

ZACHARIAH MILLER PHD

SENIOR DATA SCIENTIST

✉ zachariah.w.miller@gmail.com 🌐 <http://zwmiller.com> ☎ (270) 317-2618 📍 Chicago, IL
in <https://www.linkedin.com/in/zachariah-miller/> 🔄 zwmiller

SUMMARY

Senior data scientist focused on data-driven research and solutions. I solve interesting problems with data analysis, model building, machine learning, and software design. Passionate about using scientific techniques on "non-science" puzzles to extract answers for the real-world, understanding the world via data, and making models to improve decision making. Also passionate about writing software that doesn't suck.

SKILLS

PROGRAMMING: Python, SQL, BASH, UNIX, C++, LaTeX, Java, Spark
COMPUTING AND MODEL BUILDING: Machine Learning, AWS, Numerical Methods, Statistics, Distributed Computing, Big Data Architecture, Monte Carlo Simulations, Regression, Classification, Clustering, Database Management, MongoDB, PostgreSQL
DATA SCIENCE: SciKit Learn, TensorFlow, Deep Learning, A/B Testing, Hadoop, Graph/Network Analysis, Pandas, Hypothesis Testing, Bayesian Methodologies, Natural Language Processing, Experimental Design

EMPLOYMENT

Metis, Senior Data Scientist March 2017 - Current
Chicago, IL

- Instruction/Curriculum Design for 12-week, immersive data science/machine learning bootcamps.
- Curriculum: Python data science stack, data cleaning, supervised and unsupervised learning algorithms, natural language processing, deep learning, recommender systems, and how to apply data science techniques to solve business problems.
- Focus on end-to-end data science from data acquisition, data cleaning, database management, SQL, Python, modeling, deployment, and big data engineering.
- Designed and delivered machine learning/data analytics corporate trainings for multiple Fortune 500 companies.
- Business model optimization initiatives, focusing on data-driven insights, with Metis' in-house data.
- Work with marketing team to optimize customer segmentation and identify new markets.
- Managed projects for up to 15 students at a time, with each student working on an independent dataset and data pipeline (end-to-end process)

University of Illinois at Chicago, June 2015 - Feb. 2017
Postdoctoral Researcher - Data Analytics Specialist
Chicago, IL

- Large scale data analysis to extract physics results from petabytes of data, with C++ and Python.
- Big data techniques, including high performance cluster computing.
- Member of team that produces Monte Carlo simulations for the entire collaboration of 500+ people. Responsible for simulation design, production, and quality assurance. Co-supervisor of undergraduate research students (2016).
- Member of maintenance team for Intermediate Silicon Tracker detector (decommissioned 2016), including software management, hardware development, and hardware-software interface.

University of Kentucky, May 2010 - May 2015
Research Assistant - Data Analysis and Software Design
Lexington, KY

- Designed and built front-end GUI and back-end analyzer for a data acquisition system to interface with and record data from detectors using Java and C++. Implemented the database that stores the data.
- Wrote and maintained all analysis software for extracting high precision results from more than a million rows of data to measure neutron-induced fission rates.
- Research, design, development, and building of a prototype neutron detector for measuring particle interactions and the software for controlling the detector.

Eastern Kentucky University, Adjunct Professor Aug. 2011 - Dec. 2011
Richmond, KY

- Designed and taught Introductory Astronomy course for non-Physics Majors.

EDUCATION

University of Kentucky
PhD Nuclear Physics 2015
MS Nuclear Physics 2012

Eastern Kentucky University
BS Physics 2009
Cum Laude

PROJECTS

Curriculum Development for Data Science and ML March 2017 - Current
Developing slides, exercises, and notebooks on data science techniques. Topics range from cleaning data with Pandas and SQL to computer vision with deep learning. Developed more than 100 pieces of curriculum, including for cutting edge technologies like AWS, Spark at scale, network analysis, and many others. Also developed a deep understanding of both the theory and application of all modern machine learning algorithms in order to provide instructor notes and organize the content for maximum pedagogical use.

Machine Learning from Scratch with Python Sept. 2017 - Current
Programming many machine learning algorithms in a pedagogically useful way. Many libraries exist to use at a production level, but are necessarily opaque about how the models are built. In this library, I'm making all steps easily readable, while also making sure that the algorithms still converge to correct and useful models.
https://github.com/ZWMiller/machine_learning_from_scratch

Data Driven Marketing Strategy July 2017 - Current
Working with in-house data on our customer leads - attempting to segment our customer base into groups such that we can develop marketing strategies to address each group. Developed a pipeline to interact with our data warehouse, pull out customer records, clean the data, and cluster the various customers into sub-groups.

Particle Simulation Production and Quality Control at STAR June 2016 - Feb. 2017
Maintained and quality checked STAR's complex simulation framework. Produced simulations for the collaboration at-large and tested subsets of the simulated data. Simulation written in C++ and stored data at nanosecond time-step scale for thousands of particles, and also managed output files by writing data output across the computing cluster.

Analysis Tree Production & Measurement of Bottom Quarks at STAR Feb. 2016 - Feb. 2017
Software project to read, analyze, and compress terabytes of data into a smaller, more user-friendly format without losing useful information. Developed for use across entire analysis team. Analyzing multiple terabytes of physics data to extract information about the formation of the early universe with C++ and ROOT. Visualization of results, multiple speaking engagements on the topic, and detector maintenance and development with both software and hardware.

AWARDS

INVITED SPEAKER TO ASU-GSV ASU-GSV April 2018
ASU-GSV is the world's largest Education Tech Summit. I was invited as a speaker to inform the audience about the use of big data in Human Resources.

YOUNG RESEARCHER FELLOWSHIP Quark Matter 2017 Conference Jan. 2017
A monetary award to attend the largest high energy nuclear physics conference in the world.

OUTSTANDING TEACHING ASSISTANT AWARD University of Kentucky Arts and Sciences May 2011
Recognizes the best teaching assistant in the physics department.

VOLUNTEERING

Northwest Territory Alliance Officer/Board Member April 2015 - Current
Chicago, IL
The NWTA is a non-profit specializing in education about the American Revolutionary War and the associated era.