CHIBUZOR AUTHUR OKAFOR
LinkedIn
Portfolio

Igbojia, Ibeju-Lekki Lagos State
Email: chibuzorauthur@gmail.com
Mobile: +234 9036546571

EDUCATION

Nnamdi Azikiwe University — Anambra, Nigeria
BSc Soil Science and Land Resources Management — **August 2017 – April 2023**

Community Secondary School — Anambra, Nigeria
Senior Secondary School Certificate Examination — **September 2010 – July 2016**

SKILLS SUMMARY

Languages:	Python, SQL, R
Frameworks:	Pandas, Numpy, Scikit-Learn, Matplotlib, Keras, Pytorch, TensorFlow, NLTK, SpaCy, Gensim, FastAPI, LangChain, LangGraph, LangSmith
Tools:	RStudio, Excel, PowerPoint, Tableau, BigQuery, SQLite, Watson Studio, Dialogflow, MySQL, FAISS, PostgreSQL
Platforms:	PyCharm, Jupyter Notebook, Visual Studio Code, Heroku, Streamlit, Google Cloud Platform, DigitalOcean, Amazon
Techniques:	Clustering (K-Means, DBSCAN), Word Embedding (Word2Vec, FastText, Bag of Words, Glove, TF-IDF Vectorization, Transformers, OpenAIEmbeddings), HuggingFace, OpenAI, Unsupervised Learning
Soft Skills:	Cross-functional collaboration, Strong analytical and data interpretation, Time management, Attention to details, Excellent communication, Project management, Agile, Scrum, Problem-solving, Documentation, Adaptability, Technical Communication, Efficiency, Independence

WORK EXPERIENCE

NUPAT TECHNOLOGIES, DATA SCIENTIST INTERN — **May 2024 - July 2024**

- Conducted a competitive analysis of a competitor's company and identified necessary upgrades to their website, increasing its interactivity by 90%.
- Assisted students in debugging code, understanding programming concepts, and completing assignments
- Conducted training sessions and workshops on Python programming for new students, covering fundamental to advanced concepts
- Collaborated with other team members to improve training content and methodologies

FREELANCE DATA SCIENTIST & MACHINE LEARNING ENGINEER — **August 2024 - Present**

1. **Retrieval-Augmented Question-Answering System with Google Sheets Integration**

- I developed a Retrieval-Augmented Generation (RAG) system that integrated Google Sheets as a knowledge base for real-time question-answering, improving information retrieval accuracy.
- Implemented document embedding and FAISS indexing to enable fast and precise search of relevant answers from large datasets, enhancing response relevance.
- Optimized the system to return answers with source links, ensuring users could verify and trace the origin of information for greater reliability.
- Reduced manual data processing by automating knowledge updates into Google Sheets, improving the efficiency of maintaining the knowledge base and query accuracy.

2. **Whisper Installation and API Creation**

- Installed and configured Whisper's most accurate English language model, whisper-large.en, on a Linux-based server, creating an API to process and transcribe audio files into text.
- Developed /transcribe API endpoint to accept MPGA and MP3 file formats via POST requests or webhooks, with a focus on scalability to handle 100 weekly transcriptions of files up to 2 minutes in length.
- Validated the API using real audio files, ensuring high transcription accuracy and successful implementation without relying on OpenAI's services.
- Delivered a solution optimized for efficient server performance, ensuring that the system could handle current demands and scale for future growth.

3. **Data Visualization Tool for Financial Analysts**

- Developed a tool that allows business analysts to upload Excel files, where each tab is indexed using LLM and Retrieval-Augmented Generation (RAG) technologies.
- Implemented a querying system that generates conversational data visualizations, such as bar charts in response to user queries like "Show me Company XYZ's revenue."
- Designed the system to support multiple projects, enabling users to switch between datasets, create visualizations for different business scenarios, and ensure data accuracy through RAG.
- Provided a seamless user experience for business professionals to interact with their data, reducing time spent on manual analysis and enhancing data-driven decision-making.

