts classification thresholds based on correction patterns (target: +5-10% accuracy per quarter)

1.3 How It Solves the Threat

Problem-Solution Mapping:

Threat Dimension	Current State	Our Solution	Expected Impact
Access	71% lack reporting channels	WhatsApp interface (24M Kenyan users)	10x reach to underserved areas
Language	English-only systems exclude majority	Multi-dialect Swahili understanding	+60% accessibility
Fear	No anonymity guarantees	Zero-knowledge architecture (no PII stored)	Eliminates retaliation barrier
Speed	48-72 hour response delays	<5 min critical alert delivery	90% faster threat interdiction
Intelligence quality	Generic, incomplete reports	Agentic validation ensures actionable intel	40% increase in prosecutable leads

Threat Dimension	Current State	Our Solution	Expected Impact
Trust	Communities distrust authorities	Transparent AI + community education	Restored police-citizen cooperation

Real-World Usage Scenario:

1. **Citizen observes:** Two people counting motorcycles near school at midnight with heavy bags (Ijara, 2:00 AM)
2. **WhatsApp report:** "Nimeona watu wawili wakihesabu pikipiki karibu na shule ya Ijara usiku wa manane. Wanabeba mizigo kubwa." (In Swahili)
3. **Validator Agent:** Prompts for additional details (age estimate, weapons presence)
4. **LLM generates:** Natural follow-up in Swahili + educates user on threat indicators
5. **Escalation Agent:** Classifies as HIGH (heavy bags + late hour + school target), generates English summary
6. **Alert delivered:** SMS to security dashboard within 2 minutes with GPS coordinates
7. **Learning Agent:** Tracks outcome (false alarm vs. interdicted attack) to improve future classifications

1.4 Differentiation from Existing Solutions

Competitive Landscape Analysis:

Feature	Traditional Hotlines (999)	Usalama App	Ushahidi Platform	Our Solution
Language	English/basic Swahili	English	English	Multi-dialect Swahili + Sheng
Interface	Phone call (intimidating)	Custom app download	Web form	WhatsApp (familiar)
Availability	Business hours	24/7	24/7	24/7
Anonymity	Phone number logged	Registration required	Optional	True zero-knowledge
Intelligence	Manual operator triage	Keyword flagging	Crowdsourced mapping	AI-powered validation + escalation

Feature	Traditional Hotlines (999)	Usalama App	Ushahidi Platform	Our Solution
Context awareness	Generic questions	None	None	Understands "Boni Forest near Bargoni"
Education	One-way reporting	None	None	Teaches threat identification
Adoption barrier	High (unfamiliar, formal)	Medium (app install)	High (literacy required)	Low (messaging app they use daily)

Key Differentiator: We don't create a new system citizens must learn—we meet them where they are (WhatsApp, Swahili) and make reporting conversational.

PART 2: FEASIBILITY FOR KENYA/EAST AFRICA

Social Feasibility Assessment

Cultural & Social Fit:

Language Reality:

- 69% of Kenyans speak Swahili as first or second language
- English proficiency: 27% (mostly urban/educated populations)
- **Insight:** Security threats occur disproportionately in low-English-proficiency areas (border counties)

Technology Adoption Patterns:

- WhatsApp penetration: 89% of smartphone users (vs. 12% for custom government apps)
- Trust factor: WhatsApp perceived as private/informal (vs. official hotlines perceived as government surveillance)
- Behavioural precedent: Kenyans already use WhatsApp groups for community safety alerts ("Nyumba Kumi" neighbourhood watch)

Anonymity & Cultural Context:

- Clan/tribal dynamics: Reporting can be seen as betrayal if suspect is community member
- Fear of retaliation: Al-Shabaab history of targeting "informants"

- **Solution:** Zero-knowledge design addresses honor/shame dynamics that traditional hotlines ignore

User Acceptance Evidence:

- Preliminary focus group (n=30, Mandera County, conducted Oct 2024): 87% expressed willingness to use
- Key quote: "Ningependa kusema kitu lakini ninaogopa. Kama haiwezi kujua ni mimi, ninaweza sema." ("I'd like to say something but I'm scared. If it can't know it's me, I can speak.")

Community Education Strategy:

- Partner with Community Policing Committees (existing trust structures)
- Chatbot teaches threat identification through conversation (not just reporting)
- Posters/radio ads in Swahili: "Ripoti Usalama Kwa Simu Yako" (Report security via your phone)

PART 3: ALIGNMENT WITH SUB-THEMES/OBJECTIVES

3.1 Sub-theme v: Generative and Agentic AI

Requirement 1: Generative AI Application

How we demonstrate mastery:

- **Fine-tuned LLM:** LLaMA 2-7B adapted for Swahili security domain (demonstrates transfer learning, domain adaptation)
- **Natural language generation:** Chatbot generates contextual responses in multiple Swahili dialects (not template-based)
- **Bilingual translation:** Automatically generates English threat summaries for security analysts (cross-lingual NLP)
- **Educational content generation:** Creates personalized threat awareness tips based on user's report context

Technical depth evidence:

- LoRA fine-tuning (demonstrates understanding of parameter-efficient methods)
- Custom tokenizer extension (shows NLP preprocessing skills)
- Prompt engineering for multi-turn conversations

Requirement 2: Agentic AI Application

How we demonstrate mastery:

- **Multi-agent system:** Three autonomous agents (Validator, Escalation, Learning) with distinct goals
- **Autonomous decision-making:** Escalation Agent makes threat prioritization decisions without human input (within defined parameters)
- **Collaborative intelligence:** Agents work in pipeline (Validator → LLM → Escalation → Learning)
- **Self-improvement:** Learning Agent uses RLHF to autonomously adapt from feedback

Agentic behaviors demonstrated:

1. **Goal-directed:** Each agent has clear objective (completeness, prioritization, improvement)
2. **Autonomous:** Makes decisions within scope (e.g., "escalate as HIGH" without asking)
3. **Adaptive:** Learning Agent modifies future behavior based on outcomes
4. **Reactive:** Responds to real-time inputs (incoming reports trigger agent pipeline)

Comparison to basic chatbot (what we're NOT doing):

- Simple rule-based responses (we use fine-tuned LLM)
- Static behavior (we have learning/adaptation)
- No decision-making (our agents autonomously prioritize threats)