

Kemunto Ochwang'i

evatom86@gmail.com

<https://github.com/evalevanto>

<https://www.linkedin.com/in/kemunto-o-9a972511a/>

EDUCATION

Strathmore University

June 2015 - Nov 2018

Bachelor of Informatics and Computer Science

WORK EXPERIENCE

Microsoft Research Africa

Aug 2026 - Present

Research Engineer - Contract

PRISM AI Safety Research Fellowship

July 2026 - Present

Research Fellow

Mentored by Preeti Ravindra.

- Doing research on synchronous information flow graph monitors.

LASR LABS

Jan 2026 - Sep 2026

Research Fellow

Supervised by Rob Kirk and Alex Souly.

- Researching and building a benchmark to evaluate models' ability to discriminate when they are in evaluation and deployment settings.

Paper: [EvalDetectBench: A Benchmark for Measuring Evaluation Awareness in Frontier Language Models](#)
accepted at NeurIPS 2026 E&D track.

Lyptus Research

Feb 2026 - Mar 2026

Cybersecurity contractor

- I am contributing to the human study to measure how AI offensive cybersecurity capability is increasing over time.
-

AI Safety Grantee

June 2025 - Jan 2026

- Awarded funding by Open Philanthropy to do independent study and exploratory research in AI Security.
 - During this time I studied and passed my Offensive Security certification and attended the ARENA program at LISA.
-

ALIGNMENT RESEARCH ENGINEER ACCELERATOR

Sept 2025 - Oct 2025

ARENA 6.0 Scholar

- Participated in an in-person research engineering bootcamp focused on upskilling towards contributing to technical AI Safety.
 - For the capstone project, [my team](#) reproduced the [`Persona Vectors` paper by Anthropic](#) and studied coherence scorer and degradation on positive steering.
-

AI SAFETY ENGINEERING TASKFORCE

Jan 2025 - Mar 2025

Model Evaluations Engineer

- I built a cyber capability evaluation for the UK AI Safety Institute's suite of pre-deployment tests for frontier models. The evaluation measures accuracy improvement of the model under evaluation, afforded with memorized reflection, to successfully solving the Intercode CTF challenges in consequent clean attempts.
-

MICROSOFT

Sep 2020 - June 2025

*Software Engineer II | Data Scientist [61]
Chat Synthesis Team; People Discovery Team*

- I am currently working on evaluating manufactured data for fine tuning through feature distribution matching with real-user data. This preliminary work will allow us to measure the quality of data currently in use and provide metrics to use in tweaking configurations in our generation pipeline.
- I build AML metrics pipelines to compute engagement and relevance metrics for eyes-off curated Copilot data. These pipelines allow our team to easily monitor the quality of Copilot responses over time.
- I built and extended analysis scripts that served as a playground for offline analysis in our team. This increased agility is utilized in various projects as a framework for evaluation for shipping decision making.

- I worked on an initial framework for disambiguating apologetic responses which laid groundwork for understanding and categorizing user dissatisfaction in M365 Copilot responses.
- I conducted an in-depth analysis and reported on the reliability of key LLM-based metrics we use for evaluating M365 Copilot response quality. This report offered insights into improving the metrics and improved our understanding of the limitations of our inner evaluation framework.
- Led the work on providing contextualized answers to people queries in M365 Copilot. This was the first work towards delivering contextual people responses and has shipped to a global audience:
 - Built an offline experimentation pipeline for data ingestion, preprocessing, selection and ranking based on the user query.
 - Conducted prompt engineering for the synthesis prompt used for this scenario.
- Built and packaged a prompt-autotuning library for Language Understanding classification task. Work described in this paper: [Prompt Engineering a Prompt Engineer](#).

Software Engineer [59, 60]
People Search Team

- I implemented a prototype for ingesting contextual information to improve explainability for search retrieved file entities. I collaborated with the Context team to extend their endpoints to support this scenario. This work consequently laid groundwork for using contextual information to improve people queries responses in M365 Copilot.
- Implemented and rolled out algorithms to calculate relevance scores for ranking suggestions for chats and channels.

IBM RESEARCH

Dec 2018 - July 2020

Software Engineer
Agriculture Team

- Maize mapping, Rwanda: I built a Sentinel-1 ingestion and processing pipeline as part of a team project to detect maize growing areas in Rwanda.
- Farm boundary detection, HelloTractor: I implemented and deployed a farm boundary detection service that used tractor traces to discover farms.
- Published “ADW: Blockchain-enabled Small-scale Farm Digitization”. [Paper](#)

Software Engineering Intern
Water Team

- Internship project: I conceptualized and implemented a system for trading groundwater credits for sustainable water usage.
-

CERTIFICATIONS

Offensive Security

Oct 2025 - Dec 20256

Offensive Security Exploit Developer | [Credential](#)

COMPETITIONS AND ACHIEVEMENTS

CHASINGFLAGS CTF TEAM

Sept 2021 - Present

CTF Player, teacher and CTF organizer | [LinkedIn](#) / [Website](#)

- As part of the ChasingFlags CTF team, I solve, teach and create reverse engineering and cryptography challenges.
 - Highlights: 2nd place AfricaHackOn 2021, 1st place SheHacks (solo), 3rd place AfricaHackOn 2019 (solo).
 - Our team organizes CTFs for local university students and the community.
-

Strathmore National Math Contest

2014

Top student

- Top female student in 2014 National Math Contest. Secured full scholarship for my bachelor university studies.