4. **Applied GenAI Engineer**
   - Designed and implemented an AI-driven proposal generation platform (FastAPI + LangChain/LangGraph + OpenAI) with multi-agent workflows (intent routing, RAG retrieval, structural drafting, reflexion/critic loop) to automate RFQ/RFP response creation.
   - Built scalable retrieval architecture using PostgreSQL + pgvector and optimized document ingestion (PDF parsing, Google Drive/Docs integration, semantic chunking) to enable high-relevance, low-latency context retrieval for generation.
   - Engineered performance layer with Redis caching, Celery background/periodic workers, and database connection pooling; cut repeated query and generation latency dramatically while supporting asynchronous long-running tasks.
   - Delivered production-grade DevOps: Dockerized micro-services, blue-green deployments via Nginx, and a GitHub Actions CI/CD pipeline (tests, security scanning, load testing, automated rollback) enabling zero-downtime releases and rapid iteration.
   - Enforced robustness and observability with Pydantic-validated LLM outputs, health & metrics endpoints, cache/stat introspection APIs, structured logging, and database optimization routines (index & full-text strategy) to sustain growth and maintain reliability.

5. **AI Engineer**
   - Architected a multi-agent conversational AI platform using FastAPI, LangChain, and LangGraph; implementing LLM-driven intent routing, context summarization, and dynamic workflow branching to reduce irrelevant agent calls and improve response relevance.
   - Engineered automated chart generation and data retrieval pipeline (SQL + vector similarity search) that transforms natural language queries into structured insights and frontend-ready visualization configs, accelerating analytical turnaround time.
   - Designed and enforced robust state management with checkpointed graph execution (PostgreSQL) and background memory summarization, enabling persistent conversational intelligence and lowering token usage via strategic history compression.
   - Implemented human-in-the-loop decision agents (table selection + interrupt handling) and context sufficiency evaluation, increasing accuracy of downstream retrieval and reducing unnecessary database scans across multi-tenant datasets.
   - Refactored and stabilized asynchronous API endpoints with resilient error handling, type-safe state propagation, and structured JSON outputs (including chart configs), improving system reliability and integration readiness for production deployment.

PROJECTS

EmailGen App | DEMO **October 2025**

- Built a full-stack email automation platform (FastAPI + Next.js + Docker) with AI-ready ingestion (document & CSV uploads) and scalable async workers (Celery) for background email generation and notifications.
- Implemented secure authentication/session management and role-based admin controls, including referral tracking, subscription billing (monthly/yearly), crypto & bank payment flows, and withdrawal processing with balance ledger logic.
- Engineered robust email lifecycle: campaign workflow, scheduling, delivery status tracking, and Gmail API integration with structured notification agents for payments and withdrawals.
- Designed performant REST API layer (pagination, filtering, validation) and optimized database interactions for transactional actions (transactions, balances, withdrawals) with clear audit logging.
- Integrated real-time user support (Tawk), admin tooling, and analytics-friendly data models; standardized UX components (history tables, tabbed dashboards) for maintainable, production-grade UI.

CyberDomain Bot | LINK **October 2024**

- Developed a sophisticated search application aimed at improving information retrieval in the cybersecurity domain by leveraging advanced natural language processing techniques.
- Implemented GPT-4 for query expansion and reranking, utilizing LangChain for streamlined information retrieval and processing.
- Designed custom prompt templates that enhance search accuracy, enabling the generation of relevant query variations and prioritization of results.
  Provides researchers and professionals in cybersecurity with enhanced tools for efficient document searches, leading to improved insights and informed decision-making.

