

Kachinga Silwimba
Data Scientist • Data Engineer

Aurora, CO 80016 • kachingasilwimba@u.boisestate.edu • (+1) 208-831-0244 • kachingasilwimba.com •
linkedin.com/in/kachinga-silwimba

Eligible to work in the United States without sponsorship

Professional Summary

Data Scientist with 4+ years of experience building end-to-end analytics and machine learning solutions on large, complex datasets. Strong background in Python, SQL, cloud, and HPC, with a track record of cutting data processing time by 50%, reducing computational cost by 25%, and improving model performance and reliability. Skilled in uncertainty-aware modeling, scalable data pipelines, and MLOps, and experienced in translating technical results into clear, actionable insights for cross-functional teams in climate, energy, and risk-focused settings.

Technical Skills

Programming & Data: Python, SQL, R, MATLAB, Bash, Scala, PySpark, Linux

Machine Learning & AI: Supervised and unsupervised learning; deep learning; generative models; large language models (LLMs); TensorFlow, PyTorch, Keras, scikit-learn; model calibration; uncertainty quantification

Data Engineering & Big Data: End-to-end data pipelines; ETL; data cleaning and wrangling; Hadoop; Spark; Xarray; Dask; GPU acceleration; HPC clusters; AWS; Google Cloud

MLOps & Software Engineering: Git/GitHub; Docker; Singularity; CI/CD pipelines; MLflow; experiment tracking; containerization; performance optimization; Slurm-based workflows

Analytics & Statistics: Statistical modeling; time-series and spatiotemporal analysis; experimental design and A/B testing; model validation; SHAP; Sobol and Fourier sensitivity analysis

Visualization & Communication: Matplotlib; Plotly; Tableau; geospatial visualization (GIS, Cartopy); dashboards; technical writing; presentations for technical and non-technical audiences

Domain & Scientific Computing: Earth and environmental data; climate and earth system modeling; NetCDF/Zarr; CLM5; WRF; CDO; NCO

Experience

Boise State University, School of Computing

Boise, ID

Graduate Research Data Scientist

Aug. 2021 – Present

- Build and productionize end-to-end Python data pipelines on GPU-accelerated HPC environments (Linux, Slurm), using Docker/Singularity and Git/MLflow to cut model-simulation data processing time by 50%.
- Apply Evidential Deep Learning for uncertainty quantification (UQ), improving prediction error by approximately 20% and documenting methods for reproducibility.
- Develop SQL and Python-based analytics and visualization workflows (e.g., Plotly/Tableau dashboards) to track key performance metrics and communicate findings to collaborators.
- Design and productionize unsupervised learning workflows for spatiotemporal pattern discovery, improving key validation metrics by ~30% and supporting better pattern detection and forecasting.
- Lead cross-functional collaborations and implement MLOps practices (Docker, Git, MLflow) to deliver versioned, containerized, and traceable experiments.
- Integrate domain knowledge into model architectures to enhance simulation fidelity and decision support.
- Publish peer-reviewed work and present to technical and non-technical audiences at international venues, translating complex results into actionable insights.

NSF National Center for Atmospheric Research (NCAR), CISL Visitor Program

Boulder, CO

CISL Visiting Scholar (CVP)

Sep. 2025 – May 2026

- Build scalable Python data pipelines on NCAR HPC systems to ingest, clean, and join CLM5 simulations, hydrologic observations, and basin attributes across hundreds of basins.

- Design, train, and validate evidential deep neural network models to emulate CLM5 streamflow ensembles with calibrated uncertainty for downstream decision-making.
 - Optimize GPU-accelerated and distributed training workflows (job scheduling, I/O, memory usage) to reduce runtime relative to full CLM5 ensemble simulations.
 - Implement evaluation and monitoring scripts to track model performance (e.g., NSE, RMSE, coverage) and compare emulators against baseline hydrologic simulations.
 - Collaborate with data scientists, software engineers, and domain experts to design and implement calibration workflows that tune EDNN emulators to CLM5 ensembles and observations, improving reliability for Earth system modeling and analytics.

- Designed and implemented adaptive learning workflows to optimize CLM5 model parameters, reducing computational cost by approximately 25%.
 - Built scalable data pipelines on HPC systems to orchestrate large-scale simulations and transform high-dimensional outputs into analysis-ready datasets.
 - Applied statistical modeling, experimental design, and validation metrics to evaluate model performance and robustness across parameter and scenario combinations.
 - Collaborated with cross-functional teams to integrate advanced statistical and evaluation routines into reusable analytics and model-assessment pipelines.

Education

Boise State University Boise, ID
Ph.D. in Computing, Data Science (GPA: 4.0) Feb. 2026
From Applied AI and machine learning for data analysis. Further, I implemented data mining and machine learning

Focus: Applied AI and machine learning for large-scale Earth and environmental data and risk modeling.
African Institute for Mathematical Sciences Kigali, Rwanda
M Sc. in Mathematical Sciences, Data Science Aug. 2020 – Jul. 2021

Copperbelt University Kitwe, Zambia
B.Sc. in Physics, Computational Physics (GPA: 3.8) May 2015 – Oct. 2019

Leadership & Activities

SIAM (Society for Industrial and Applied Mathematics) BSU Chapter Boise, ID
Vice President Apr. 2024 – May 2025

- Organize seminars and workshops featuring industry experts, connecting students with professionals.

- Manage a \$5,000+ annual budget, ensuring financial transparency and efficiency.

LEAP (Learning the Earth with Artificial Intelligence and Physics) New York, NY (Remote)
LEAP Tier 2 Member Jan. 2024 – Present

- Evaluate models against observational data, ensuring scientific integrity and improved prediction reliability.

Awards & Certifications

Awards:

- Bill Anderson Fund Fellow (2024), William Averette Anderson Fund, USA.
 - ASP Graduate Visitor Program Fellow (2024), NSF NCAR, Boulder, CO.

Certifications:

- IBM Machine Learning Certificate, IBM Digital-Nation (2020).
 - NASA Transform to Open Science (TOPS) Open Science Certificate (2024).
 - LEAP Momentum Bootcamp in Climate Data Science (2024).
 - Responsible Conduct of Research, CITI Program (2024).
 - CLM5 Point Simulations and WRF Tutorial Training, NCAR (2022).