

JACOB JOHN

Software Engineer | AI Data Annotation and Model Evaluation Specialist

Nairobi County, Kenya | jacobkasunzu1@gmail.com | linkedin.com/in/jacobjohn01 | Available for Remote Roles

PROFESSIONAL SUMMARY

Research-focused software engineer and AI data specialist with hands-on experience in machine-learning data pipelines, multimodal annotation, model evaluation, Python development, and human-feedback workflows. Brings a rigorous, evidence-driven approach to dataset quality, rubric design, error analysis, and reproducible technical documentation. Experienced collaborating remotely with cross-functional teams to improve the usefulness, safety, and reliability of AI systems.

TECHNICAL SKILLS

Programming	Python, SQL, R, Java, C++, JavaScript
AI and ML	PyTorch, TensorFlow, Keras, Hugging Face Transformers, NLP, LLM evaluation
Data Annotation	Text, audio, image and video annotation; quality assurance; taxonomy design; rater calibration
Engineering	REST APIs, data preprocessing, validation workflows, Git/GitHub, Linux, Docker fundamentals
Data Systems	PostgreSQL, MySQL, dataset curation, exploratory data analysis, performance reporting
Research Practice	Experimental design, error analysis, prompt evaluation, technical writing, reproducible documentation

PROFESSIONAL EXPERIENCE

Machine Learning Specialist | Outlier

Nov 2023 - Present

Remote

- Evaluate AI-generated responses against detailed quality, reasoning, and instruction-following criteria, producing actionable feedback for model improvement.
- Contribute to structured training-data workflows, including prompt design, response annotation, preference ranking, and failure-mode identification.
- Apply careful quality-control practices to improve consistency, clarity, and relevance across AI-training datasets.
- Work independently in an asynchronous environment while documenting decisions and adhering to evolving evaluation guidelines.

Freelance Developer Consultant (Python) and AI Tutor | Mindrift

Feb 2023 - Sep 2024

Remote

- Developed structured prompts and human-interaction scenarios to help train and evaluate generative AI systems.
- Annotated and reviewed text, audio, and video data, emphasizing label consistency, contextual correctness, and useful feedback loops.
- Supported data-curation activities and technical documentation for model-training and validation workflows.
- Translated complex technical concepts into clear, user-focused guidance for diverse audiences.

Senior Software Development Engineer | JHUB Africa

Mar 2024 - Aug 2024

Nairobi County, Kenya

- Supported AI and simulation research through data gathering, data-quality review, and engineering collaboration.
- Designed annotation-support workflows that improved the organization and usability of data collected for model development.
- Worked with research and technical stakeholders to integrate real-world data into infectious-disease simulation initiatives.
- Provided data-driven observations that supported iterative model and workflow improvements.

Software Engineer Intern | TechWorks Ltd.

Jun 2022 - Aug 2022

Nairobi, Kenya

- Supported mobile-application development, user testing, and documentation of technical workflows.
- Assisted with dataset preparation and label-assignment checks used in AI-model validation activities.
- Created clear workflow documentation to improve coordination between development and data-annotation teams.

Junior Data Scientist | Kilimanjaro Solutions

Jan 2021 - May 2022

Nairobi, Kenya

- Contributed to data-labeling and dataset-curation activities for machine-learning applications.
- Prepared structured training data for consumer-behavior and supply-chain analysis use cases.
- Supported data-quality checks and organization of large datasets for training, testing, and validation.

SELECTED ENGINEERING AND RESEARCH WORK

LLM Evaluation and Annotation Workflows

Applied structured rubrics to assess model outputs for factuality, instruction adherence, reasoning quality, and user usefulness. Documented error patterns to support targeted model improvement.

Multimodal Dataset Quality Assurance

Reviewed annotations across text, audio, image, and video tasks, focusing on ambiguity resolution, consistency checks, and adherence to labeling instructions.

Python Data Preparation and Analysis

Used Python-based workflows to organize, clean, validate, and analyze datasets for AI-training and research use cases.

AI Interaction Scenario Design

Created prompt-based scenarios that capture realistic human needs and edge cases, helping surface model behavior across varied contexts.

EDUCATION

Bachelor of Science in Computer Science

2022 - 2026

Jomo Kenyatta University of Agriculture and Technology (JKUAT) | Kenya

PROFESSIONAL DEVELOPMENT

- Python for Data Science (2024)
- Machine Learning Specialization - Coursera (2023)
- AI and Deep Learning - Udemy (2023)
- Advanced Data Structures and Algorithms - edX (2023)

RESEARCH INTERESTS

Reliable AI systems, human feedback and preference learning, LLM evaluation, data-centric AI, natural language processing, scalable annotation systems, AI safety evaluation, and quality assurance for multimodal datasets.

ADDITIONAL INFORMATION

- Languages: English (fluent), Spanish (conversational), French (basic reading and writing).
- Remote collaboration: Comfortable working across time zones with Slack, Zoom, GitHub, and shared documentation tools.
- Work setup: Reliable internet connection and computer equipment for remote technical work.