

Isaac Kobby Anni

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

Ohio, USA | (419) 378-9889 | isaacka@bgsu.edu | LinkedIn | isaackobby.com | Github

Summary

Ph.D. candidate in Data Science with experience developing machine-learning, NLP, and agentic AI systems in Python. Builds reproducible pipelines for model training, feature engineering, controlled execution, leakage detection, and robust evaluation. Applied background includes credit-risk modeling, student-success analytics, computer vision, and research collaboration. Combines quantitative training in data science, statistics, mathematics, and finance with strong technical communication and cross-functional problem solving.

Education

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

Technical Skills

Programming: Python, SQL, R, C++, Java
Machine Learning and AI: Small language models, agentic AI, NLP, transformers, predictive modeling, feature engineering, classification, clustering, time series, computer vision
Frameworks: PyTorch, TensorFlow, Hugging Face, scikit-learn, pandas, NumPy, XGBoost, OpenCV
Tools: Git, Docker, Jupyter, Streamlit, Power BI, Tableau, MongoDB; interactive AI tooling for coding, debugging, experiment planning, and technical documentation
Methods: Cross-validation, bootstrap evaluation, statistical modeling, experimental design, causal inference, model interpretation, leakage detection, reproducible experimentation

Machine Learning and AI Projects

SLM-First Automated Feature Engineering and Validation *Ph.D. Research, 2024–Present*

- Designed a Python workflow in which specialized small language models generate executable features that must pass controlled execution, target-exclusion, leakage, and downstream predictive-value checks.
- Evaluated candidate features with repeated randomized splits, cross-validation gates, fixed held-out tests, and bootstrap robustness analysis across multiple predictive models.
- Produced 23 accepted SLM-generated features across 12 classification datasets; 20 achieved a bootstrap success rate of 1.0 under the study’s prespecified threshold.
- Developed numerical, temporal, and categorical specialist agents and structured experiment metadata for traceable leave-one-dataset-out evaluation.

SLM-First Data Cleaning and Leakage Detection *Research Project, 2026*

- Built a proposal-and-validation pipeline that executes model-suggested cleaning operations under explicit constraints and accepts them through cross-validated downstream evaluation.
- Implemented automated tests for system behavior and failure handling, with 21 tests passing in the current project suite.

Efficient and Privacy-Aware Agentic AI Systems *Dissertation Research*

- Investigate specialized small language models as cost-efficient, locally deployable components for automated data science, evaluating accuracy, latency, memory, privacy, and robustness.
- Develop approaches for corpus provenance, model specialization, validation, and semantic coordination among task-specific AI agents.

Experience

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

- Develop Python, SQL, and Power BI pipelines and analytical products supporting student-success, retention, and institutional decision-making.

- Build predictive and statistical analyses over complex educational data and communicate actionable findings to university stakeholders.
- Maintain reporting and dashboard workflows with attention to data consistency, repeatability, privacy, and secure access.

Research Fellow, NLP / Research Collaboration

Computational Language Understanding / ICRISAT

Mar. 2022–Aug. 2022

Remote

- Contributed to reproducible NLP workflows involving text data, annotation, information extraction, and experimental evaluation across an international research collaboration.
- Supported research that resulted in a peer-reviewed ACL workshop publication on subjective-view annotation and model training.

Data Scientist Jr.

Baamtu Technologies

Oct. 2021–Mar. 2022

Dakar, Senegal

- Developed deep-learning and preprocessing workflows for satellite-image segmentation, achieving 89% intersection over union in project evaluation.
- Integrated model outputs into interactive analytical tools supporting applied, user-facing decision workflows.

Data Scientist Jr.

Obertys Technologies

Jul. 2021–Mar. 2022

Dakar, Senegal

- Built and evaluated credit-scoring workflows across 15 datasets, combining 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 while communicating modeling tradeoffs to business and technical stakeholders.

Selected Publications and Research Activity

- **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., Hyland, C., Alcock, K., Beal Cohen, A. A., Kanyamahanga, H., **Anni, I. K.**, and Surdeanu, M. (2023). “Annotating and Training for Population Subjective Views.” *WASSA, ACL*, pp. 416–430.
- Accepted poster presentation on SLM-first automated feature engineering, CMD-IT/ACM Richard Tapia Celebration of Diversity in Computing, September 2026.