

# VICTOR OWINO

Newark, NJ 07106 | (862) 899-0511 | victorowino@gmail.com  
linkedin.com/in/victorowinoke | github.com/VictorOwinoKe

## SUMMARY

---

Data Scientist and Computational Linguistics with solid experience in deep learning, NLP, data analytics, text mining, and linguistics corpus annotation. Experienced in building predictive models, NLP pipelines, LLM evaluation frameworks, data pipelines, ETL workflows, and BI dashboards using Python, SQL, PyTorch, Hugging Face Transformers, Power BI, and data pipelines. Strong background in translating and presenting complex data into actionable insights for cross-functional stakeholders, decision-making, and operational management.

## EXPERIENCE

---

### Research Assistant – NLP and Deep Learning | 09/2025 - 05/2026

Montclair State University - Montclair, NJ, USA

- Conducted NLP and biomedical AI research with Dr. Liu Hao, focusing on Pragmatic Inference Chain (PIC) reasoning with large language models for adverse drug event detection in clinical notes, and on deep learning approaches for predicting blood transfusion adverse events using MIMIC-IV patient records (EHR).
- Collaborated with Dr. Anna Feldman in the NLP Lab on a multilingual euphemism project, leading Swahili euphemism data annotation and contributing to a cross-lingual benchmark evaluating LLM understanding of euphemisms, pragmatic meaning, and machine translation quality.
- Contributed to three research projects submitted for conference paper and poster presentation, with review feedback expected by June 2026.

### Analytical Data Engineer | 09/2023 - 12/2024

Kidogo Innovations Limited - Nairobi, Kenya

- Designed dimensional and relational data models in Google BigQuery to support analytics dashboards, KPI reporting, and self-service analysis for program stakeholders.
- Built Python and SQL ETL/ELT workflows across data lake and warehouse environments, improving data consistency, refresh reliability, and reporting turnaround.
- Developed Power BI dashboards for national and sub-national monitoring, enabling leadership to track performance trends, operational bottlenecks, and program outcomes.

### Health Data Analyst | 10/2021 - 08/2023

Kenyatta National Hospital - CONNECT Program - Nairobi, Kenya

- Built ETL pipelines integrating facility-level data into centralized analytical datasets for routine reporting and performance tracking across 22 facilities.
- Developed dashboards and executive reporting models for senior leadership, CDC teams, and Ministry of Health stakeholders, reducing manual reporting effort by 45%.
- Presented analytical findings through dashboards, performance reviews, and executive reports to strengthen evidence-based decision-making and monitoring cycles.

### Data Science Associate | 01/2021 - 09/2021

University of Nairobi Institute of Tropical and Infectious Diseases (UNITID) - Nairobi, Kenya

- Conducted statistical analysis, data cleaning, and feature engineering with Python, R, and SQL to support infectious disease surveillance and public health research.
- Built analysis-ready datasets and visualizations in R, Excel, and Power BI, improving data exploration and insight delivery for researchers and public health stakeholders.
- Co-developed machine learning and statistical models to identify key drivers of zoonotic MERS-CoV transmission in Kenya, supporting prevention-focused research.

## EDUCATION

---

### Master of Science, Computational Linguistics | Expected 05/2026

Montclair State University - Montclair, New Jersey

- Relevant coursework: Machine Learning, Deep Learning, Natural Language Processing, Special Topics in Generative AI; GPA.

### Bachelor of Science, Computer Science | 03/2021

University of Eldoret - Uasin Gishu, Kenya

- Relevant coursework: Database Management Systems, Object-Oriented Programming, Data Structures and Algorithms.

## RELEVANT PROJECTS AND RESEARCH

---

### **BERT NER for Clinical Trials Eligibility (CHIA)**

**Technologies:** Python, PyTorch, Hugging Face Transformers, BERT, BioBERT, ClinicalBERT, NLP

- Fine-tuned BERT, BioBERT, and ClinicalBERT models for medical named entity recognition, extracting Conditions, Drugs, and Procedures from clinical text using BIO tagging.
- Implemented subword-token alignment and exact-boundary evaluation, achieving entity-level F1 of 0.85 with per-entity and token-level error analysis.

### **Credit Risk Analytics & Default Prediction**

**Technologies:** Python, SQL, scikit-learn, XGBoost, Power BI, EDA, risk analytics

- Built analytical datasets and engineered risk features from loan and customer financial data using Python and SQL.
- Trained Logistic Regression, Random Forest, and XGBoost models using ROC-AUC, precision, and recall; created Power BI dashboards for portfolio risk monitoring.

### **Healthcare Insurance Fraud Detection & Real-Time Analytics Dashboard**

**Technologies:** Power BI, Power Query, SQL, Python, healthcare claims analytics

- Cleaned, transformed, and enriched healthcare claims data containing procedures, diagnoses, and billing amounts using SQL, Python, and Power Query.
- Designed a Power BI dashboard to monitor suspicious claims patterns and support data-driven fraud analytics for healthcare insurance workflows.

## CERTIFICATIONS

---

- Accredited Python Data Engineer
- IBM Certified Specialist: Data Analysis with Python
- Microsoft Certified: Data Analyst Associate

## SKILLS

---

**Programming & Querying:** Python, R, SQL, Bash

**Machine Learning & Deep Learning:** scikit-learn, XGBoost, TensorFlow, PyTorch, model evaluation, feature engineering, classification, regression

**NLP & LLMs:** Hugging Face Transformers, BERT, BioBERT, ClinicalBERT, spaCy, NLTK, prompt engineering, LLM evaluation, text classification, NER, semantic similarity

**Data Engineering:** ETL/ELT pipelines, workflow orchestration, data lakes, data warehousing, REST APIs, JSON, CSV, Parquet

**Databases & Cloud:** PostgreSQL, MySQL, Google BigQuery, Neo4j, AWS, Azure, GCP

**Analytics & Visualization:** Power BI, Tableau, Excel, dashboard development, KPI reporting, statistical analysis

**Tools:** Git, Jupyter Notebook, Google Colab, VS Code