

ONDIRA NABOTH

Email: ondiranaboth@gmail.com

PROFESSIONAL SUMMARY

AI Data Evaluator and AI Red Teamer with hands-on experience annotating, reviewing, and stress-testing multilingual (Swahili & English) datasets and large language models. Strong focus on safety evaluation, adversarial prompt design, edge-case discovery, policy compliance, and model robustness. Technical background in information systems, AI fundamentals, and cybersecurity enabling structured, specification-driven evaluation work.

CORE EXPERTISE

- Text, audio, and multimodal data annotation
- LLM evaluation and output quality assessment
- Adversarial prompt engineering and jailbreak testing
- AI safety and policy compliance testing
- Intent classification and sentiment analysis
- Named entity recognition and semantic consistency checks
- Transcription and speech-quality evaluation (Swahili)
- Localization and cultural relevance validation
- Edge-case identification and failure mode analysis
- Structured evaluation workflows and QA documentation

PROFESSIONAL EXPERIENCE

AI Red Teamer (Contract) – Independent / Remote (Sept 2025 – Present)

- Designed adversarial prompts to test model robustness and safety boundaries.
- Identified jailbreak patterns, hallucinations, and unsafe output behaviors.
- Conducted policy compliance testing across safety categories.
- Documented failure cases and provided structured mitigation feedback.
- Evaluated model alignment against specification and safety standards.

AI Language Data Evaluator (Contract) – Independent / Remote (Jan 2024 – Present)

- Evaluated Swahili and English datasets for semantic accuracy, intent alignment, and tone preservation.
- Reviewed text and audio data for entity consistency, grammatical correctness, and contextual clarity.
- Assessed sentiment and intent interpretation accuracy.
- Conducted QA checks to identify inconsistencies, edge cases, and guideline deviations.
- Validated localized content for AI training readiness.

LLM Operations & Evaluation Specialist (Contract) – Zone1Kenya (Oct 2024 – Oct 2025)

- Supported deployment and monitoring of large language models.
- Assisted in evaluation pipelines for model output quality and reliability.
- Maintained structured workflows for dataset versioning and evaluation feedback.

Cybersecurity Intern – National Cereals Board (Jan 2024 – Aug 2024)

- Monitored system logs and alerts to identify anomalies.
- Assisted in vulnerability assessments and compliance checks.
- Applied procedural accuracy transferable to data quality and AI evaluation work.

EDUCATION

Bachelor of Commerce – Information Systems

Egerton University

March 2020 – November 2025

CERTIFICATION

AI Associate Certification – Oracle University (Aug 2025 – Nov 2025)

LANGUAGES

English – Fluent

Swahili – Fluent