Zachary Jacokes

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

Executive Summary

Data scientist and systems architect specializing in experimentation, causal inference, and extracting actionable structure from large-scale, high-dimensional datasets. Expertise in designing robust experimental frameworks, developing and validating metrics under noisy and heterogeneous data conditions, and building reproducible analytical pipelines that generalize across populations and data sources. Experienced cross-functional collaborator with a track record of translating ambiguous research problems into measurable outcomes. Seeking to apply rigorous quantitative methodology to identity and commerce challenges at global scale.

Education

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

Doctoral Research

- **Designed and executed experimentation pipelines** for evaluating machine learning models on high-dimensional datasets across multi-site environments (500+ subjects), including power analysis, effect size estimation, and reproducible cross-cohort validation.
- **Developed metrics and validation frameworks** using nested cross-validation, domain harmonization, and statistical testing strategies to ensure model performance generalizes across heterogeneous data sources, akin to identity signal validation across diverse user populations.
- **Built representation learning frameworks** to discover reproducible latent structure linking individual-level signals across disparate datasets, informing how varied observations can be unified into coherent individual profiles.
- **Built scalable HPC pipelines** for large-scale data processing and model experimentation across distributed compute environments, reducing iteration cycles from days to hours.

Technical Skills

- **Experimentation and inference**: A/B testing methodology, experimental design, causal inference, (observational and experimental; collider bias mitigation), power analysis, multiple comparisons correction, regression, classification, dimensionality reduction
- **Machine learning and modeling**: supervised and unsupervised learning, deep learning (PyTorch, TensorFlow), representation learning, time-series modeling, model evaluation and validation
- **Programming and tools**: Python, R, SQL, Bash, Slurm, Git/GitHub
- **Data infrastructure**: reproducible ML pipelines, containerization (Docker/Singularity), HPC workflow orchestration, large-scale data management

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