

JACK GRAY

Clinical Research Solutions Specialist

✉ contact@jackgray.nyc
☎ 929-409-5452

🌐 linkedin.com/in/johnhgrayiii
🔗 github.com/jackgray

🌐 jackgray.nyc
📷 @latent.to.life

📍 Brooklyn, NY



EXPERIENCE

Lead Data Solution Engineer

Center for Neuropsychomaging of Addiction and Related Conditions (NARC) | Dr. Rita Goldstein

📅 April 2022 – Ongoing

📍 Icahn School of Medicine at Mount Sinai

Fulfill all technological needs for several multi-\$M clinical trials to investigate efficacy of novel pharmacologic, cognitive, and digital therapeutic treatments to opiate and cocaine use disorders

Designed & implemented collection of **self-managed IaC solutions** to manage & process clinical data in a metal/cloud hybrid environment

DISTRIBUTED COMPUTING + NODE PROVISIONING | Flatcar Linux/systemd + Terraform + Make + Equinox Metal + Azure + Nomad

- Code to auto-generates new role-specific cluster node images from templates and TF var inputs
- Pre-loads systemd units, network settings, and download required binaries for instant cluster-joins
- Evaluated alongside K8's and selected for minimal maintenance overhead and system resource efficiency

DISTRIBUTED OBJECT STORE | MinIO S3 + Bare-metal + Docker

- Central storage backbone for data collection endpoints and ETL processes
- Event streaming, erasure-encoding, retention compliance, IAM, versioning, encryption, VPN protected

BIG DATA SEARCH/VIZ | Elastic stack: Filebeat + Metricbeat + Logstash + Elasticsearch + Kibana + Grok + regex + Python + Docker

- Dynamically scans and indexes databases and text files from disparate configurable sources
- Offers no-code full-text & SQL **queries** on millions of unstructured documents with <15ms latency
- Rapidly create visualization dashboards with drag & drop or Vega syntax to monitor KPI's

DATA PIPELINE MANAGEMENT | Kestra (Airflow alternative)

- Automated processing pipelines for MRI, audio/NLP, genetic, and psychometric data with pandas, ffmpeg, WhisperAI, matplotlib, bash
- Evaluated alongside Airflow and selected for its container-preferred architecture, S3 integration, and low overhead

DB / ETL SOLUTIONS | Python/Pandas/Elasticsearch Client + Bash + S3 API + Webhooks + ArangoDB + PostgreSQL + REST APIs

- Extracted clinical assessment data from survey distribution platform, modeled and loaded into graph model in ArangoDB collections
- Wrote DAGs to automate stages of MRI/behavioral analysis including data downloads, conversion, relabeling, pre-processing, temporal alignment, validation, and subsequent analysis processing stages

WEB/APP DEVELOPMENT | React + Express + NGINX + Azure + Cloudflare + SSL/TLS

- Audio Recording App – records and transcribes audio diary entries for NLP analysis from mobile web browser

BROWSER-BASED VIRTUAL MACHINES + CUSTOM OS | Terraform + Coder workspaces + Apache Guacamole

- Users spin up their own ephemeral VMs pre-configured with analysis tools and S3 mounts per IAM policies
- Scalable – lifted compute resource limitations

All solutions were thoroughly evaluated & tested alongside popular industry standards, vetted for specific needs, scale, cost, and maintenance overhead, and configured to run in self-managed/free **Docker** instances with zero recurring costs while allowing seamless expansion to the cloud or paid/managed instances

MRI Technical Specialist / Data Pipeline Developer

Columbia University Department of Psychiatry | MRI Center

📅 June 2021 – April 2022

📍 New York State Psychiatric Institute

Developed ETL solutions for multi-modal imaging data between imaging console, XNAT database, and compute cluster serving department of MRI researchers

- Coil-noise detection pipeline sends **Slack notification** when early signs of a specific type of image degradation are detected, informing MRI compute team immediately & affording a chance to fix the problem gracefully in scanner off-hours preventing massive data loss
- Pipeline to pull, label, validate, and pre-process selected neuroimaging data in a multi-user environment with built-in auth service
- Automated MR Spectroscopy conversion and network transfer of proprietary GE imaging formats

MRI Research Tech

Columbia University Department of Psychiatry | MRI Center

📅 January 2016 – June 2021

📍 New York State Psychiatric Institute

Fervently studied web development, React, Next, GraphQL, AWS, distributed system design, progressive web-app principles, hosting, TLS/SSL, through a personal project, *GovTracker*, a one-way social network to track politician activity and events scraped from multiple sources

(Also managed equipment and data collection sessions for department, consulted researchers technically, debugged psychometric code, handled imaging data transfers)

EDUCATION

B.E. in Electrical & Computer Engineering

Minors in Neuroscience + Music

📅 2009 – 2014

📍 Auburn University