Zachary Jacokes, Ph.D.

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

SUMMARY

Neuroscientist, data scientist, and machine learning researcher with deep expertise in large-scale neural signal modeling, multi-site clinical data systems, and reproducible analytical pipelines. Experienced in turning complex, high-dimensional neural and clinical datasets into rigorous, decision-relevant evidence for scientific stakeholders. Background includes neurological biomarker discovery, patient phenotyping, brain-behavior modeling, data harmonization, and clinical research infrastructure across federally funded multi-site studies. Strong publication record in neurodevelopmental disorders, with practical experience collaborating across clinical, engineering, and research teams to ensure data curation, analysis, and interpretation hold up under scientific scrutiny.

EDUCATION

University of Virginia: Ph.D. in Data Science — Fall 2021 – Spring 2026
Member of Inaugural School of Data Science Ph.D. Cohort

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

DOCTORAL RESEARCH

- Designed and executed machine learning pipelines for high-dimensional, multi-site neuroimaging and EEG datasets (500+ subjects across multiple acquisition environments), enabling reproducible cross-cohort analysis and robust biomarker evaluation.
- Developed spectral embedding and representation learning frameworks to identify stable latent neural structure across heterogeneous populations (directly applicable to biomarker discovery and patient phenotyping in clinical EEG settings; publication under review).
- Applied harmonization methods (Neuro ComBat, nested cross-validation, power analysis, effect size estimation) to correct for site and scanner effects across cohorts, ensuring validity of derived neural endpoints across populations.
- Built interpretable models linking connectivity and phenotypic patterns to behavioral and clinical outcomes through topography-aware brain-behavior integration (publication under review; available on medRxiv).
- Operationalized scalable HPC pipelines (Slurm, Docker/Singularity) for automated ingestion, de-identification, preprocessing (fMRIPrep, AFNI), and BIDS-structured output, reducing iteration cycles from days to hours.
- Documented dataset provenance, idiosyncrasies, and quality characteristics across heterogeneous multi-site acquisitions; built validation frameworks to ensure fitness-for-use prior to downstream analysis.

EXPERIENCE

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

Clinical EEG & Neuroimaging Data Infrastructure

- Architected and deployed a multi-site clinical data platform (REDCap) supporting 500+ participants, 5+ research sites, and 30+ standardized behavioral instruments, which functioned as the longitudinal data backbone for a federally funded neurodevelopmental study.
- Designed automated scoring pipelines computing summary metrics, T-scores, and sex-normed clinical scales across instruments, enabling endpoint derivation for downstream statistical and ML analysis.
- Implemented validation and constraint logic (range checks, type enforcement, PHI safeguards) and led audit response efforts adapting practices to meet HIPAA and NIMH/NDA compliance requirements.
- Designed and implemented end-to-end automated pipeline for ingestion, de-identification, and preprocessing of multi-modal data (fMRI, DTI, structural MRI, EEG) with BIDS-structured outputs for downstream scientific use.
- Built parallelized HPC workflows enabling petabyte-scale dataset processing with reproducible, version-controlled outputs.

Scientific Communication & Team Leadership

- Authored comprehensive data documentation and trained clinical and research staff across sites on system use, data standards, and quality requirements.
- Coordinated cross-functional project teams spanning data engineering, quality control, and scientific analysis across multi-site research initiatives.
- Mentored two undergraduate researchers in ML techniques, imaging pipeline development, and research design; both contributed to published and conference work.

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

- Coordinated multi-site neuroimaging data collection, harmonization, and distribution for the GENDAAR Research Consortium: a multi-institution longitudinal neurodevelopmental study requiring standardized data governance across sites.
- Developed MRI quality control protocol using factor analysis for cross-site data consistency; presented at OHBM 2017 and 2018 and adopted as lab standard.
- Designed reproducible statistical analysis workflows in Python, R, and SPSS adopted as lab standards; published multiple first- and co-authored papers on neuroimaging, multi-site data challenges, and structural brain abnormalities.
- Gained extensive experience presenting complex analytical results to interdisciplinary scientific audiences including clinicians, engineers, and academic collaborators.

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

- Developed Python-based data manipulation and analysis tools; contributed to experimental design and behavioral neuroscience methodology.

TECHNICAL SKILLS

Neuroscience & Signal Analysis: fMRI/DTI preprocessing, spectral embedding, connectivity modeling, EEG analysis, BIDS, fMRIPrep, AFNI, NiLearn, multi-modal data integration

Statistics & ML: Supervised and unsupervised learning, dimensionality reduction, representation learning, deep learning (PyTorch/TensorFlow), time-series modeling, causal inference, harmonization (Neuro ComBat), nested cross-validation, power analysis

Programming & Tools: Python, R, Bash; NiLearn, fMRIPrep, AFNI; version-controlled collaborative codebases (Git/GitHub); reusable, documented, tested scientific code

Data Systems & Infrastructure: REDCap, relational database design, ETL pipelines, large-scale data validation, HPC (Slurm), containerization (Docker/Singularity), AWS S3, Globus

Scientific Communication: Peer-reviewed publications, conference presentations (OHBM 2016–2023), cross-functional stakeholder communication, statistical analysis plans

SELECTED PUBLICATIONS

1. **Jacokes Z**, Beeler-Duden S, Lawson S, et al. Autism Sensory Profiles Predict Stimulus-Evoked Insula Connectivity. MedRxiv (preprint). — *Topology-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. Biocomputing 2025. World Scientific; 2024:614–630. — *Representation learning for biomarker identification in high-dimensional clinical data.*
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 for patient classification and cross-cohort generalization.*
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. — *Age-stratified subgroup analysis; phenotype-biomarker association.*
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. — *Multimodal neural biomarker characterization and interpretability.*

6. 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. — *EM framework for noisy ground truth inference in longitudinal time-series data.*

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 neuroscience funding allocation and strategic direction
- Neurodata Interest Group — Founding member; led biweekly discussions on seminal publications in neuroscience and data science