

# Mahi Aminu Aliyu

Along Dass Road, Bauchi, Nigeria

mahigital@gmail.com | a-l-mahi.github.io | github.com/A-l-Mahi

## RESEARCH INTERESTS

My research focuses on how machine learning and deep learning models learn meaningful representations, with a particular emphasis on causal representation learning, inductive biases, and reinforcement learning.

## EDUCATION

**M.Sc., Computer Science**, Abubakar Tafawa Balewa University, Nigeria 2023 – Present

*Thesis*: Location-agnostic Water Quality Model

*Advisors*: Prof A.Y. Gital & Dr B.I. Ya'u

**B.Sc., Computer Science**, Al-Qalam University Katsina, Nigeria 2018 – 2023

*Thesis*: Water Quality Prediction Model

*Advisor*: Dr Mansir Abubakar

## EXPERIENCE

**Independent Researcher** 2024 – Present

*ML Collective (Remote)*

Research focus: Causal Representation Learning, ML Theory and Optimization, Multimodal AI, Multilingual datasets.

**Data Engineer & Analyst** 2023 – Present

*Horizon IT Services (Part-time)*

- Curated a low-resource Hausa dataset by extracting and cleaning 500+ text entries from Al-Barka radio station's news archives
- Applied advanced data cleaning techniques to transform 2,000+ raw Hausa news entries into a structured CSV dataset
- Collaborated on dataset enhancement by introducing categorization and optional labeling, improving data usability by 30%

## PUBLICATIONS & MANUSCRIPTS

[1] **Towards Robust Generalization in African AI: Causal Inference and Domain Shift Mitigation** [Manuscript]

*Mahi Aminu Aliyu*, Chisom Chibuike, Fatimo Adebajo, Omokolade Awosanya, Samuel Oyeneye

[2] **Water Quality Prediction using USGS Data: A Machine Learning Approach for Automated Assessment Systems** [Under Review '25]

Mansir Abubakar, *Mahi Aminu Aliyu*, Fatimo Adebajo, Armaya'u Umar Zango, Rukayya Ismail

[3] **AfriCaption: Establishing a New Paradigm for Image Captioning in African Languages**, openreview.net [Under Review '25]

Mardiyyah Oduwole, Prince Mireku, Fatimo Adebajo, Oluwatosin Olajide, *Mahi Aminu Aliyu*, Jekaterina Novikova

## CONFERENCES & SUMMER SCHOOLS

**Machine Learning Summer School (MLSS) – Participant** 2025

Selected to attend advanced ML training led by top global researchers

**Deep Learning IndabaX Nigeria – Participant** 2024

Participated in African-led ML training, discussions, and community-building events

## SELECTED PROJECTS

**Cross-Lingual Aspect-Based Sentiment Analysis in African Languages**

- Translated SemEval-2014 to 17 African languages (afri-SemEval-2014) with aspect alignment heuristics
- Fine-tuned XLMR, AfroXLMR-large, and AfroXLMR-large-76L on aspect-based sentiment task
- Designed and evaluated causal vs non-causal variants to compare generalization performance
- Causal models showed superior OOD performance with fewer training steps

**Dono Health**

- Led development of inclusive insurance MVP with Health Center, Vendor, and Insured modules

- Designed serverless AWS backend with Lambda, API Gateway, and CDK
- Integrated Flutterwave for payments and Africa's Talking USSD API for voucher-based funding

### SageBot

- Built AI tool that summarizes technical docs for developers, cutting review time by 40%
- Leveraged LLMs (ChatGPT, LLaMA-3), Hugging Face, and Langchain for high summarization accuracy
- Improved developer workflow efficiency by enabling 50% faster comprehension of complex documents

## TALKS & PRESENTATIONS

---

### MLSS-Senegal '25

2025

Poster presentation: "Towards Robust Generalization in African AI: Causal Inference and Domain Shift Mitigation"

### ML Collective #25 Research Jam

2024

Oral Presentation: "Pretrained Models for Water-based Applications"

## COMMUNITY SERVICE

---

### City Lead, Data Science Nigeria

2024 – Present

- Organize AI workshops, hands-on sessions, and reading groups for 150+ community members
- Promote AI adoption and capacity building across underserved sectors

### DeepLearning.ai Ambassador

2023 – Present

- Facilitated knowledge sharing and engaged 100+ peers as an active community ambassador
- Hosted "Growing an AI Community" event in Bauchi with 50+ participants
- Organized and participated in 5 cross-community events to strengthen the regional AI ecosystem

### Lead, Google Developer Student Clubs

2021 – 2022

- Led GDSC community, organizing 10+ workshops and technical sessions impacting 350+ students
- Applied Google technologies to address UN SDGs, delivering solutions to 5 local communities

## TECHNICAL SKILLS

---

**Programming Languages:** Python, C++, R

**ML/AI Frameworks:** PyTorch, TensorFlow, Hugging Face, Scikit-learn

**Web Frameworks & Tools:** Django, Flask, Git/GitHub

**Cloud Platforms:** Microsoft Azure, AWS & GCP

**Operating Systems:** Linux

**Data Analysis:** Pandas, NumPy, Matplotlib, Seaborn

## RELEVANT COURSES

---

- Machine Learning Specialization, Coursera – DeepLearning.ai & Stanford 2024
- Statistical Inference, Coursera – Johns Hopkins University 2023
- Data Science: Foundations using R Specialization, Coursera – Johns Hopkins University 2022