

# Zachary Jacokes

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

## SUMMARY

---

Data scientist and systems engineer specializing in designing and operationalizing data pipelines for complex, multi-site, and regulated environments. Experience building end-to-end data infrastructure for clinical and neuroimaging datasets, including ingestion, validation, transformation, and large-scale analysis. Combines statistical rigor with pragmatic engineering to translate noisy, heterogeneous data into reliable, scalable systems.

## EDUCATION

---

**University of Virginia** — Ph.D. in Data Science (Expected 2026) Fall 2021 – Spring 2026

*Member of Inaugural School of Data Science Ph.D. Cohort*

**Emory University** — B.A. in Psychology Fall 2009 – Spring 2013

## EXPERIENCE

---

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

### ***Clinical Data Infrastructure (REDCap-Based)***

- Architected and deployed a multi-site clinical data platform using REDCap to support 20+ standardized behavioral instruments across 4+ research sites, ensuring data integrity for over 500 participants
- Designed database schema from scratch to align digital capture with paper-based clinical instruments while meeting NIH/NDA reporting requirements
- Implemented automated scoring pipelines including computation of summary metrics, T-scores, and sex-normed clinical scores across instruments
- Built extensive validation and constraint logic (range checks, type enforcement, PHI safeguards) to ensure data integrity at scale
- Developed branching logic and conditional workflows to dynamically adapt surveys based on participant responses
- Managed role-based access controls for clinicians, research assistants, and analysts across sites, supporting secure and compliant data workflows
- Led audit response efforts, adapting data validation and PHI handling practices to meet HIPAA compliance requirements
- Authored comprehensive documentation and trained clinical and research staff on system use and data standards

### ***Neuroimaging Data Pipeline (End-to-End Automation)***

- Designed and implemented an automated pipeline for ingestion, de-identification, and preprocessing of multi-modal imaging data (fMRI, DTI, structural MRI, EEG)
- Built scheduled ingestion workflows to detect new DICOM data and trigger processing pipelines on a rolling basis
- Automated de-identification of imaging data to ensure HIPAA-compliant handling of sensitive clinical information
- Standardized all outputs into BIDS format and integrated preprocessing using fMRIPrep within HPC environments (Slurm)
- Enabled scalable downstream analysis by structuring outputs for machine learning and statistical pipelines across petabyte-scale datasets

### ***Team Leadership & Mentorship***

- Coordinated small project teams developing data engineering, quality control, and analysis workflows across multi-site research initiatives
- Mentored two undergraduate researchers: guided development of imaging analysis pipelines, capstone research plans, and ML techniques (LDA, logistic regression, tree-based methods, genomic analyses)

**Programmer/Analyst**, University of Southern California

Fall 2015 – Fall 2019

- Coordinated multi-site neuroimaging data collection and distribution pipelines for the GENDAAR Research Consortium
- Designed reproducible statistical analysis workflows adopted as lab standards
- Developed MRI quality control protocols using factor analysis for cross-site data consistency, presented at OHBM 2017 and 2018
- Published multiple first-author and co-authored papers on neuroimaging, multi-site data challenges, and structural brain abnormalities

**Research Assistant**, Yerkes National Primate Research Center Summer 2014 – Summer 2015

- Developed Python-based data manipulation tools for experimental research with principal investigators
- Contributed to experimental design and methodology development for behavioral neuroscience studies

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

- Designed and implemented a novel research experiment in collaboration with computer scientists at the Child Study Lab
- Developed proficiency in coding and video annotation for behavioral data collection

**Research Assistant**, Marcus Autism Center — Early Intervention Program Spring 2013

- Conducted Discrete Trial Training with children with autism; supported functional communication and sign language instruction

## DOCTORAL RESEARCH

---

- Designed machine learning experimentation pipelines for high-dimensional, multi-site datasets (500+ subjects), enabling robust cross-cohort validation and reproducible evaluation of model generalization
- Developed representation learning and spectral embedding frameworks to identify stable latent structure across heterogeneous clinical and neuroimaging data (publication under review)
- Implemented harmonization and validation strategies (ComBat, nested cross-validation, power analysis, effect size estimation) to ensure generalization across sites, scanners, and populations
- Built scalable, automated HPC pipelines (Slurm, Docker/Singularity) for large-scale data processing and model experimentation, reducing iteration cycles from days to hours
- Developed interpretable models linking circuit-level neural activity to behavior through topography-aware brain–behavior data integration

## TECHNICAL SKILLS

---

- **Programming:** Python, R, SQL, Bash
- **ML & Statistics:** supervised/unsupervised learning, deep learning (PyTorch/TensorFlow), neural network implementation from first principles (NumPy), dimensionality reduction, representation learning, time-series modeling, experimental design, A/B testing methodology, causal inference
- **Data Systems:** REDCap, relational database design, schema design, data validation pipelines, large-scale data management
- **Infrastructure:** HPC (Slurm), containerization (Docker/Singularity), reproducible ML pipelines, AWS S3, Globus, Git/GitHub
- **Neuroimaging:** BIDS, fMRIPrep, AFNI, NiLearn

## SELECTED PUBLICATIONS

---

1. Jacokes Z, Beeler-Duden S, Lawson S, et al. Autism Sensory Profiles Predict Stimulus-Evoked Insula Connectivity. MedRxiv (preprint).
2. Jacokes Z, Adoremos I, Hussain AR, et al. Unsupervised Dimensionality Reduction Techniques for the Assessment of ASD Biomarkers. Biocomputing 2025. World Scientific; 2024:614–630.
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.

4. Ressa HJ, Newman BT, Jacokes Z, et al. Widespread associations between behavioral metrics and brain microstructure in ASD suggest age mediates subtypes. *Imaging Neuroscience*. 2025;3.
5. Newman BT, Jacokes Z, Venkadesh S, et al. Conduction velocity, G-ratio, and extracellular water as microstructural characteristics of ASD. *PLoS ONE*. 2024;19(4):e0301964.
6. Van Horn J, Jacokes Z, Newman B, Henry T. Is Now the Time for Foundational Theory of Brain Connectivity? *Neuroinformatics*. 2023:1–3.
7. Gupta R, Audhkhasi K, Jacokes Z, Rozga A, Narayanan S. Modeling Multiple Time Series Annotations as Noisy Distortions of the Ground Truth. *IEEE Trans Affective Comput*. 2018;9(1):76–89.
8. Irimia A, Lei X, Torgerson CM, Jacokes Z, et al. SVMs, MDS, and MRI reveal structural brain abnormalities associated with ASD and sex. *Front Comput Neurosci*. 2018;12:93.

*Full publication list: 14 journal articles, 1 book chapter, 15+ conference abstracts (OHBM 2016–2023)*

## **UNIVERSITY SERVICE & LEADERSHIP**

---

- University of Virginia Raven Society — First inductee from the School of Data Science; member of Selection Committee
- UVA Brain Institute — Consulted on funding allocation for neuroscience projects and strategic direction
- Neurodata Interest Group — Founding member; led biweekly discussions on seminal publications in neuroscience and data science