CLEMENT SAMPEBGO

1 University Avenue, Berekuso
Phone numbers: +233 50 384 8241/+226 77 54 07 92.
Email: clement.sampebgo@ashesi.edu.gh
LinkedIn: www.linkedin.com/in/clement-sampebgo

EDUCATION

Ashesi University — **Berekuso, Eastern Region**
BSc. Computer Science — **Expected Date of Graduation: Jun 2027**
Cumulative GPA: na/4.00

RESEARCH & PUBLICATIONS

Global PIQA: Evaluating Commonsense Reasoning Across 100+ Languages and Cultures **2026**
Contributing Author | Accepted, NeurIPS 2026 Evaluations & Datasets Track (poster) | arXiv:2510.24081

- Helped build the Zarma dataset (low-resource West African language) for a 141-variety benchmark created by 350+ researchers from 65+ countries, working with M. K. Keita (Rochester Institute of Technology).
- Prepared prompts, filtering and formatting for LLM-generated examples and reviewed French seed examples for cultural relevance.
- Zarma set proved among the hardest: one of only 8 languages where the best model scored below 80% (79%).

ACHIEVEMENTS & AWARDS

Paper accepted at NeurIPS 2026 (Evaluations & Datasets Track, 25.8% acceptance rate) **September 2026**
Recipient, The MasterCard Foundation Scholar Program, Ashesi University **2023 – 2027**
Recipient, The French Government Scholarship, CESI Graduate School of Engineering, France **2025 – 2026**
Nominated for an Exchange Program, CESI - Aix-en-Provence Campus, France **2024 – 2025**
Dean's List, Semester 1, 3, Ashesi University **2023 – 2025**

WORK EXPERIENCE

Extern, Pfizer Advanced: AI-Powered Document Insights & Data Extraction, **July 2026 – October 2026**

- Developed Python-based document processing pipelines for pharmaceutical PDFs using Tesseract and PaddleOCR for OCR, extraction, and layout-aware processing.
- Built RAG workflows with LlamaIndex and open-source LLMs (Mistral, Phi-2), optimizing document chunking and retrieval strategies.
- Designed document classification and routing workflows to split bundled PDFs, identify document types, and direct files to appropriate extraction pipelines.
- Built a Gradio RAG chatbot and conducted end-to-end evaluation of OCR accuracy, retrieval quality, and routing performance, documenting optimization strategies and deployment recommendations.

Extern, SQL & Database Architecture Externship with Breaking Games **July 2026 - September 2026**

- Designed a normalized analytics database in SQLite, including dimension and fact tables linked via primary and foreign keys to unify six disconnected e-commerce and marketing data sources into a single schema.
- Wrote SQL queries (JOINs, GROUP BY, HAVING, window functions, correlated subqueries) to diagnose real data-quality issues, including slug-mismatched product attribution and bot/fraud traffic, recalculating a cart-abandonment rate from a naive 91.9% down to a verified 37.6%.
- Built a denormalized reporting layer and a static HTML/Chart.js dashboard presenting five KPIs, a product leaderboard, and channel-performance visualizations for non-technical stakeholders.
- Delivered a data-backed Q4 marketing and inventory recommendation, including a reorder plan and an ad-spend reallocation proposal, defended in a stakeholder presentation with anticipated pushback and stated limitations.

National Geographic Society + TNC, Marine Conservation Externship, **September 2025 - November 2025**

- Applied GIS, spatial analysis, and mapping techniques to assess environmental risks and biodiversity trends.
- Used data cleaning, visualization, and exploratory analysis to support evidence-based conservation decisions.
- Created an **ArcGIS StoryMap** integrating geospatial data, and analytical findings to communicate key patterns.
- Developed a concise research brief with data-driven recommendations for restoring the Tanghin Dam ecosystem.

Beats by Dre, Consumer Insights Data Analytics Externship **June 2024 - July 2024**

- Conducted sentiment analysis of consumer reviews using Python and data science libraries.
- Utilized Exploratory Data Analysis (EDA) to identify consumer behavior trends.
- Developed data visualizations to support strategic brand decision-making.

VOLUNTEER & CO-CURRICULAR ACTIVITIES

Youths4Change Initiative member, Ashesi University **December 2024 – Present**

- Mobilize peers to drive social impact initiatives addressing community challenges.
- Organize awareness campaigns and educational programs for underrepresented youth empowerment.

Peer Tutor, Ashesi University **Jun 2024 - Present**

- Facilitate interactive learning sessions in Python for *Introduction to Computing and Information Systems*.

TRAINING & CERTIFICATION

Prompt Engineering for ChatGPT – Vanderbilt University **December 2024 – January 2025**

- Designed optimized prompts for AI performance in data analysis, problem-solving, and creative writing.

Google, Data Analysis Professional Certificate **October 2024 – January 2025**

- Mastered data cleaning, visualization, and statistical analysis using R programming.
- Developed expertise in data-driven decision-making processes.

SKILLS & INTEREST

Languages: Bilingual in French (native) and English (professional proficiency).
Technologies: Microsoft Tools, Google Workspace, RStudio, Tableau, Jupyter Notebook, DB Browser for SQLite.
Technical Skills:

- **Programming:** Python, SQL, Java, C++, C, JavaScript, R.
- **AI & Machine Learning:** OCR, Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP), LlamaIndex, document intelligence, LLM evaluation, benchmark construction, multilingual NLP
- **Databases:** SQLite, relational database design, dimensional modeling, star schema
- **Data Analysis & Visualization:** Pandas, Matplotlib, Tableau, Chart.js, exploratory data analysis
- **Web Development:** HTML, CSS, React, Next.js
- **Tools:** Jupyter Notebook, RStudio, DB Browser for SQLite, Git

Soft Skills:

- Adaptability, Leadership, Communication, Critical Thinking, Time Management.
- Collaboration, Problem-Solving, Attention to Detail, Organization, Presentation.

Sports: Gym, Endurance Running, Pickleball.