12 Behavioral Data Analysis Platform: Building Research Infrastructure for Animal Models of OCD (SE/CS)

Supervisor(s): Dr. Henry Szechtman and Dr. Anna Dvorkin-Gheva
Organization: Dept. of Psychiatry and Behavioural Neurosciences, and Dept. of Pathology and Molecular Medicine, McMaster
E-mail(s): szechtma@mcmaster.ca, dvorkin@mcmaster.ca
Selected: Available

Description

Project Overview

This project focuses on creating a comprehensive data analysis platform for one of the most extensive behavioral neuroscience datasets—nearly 20,000 trials of rat behavioral data from studies of Obsessive-Compulsive Disorder (OCD). The dataset contains rich spatial-temporal tracking data (x, y, t coordinates), corresponding video recordings of rat behavior, and additional research files that researchers may use to study psychiatric disorders.

The Challenge

While this valuable dataset exists in a public repository, there is currently no user-friendly way for researchers to search, filter, download, or analyze specific subsets of data. Moreover, researchers cannot easily view behavioral trajectories alongside their corresponding videos, a feature that would add to the dataset's scientific utility.

Project Goals

Your team will build a modern web-based platform that makes this behavioral dataset truly accessible to the global research community:

1. **Database Design & Backend Development**
    - Design a robust PostgreSQL database schema for behavioral data, metadata, video files, and additional research files
    - Implement REST API endpoints for data queries and downloads
    - Create efficient indexing for large-scale spatial-temporal data and file management system for video content
2. **Interactive Web Interface**
    - Build a React-based frontend with intuitive search and filtering capabilities
    - Implement natural language querying—allow researchers to ask questions like "show me trials with strong checking behavior after 5 injections" or "find sessions where rats showed compulsive patterns"
    - Create data visualization tools for trajectory plotting and behavioral metrics
    - Implement synchronized video playback—allow researchers to view behavioral trajectories alongside corresponding video recordings
    - Design user-friendly interfaces for researchers without programming expertise
3. **Data Processing Pipeline**
    - Develop Python tools for processing coordinate data into meaningful behavioral measures
    - Implement algorithms for detecting key behavioral patterns (homebase behavior, checking routes, exploration metrics)
    - Create automated analysis workflows for common research tasks
    - Design extensible architecture for future video analysis capabilities

Technical Stack

- **Backend:** Python (FastAPI), PostgreSQL, Redis caching
- **Frontend:** React, modern JavaScript, data visualization libraries
- **Data Processing:** Python (NumPy, Pandas, scientific computing libraries)
- **NLP Integration:** Language model APIs for natural language querying
- **Media Handling:** Video streaming, file management systems

Why This Project Matters

- **Real Impact:** Your work will be used by OCD researchers worldwide, potentially accelerating discoveries that help millions of people
- **Open Science:** You'll contribute to making scientific data more accessible and reusable
- **Technical Challenge:** Handle big data problems with spatial-temporal complexity, large video files, and AI-assisted querying
- **Professional Experience:** Build a production system that serves an international research community

Learning Outcomes

- Full-stack web development with modern technologies
- Database design for scientific data and multimedia content management
- API development and optimization for large file handling
- Data visualization, video integration, and user experience design
- Working with real scientific datasets and research workflows

Deliverables

- Functional web platform deployed and accessible online
- Natural language query system for intuitive data exploration
- Synchronized trajectory-video viewing capabilities
- Python analysis tools packaged as open-source libraries
- Extensible file management system for diverse research data
- User documentation and tutorials for researchers

Team Structure (Suggested)

- Backend/Database specialist
- Frontend/UI developer
- Data processing/algorithms developer
- NLP/AI integration specialist
- Media/video systems developer
- DevOps/deployment specialist

Links

- GigaScience Article
- FRDR Szechtman Lab Repository