Breast Cancer Diagnosis Prediction | LINK **September 2024**

- Developed a predictive model for breast cancer diagnosis using Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel, achieving an accuracy of 93%.
- Conducted data preprocessing, feature selection, and model training on a dataset containing tumor characteristics to classify tumors as malignant or benign.
- Evaluated model performance using metrics such as accuracy, precision, recall, and F1-score, demonstrating high robustness with a weighted F1-score of 0.93.
- Provided insights for healthcare professionals to aid in early cancer detection and improve patient outcomes through accurate diagnosis.

Customer Segmentation | LINK **July 2024**

- Implemented three clustering methodologies - K-Means Clustering, Hierarchical Clustering, and DBSCAN - to analyze customer profiles within a supermarket dataset.
- Applied clustering models to pinpoint distinct customer segments, enabling targeted strategies tailored to specific customer needs. Highlighted the effectiveness of KMeans Clustering in delineating clusters and DBSCAN in identifying outliers
- Analyzed customer spending patterns to devise customized strategies for membership enrollment and marketing campaigns. Prioritized targeting high-spending customers for membership enrollment while devising specialized promotions to engage low-spending segments.
- Translated complex analytical findings into actionable insights for business decision-making. Provided clear recommendations for future strategies based on insights derived from clustering analysis.

CERTIFICATES

Google Prompting Essentials | **CERTIFICATE** **September 2025**

- Prompt Engineering & LLM Applications – Designing effective prompts, applying structured prompting techniques, and leveraging large language models (LLMs) for real-world tasks.
- Data Analysis & Visualization with AI – Using prompting to accelerate data analysis, generate insights, and create clear visualizations and presentations.
- Complex Problem-Solving & Ideation – Applying AI-driven prompt patterns to brainstorm solutions, synthesize data, and tackle business challenges creatively.
- AI Agent Design & Automation – Building prompt-based AI agents for role-play, workflow automation, and expert feedback simulations.

Python and Statistics for Financial Analysis | **CERTIFICATE** **May 2024**

- Skills to import, pre-process, save, and visualize financial data using Python's pandas DataFrame.
- Techniques to manipulate existing financial data by generating new variables from multiple columns.
- Application of important statistical concepts such as random variables, frequency distributions, population and sample analysis, confidence intervals, and linear regression in financial contexts.
- Building a trading model using multiple linear regression and evaluating its performance using various investment indicators.

IBM AI Engineering (IBM) | **CERTIFICATE** **April 2024**

- Explored the concepts of machine learning, deep learning, and neural networks, including various ML algorithms such as classification, regression, clustering, and dimensional reduction.
- Implemented supervised and unsupervised machine learning models utilizing SciPy and Scikit-Learn libraries.
- Deployed machine learning algorithms and pipelines efficiently on Apache Spark for scalable and distributed computing.
- Developed deep learning models and neural networks using industry-standard frameworks like Keras, PyTorch, and TensorFlow to tackle complex data challenges effectively.

Google Advanced Data Analytics (Google) | **CERTIFICATE** **April 2024**

- Investigated the diverse roles of data professionals within organizational structures to understand their impact on data-driven decision-making processes.
- Developed expertise in creating compelling data visualizations and applying statistical methods to explore and analyze complex datasets effectively.
- Applied advanced regression and machine learning techniques to construct predictive models and extract valuable insights from data, enhancing decision-making capabilities.
- Demonstrated proficiency in communicating insights derived from data analysis to stakeholders through clear and concise presentations and reports, facilitating informed decision-making processes within the organization.

Google Data Analytics (Google) | **CERTIFICATE** **October 2023**

Developed a comprehensive understanding of the data life cycle and various stages involved in data analysis.

- Mastered data visualization techniques, proficient in creating insightful dashboards, presentations, and utilizing popular visualization platforms to effectively communicate data findings.
- Skilled in data cleaning and organization techniques for analysis, proficient in conducting thorough analyses and calculations using spreadsheets, SQL, and R programming.
- Learned essential analytical skills encompassing data cleaning, analysis, and visualization, proficient in utilizing key tools such as spreadsheets, SQL, R programming, and Tableau to extract actionable insights from complex datasets.