SYDNEY ANUYAH

+1(463) 269-6271 \diamond Indianapolis, IN (open to relocation)

sydneyanuyah@gmail.com \diamond LinkedIn \diamond GitHub \diamond Google Scholar \diamond Tableau

OBJECTIVE

I am a Data professional skilled in both Data Science and Data Engineering with experience in research science, specializing in computer vision, natural language processing and big data analytics. My research focuses on understanding causal relations in NLP and furthermore negating biases in text generation

EDUCATION

 PhD in Data Science , Indiana University, Indianapolis
 2024 - 2028

 Research Interests: Deep Learning, Natural Language Processing, Large Language Models, Causality and Correlation, Computer Vision
 2022 - 2024

 Master of Science in Applied Data Science, Indiana University-Purdue University, Indianapolis
 2022 - 2024

 Relevant Skills Gained: Calculus and Distributed Systems: PySpark Machine Learning, Deep Learning, Reinforcement Learning, Computer
 Vision, Natural Language Processing, Data Mining, Database Systems

 Master of Science in Financial Engineering, WorldQuant University, New Orleans
 2021 - 2023

 Relevant Skills Gained: Machine Learning in Finance, Portfolio Optimization with Python, Financial Data and Markets, Discrete and
Continuous-time Stochastic Processes
 CGPA: 3.69/4.00

 Bachelor of Science in Electrical and Electronics Engineering, University of Lagos, Nigeria
 2014 - 2020

 Relevant Skills Gained: Engineering Mathematics, Switching And Logic System, Computer Programming, Engineering Statistics & Computer
 Statistics & Computer

 System
 CGPA: 3.46/4.00

TECHNICAL SKILLS

Programming Languages	Python, R, SQL, C++, MATLAB
Technical Tools	Visual Studio, VS Code, Git, Docker, Azure DevOps, Apache Spark, Mixpanel
Computer Vision Frameworks	YOLO, RCNN, ONXX, TensorFlow, Vision Transformer models, PyTorch,
Natural Language Processing Frameworks	NLTK, SpaCy Gensim, Hugging Face Transformer models BERT, Flair
Machine Learning Frameworks	PySpark, RDDs, Scikit-learn, Keras
Data Visualization	Tableau, Power BI, matplotlib
Cloud Platforms:	AWS, Azure Cloud Services, Google Cloud Services,
Spreadsheets	Microsoft Excel, Google Sheets

OTHER SKILLS

- Understand software engineering architectural process including design, testing, debugging, documentation and deployment.
- Proficient in creating statistical models, analyzing data, and utilizing techniques for causal inference.
- Skilled at simplifying and presenting intricate data findings to anyone without technical backgrounds
- Excellent problem-solving, analytical, communication, collaboration, and interpersonal skills;

EXPERIENCE

Data Science Graduate Assistant - Dr. Chakraborty

Luddy School of Informatics, Computing, and Engineering

- Spearheaded an individual research project titled "Can Deep Learning Large Language Models be Used to Unravel Knowledge Graph Creation?" which holds potential to revolutionize how we extract and visualize connections in GERD-related literature, enhancing our understanding and application of this knowledge. The paper detailing our work can be found here.
- Collaborated with lab partners to ensure timely running of experiments, Data Transformation, ETL processes which includes database management and data parsing
- Led weekly discussion sessions, clarifying complex topics and enhancing student comprehension for Cloud Computing classes
- Graded assignments, quizzes, and exams, providing timely and constructive feedback to students studying Cloud computing

Data Science Intern

Indiana IOT Laboratory

• Responsible for the installation, operations, and data analytics of Vision AI platform within an Indianapolis-area manufacturer's existing operation.

Data Science Intern Amazon May 2023 - September 2023 Bellevue, WA

January 2024 - May 2024

Fishers, IN

August 2022 - Present Indianapolis, IN

- Developed Project Toucan, a computer vision model for Amazon warehouses which saved the company \$1.2 Million anually per site. Performed a lot of ground training and model finetuning.
- Created Machine Learning models which made optimized the entire work flow of Amazon fulfilment centers.
- Got my internship extended because of the enormous progress I added to the team.

Data Science Research Assistant - Dr. Davida Bolchini

Luddy School of Informatics, Computing, and Engineering

• Led the programming aspect of our lab collaboration to develop the "Question Generating Dataset (QGD)" using Siddharth Dushantha's Sherlock module. Our innovative AI methodology attracted considerable attention at an AI conference in Thailand. The paper detailing our work can be found here.

Data Engineer

Chaka Technologies

- Developed, constructed, tested, and maintained data architectures, aligning these architectures with business requirements.
- Used Python, MSSQL, Mixpanel, Azure, and many others for data reporting thereby identifying ways to improve data reliability, efficiency, and quality.
- Created automated pipelines using Apache Airflow hosted using Docker containers for automatic stored procedures in SQL

Data Analyst

Edan Investments

December 2019 - May 2021 Lagos, Nigeria

April 2023 - May 2024

July 2021 - June 2022

Lagos, Nigeria

Indianapolis, IN

- Collaborated with the data science team to develop predictive models for customer churn analysis using machine learning algorithms.
- Conducted exploratory data analysis on large data sets using SQL, Python, and data visualization tools.

CURRENT PROJECTS

Do Transformer Models Understand Cause-Effect Relationships?

An Empirical Study of Causal Relation Extraction Transfer: Design and Data

Do Large Language Models like BART and GPT-2 have inherent biases towards the African and Black population?

PAST PROJECTS AND PUBLICATIONS

Can Deep Learning Large Language Models be used to unravel Knowledge Graph Creation. Published with ACM The paper can be found here.

Question Generating Dataset (QGD) which takes datasets from IndyHub and analyses them to generate insightful analytical questions. The paper can be found here.

Deploying an accurate object detection model for Amazon Fulfilment centers. The paper can be found here.

Comparing Automated Medical Diagnosis of Breast Cancer using CNN model and Vision Transformers (VIT) models which resulted in VIT performing slightly better with a 90% accuracy against the 87% baseline accuracy. The paper can be found here.

Built a passlock, self-timing solar inverter and wrote the methodology in a research paper. The small inverter was used to power some sockets and charge multiple devices directly from the sunlight. The paper can be found here.

Comparative Analysis of 3D-CNN Models, GARCH-ANN, and VAR Models for Determining Equity Prices which resulted in the identification of the most accurate model for determining equity prices. Currently under review for publication

POSTERS

What Causes Causality? Exploring Causal Expression and Understanding of Language Models.

Do Large Language Models like BART and GPT-2 have inherent biases towards the African and Black population?.

MEMBERSHIP/EXTRACURRICULAR

- Member, Association for Computing Machinery (ACM)
- Member, National Society of Black Engineers, Indianapolis Chapter (NSBE)
- Member, Institute of Electrical and Electronics Engineers (IEEE)
- Member, Honor Society Club

SCHOLARSHIPS AND AWARDS

Luddy School of Informatics Scholarship Funded by Indiana University under the graduate program.