Zachary Jacokes

(615) 604-7498 | zjacokes@gmail.com | GitHub | Google Scholar

Executive Summary

Data scientist and systems architect specializing in extracting structure from complex, high-dimensional datasets and translating insights into scalable operational tools. Experienced designing automated pipelines, optimizing data workflows, and leading cross-functional projects across multi-site research environments. Known for turning ambiguous problems into measurable improvements through experimentation, analytics, and pragmatic engineering.

Education

Emory University, B.A. in Psychology *Fall 2009 – Spring 2013*
University of Virginia, Ph.D. in Data Science (in progress) *Fall 2021 – Spring 2026*

Doctoral Research

- Designed and implemented machine learning pipelines to analyze high-dimensional datasets across multi-site environments (500+ subjects), enabling reproducible analysis and cross-cohort validation.
- Developed embedding and dimensionality reduction optimization frameworks to identify reproducible latent structure across cohorts (publication under review).
- Implemented cross-validation and harmonization strategies (e.g., ComBat, nested CV) to ensure robust domain generalization across sites and scanners.
- Built automated and reproducible HPC data pipelines to process large datasets across distributed compute environments, reducing manual analysis time and enabling scalable experimentation.

Systems & Automation

- Designed automated pipelines for multi-site data ingestion, validation, and analysis
- Built reproducible containerized workflows in Docker and Singularity for distributed HPC environments
- Developed statistical validation frameworks to ensure robust model performance across heterogeneous datasets
- Implemented parallelized data processing pipelines reducing analysis time from days to hours

Technical Skills

- **Machine learning & data science:** traditional machine learning and statistical analysis, experimental design, deep learning (PyTorch/TensorFlow), neural network implementation from first principles (NumPy), time-series modeling and representation learning
- **Programming & tools:** Python, R, Bash, Slurm, NiLearn, fMRIPrep, AFNI
- **Data & workflow infrastructure:** Git/GitHub, reproducible pipelines, containerization (Docker/Singularity)

Work Experience

Senior Data Specialist, University of Virginia *Fall 2019 – Fall 2021*

- Developed, implemented and maintained scalable database systems for longitudinal, multi-site research projects, ensuring data integrity for over 500 participants
- Managed survey-based data acquisition efforts using secure transfer software
- Reduced manual error and increased validation efficiency by developing automated data processing and quality control pipelines
- Implemented data analysis workflows for parallel use on a high-performance computing cluster, enabling processing of petabyte-scale datasets by reducing computation time
- Contributed to HIPAA compliance documentation and managed data-access workflows
- Coordinated small project teams developing data engineering, quality control, and analysis workflows across multi-site research initiatives
- Published original research in data science and neuroscience

Programmer/Analyst I, University of Southern California *Fall 2015 – Fall 2019*

- Coordinated data collection and dissemination for multi-site studies
- Designed and executed reproducible statistical workflow, adopted as the lab standard
- Created data-driven MRI quality control protocol, adopted as the lab standard
- Gained experience with administration and scoring of neuropsychological tests
- Published multiple scientific papers, posters, and abstracts in neuroscience

Research Assistant, Georgia Institute of Technology *Summer 2013 – Spring 2014*

- Designed and implemented novel research experiments
- Learned to use basic coding programs and languages

Significant Publications

1. **Jacokes Z**, Beeler-Duden S, Lawson S, et al. Autism Sensory Profiles Predict Stimulus-Evoked Insula Connectivity. In: *MedRxiv* (preprint).
   — Topography-aware brain-behavior data integration and synthesis

2. **Jacokes Z**, Adoremos I, Hussain AR, et al. Unsupervised Dimensionality Reduction Techniques for the Assessment of ASD Biomarkers. In: *Biocomputing 2025*. WORLD SCIENTIFIC;2024:614-630.
   — Representation learning for identification of high-dimensional biomarkers

3. **Jacokes Z**, Jack A, Sullivan CAW, et al. Linear discriminant analysis of phenotypic data for classifying autism spectrum disorder by diagnosis and sex. *Front Neurosci*. 2022;16:1040085.
   — Applied ML framework for generalizable classification

4. Newman BT, **Jacokes Z**, Venkadesh S, et al. Conduction velocity, G-ratio, and extracellular water as microstructural characteristics of autism spectrum disorder. Bray S, ed. *PLoS*

ONE.2024;19(4):e0301964.
— Microstructural signal modeling; multimodal integration and interpretability

5. Gupta R, Audhkhasi K, **Jacokes Z**, Rozga A, Narayanan S. Modeling Multiple Time Series Annotations as Noisy Distortions of the Ground Truth: An Expectation-Maximization Approach. *IEEE Trans Affective Comput*. 2018;9(1):76-89.
— EM framework for noisy time-series ground truth inference