Isaac Kobby Anni

Ph.D. Candidate in Data Science | Machine Learning & AI Research

Bowling Green, Ohio | Open to Relocation isaacka@bgsu.edu | LinkedIn | Portfolio | GitHub

Summary

Ph.D. candidate and machine learning researcher developing reproducible AI and data-science systems in Python. Experienced in predictive modeling, language models, automated feature engineering, experimental design, data pipelines, leakage prevention, and outcome-based model evaluation. Combines applied research with financial-risk modeling, educational analytics, NLP, and experience translating technical results for interdisciplinary and business stakeholders.

Education

Bowling Green State University — Ph.D. in Data Science, Expected Dec. 2027 — Bowling Green, Ohio
Bowling Green State University — M.S. in Data Science, May 2024 — Bowling Green, Ohio
African Institute for Mathematical Sciences — M.S. in Mathematical Sciences, July 2022 — Dakar, Senegal
University for Development Studies — B.Sc. in Financial Mathematics, June 2019 — Ghana

Technical Skills

Programming and Data: Python, SQL, R, pandas, NumPy, SciPy, MongoDB
Machine Learning and AI: scikit-learn, PyTorch, TensorFlow, XGBoost, Hugging Face, language models, NLP, transformers, tabular ML, classification, clustering, time series, computer vision
Research Engineering: Git, Docker, Linux, SLURM, high-performance computing, pytest, JSON Schema, Jupyter, Streamlit, Power BI, Tableau
Methods: Experimental design, cross-validation, paired bootstrap evaluation, hypothesis testing, distribution-shift evaluation, leakage detection, model interpretation, causal inference, reproducible experimentation

Research and Data Science Experience

Research Assistant — *Aug. 2026–Present*
Human–AI Collaboration Research, Bowling Green State University — Bowling Green, Ohio

- Conduct research under Dr. Md Main Uddin Rony on trustworthy human–AI collaboration, including trust calibration, reliance, oversight, and risks arising from AI-assisted decision-making.
- Design structured taxonomies, empirical studies, and reproducible analysis workflows for evaluating AI-supported decisions across interdisciplinary evidence.
- Communicate research methods and findings through manuscripts, presentations, and collaborative technical materials for technical and nontechnical audiences.

Graduate Assistant / Data Scientist — *Aug. 2022–Apr. 2026*
Student Success Analytics & Technologies, Bowling Green State University — Bowling Green, Ohio

- Developed Python, SQL, and Power BI pipelines supporting retention modeling, reporting automation, and institutional decision-making across complex student datasets.
- Built and evaluated a student-persistence model on approximately 3,580 records using eight-fold cross-validation, achieving ROC-AUC of 0.735 while explicitly assessing minority-class performance.
- Delivered dashboards and statistical analyses to university stakeholders while maintaining data quality, privacy, reproducibility, and role-based access controls.

Research Fellow, Natural Language Processing — *Mar. 2022–Aug. 2022*
Computational Language Understanding / ICRISAT — Remote

- Developed reproducible NLP workflows involving text annotation, information extraction, semantic similarity, prompt-based extraction, and experimental evaluation in an international research collaboration.
- Presented progress to interdisciplinary collaborators and supported research resulting in a peer-reviewed ACL workshop publication on subjective-view annotation and model training.

Selected Machine Learning and AI Research

Trustworthy Evaluation of Language-Model-Driven AutoML *Dissertation Research, 2026–Present*

- Developed an evaluation framework for LM-generated AutoML actions spanning validity, incremental predictive value, stability, evaluator generality, robustness, and decision cost.
- Implemented leakage-safe matched baseline–candidate experiments with logistic regression, random forest, and XGBoost using cross-validation, paired bootstrap analysis, and held-out evaluation.
- Built Python workflows for evidence tracking, model-output validation, lifecycle decisions, and dataset-family isolation, supported by more than 300 automated tests.

Language-Model-Driven Automated Feature Engineering *Ph.D. Research, 2024–Present*

- Designed a Python workflow in which language models generate executable features subject to controlled execution, target-exclusion, leakage, and downstream predictive-value checks.
- Produced 23 accepted model-generated features across 12 classification datasets; 20 achieved a bootstrap success rate of 1.0 under the study's prespecified robustness threshold.
- Developed numerical, temporal, and categorical specialist agents with traceable leave-one-dataset-out evaluation across multiple predictive models.

Credit-Risk Modeling and Decision Support *Graduate Research and Applied Project*

- Built and evaluated credit-scoring workflows across 15 financial datasets using feature engineering, predictive modeling, experimentation, and model interpretation.
- Improved approval accuracy by 20%, reduced false positives by 10%, and contributed to an estimated $55K in savings through model-driven decision support.
- Completed an M.S. thesis on structural approaches to credit-risk modeling, connecting mathematical methods with applied financial decision-making.

Automated Data Cleaning and Leakage Detection *Research Project, 2026*

- Built and tested a constrained proposal-and-validation pipeline for model-suggested cleaning operations, downstream evaluation, leakage prevention, and failure handling.

Selected Publications and Presentations

- **Anni, I. K.** and Dasigi, V. G. (2025). "Analysis of Document Representation for Text Classification." *Advances in Information and Communication*, Springer, pp. 261–273.
- Alexeeva, M., et al., including **Anni, I. K.** (2023). "Annotating and Training for Population Subjective Views." *WASSA, ACL*, pp. 416–430.
- **Anni, I. K.** (2026). "SLM for Automated Feature Engineering." Poster presented at the CMD-IT/ACM Richard Tapia Celebration of Diversity in Computing, Atlanta, Georgia.