Timothy **Sheehan**

Data Science | Research (292 Citations; H-Index 7)

timothysheehanc@gmail.com | TimCSheehan | timothy-sheehan | Timothy C. Sheehan

Skills

analyses	ML: supervised {regression, classification, XGBoost}, unsupervised {k-means, VAE, PCA}, and contrastive learning. Hypothesis testing: (parametric and non-parametric), Bayesian inference, Fourier analysis, autoregression
packages	matplotlib, pandas, numpy, PyTorch (lightning), transformers (huggingface), seaborn, scikit-learn, OpenCV
languages	python, R, SQL, JavaScript/Typescript/HTML
tools	Docker, AWS EC2/S3, LakeFS, git, Google Cloud functions/scheduler, postgreSQL, LangChain
non-professional	figure drawing, long-distance paddleboarding, alpine skiing, open-water swimming

Experience

Umpire Bias Web App — *Remote*

Machine Learning | Web Development — *April 2024 - Present*

- Built interactive website for visualizing strike zones biases of MLB umpires *(umpirebias.com)* implementing feedback from web-developers and sports analysts. Queried data in **postgreSQL** database utilizing client and server side actions (**React**).
- Scheduled (**Google Cloud Scheduler**) continuous training and deployment of boosting (**XGBoost**) models to predict pitcher performance (#K/#BB) via containerized **Cloud Function**. Utilized **Bayesian Optimization** to chose meta-parameters (*repo*).
- Deployed riddle telling chatbot using **LangChain**, prompt engineering, and ChatGPT API calls (*chatApp*).

Cumulus Neuroscience — *Boston, MA*

Data Scientist — *May 2023 - April 2024 (loss of funding)*

- Scientific lead on clinical trial partnership with large pharma company (>$1M contract). Collaborated with members of analytics and science team to deliver timely reports to customers on data quality and target effect sizes.
- Quantified drug response using **mixed-effects models**, increased reliability (intraclass correlation, **ICC**) of key metric by >20%.
- Database management for clinical data **(AWS S3, lakeFS)**. Versioning and containerization of production code **(git, Docker)**.
- Created **ETL pipeline** to merge multi-modal data streams (EEG, eye-tracking, genetic markers) into existing structured database.
- Trained variational autoencoder (**VAE**) in **PyTorch** to harmonize dimensionality of EEG datasets from disparate sources.

Perceptive Automata — *Boston, MA*

Data Science Intern — *Summer, 2022*

- Designed experiments to measure human judgements of pedestrian intent for human in the loop (**HITL**) ML models (autonomous driving risk). Improved prompts and interface using **jsPsych** and wrote data to **MySQL** database.
- Deployed **image segmentation** & **pose estimation** computer vision algorithms to label pedestrian interactions (**OpenCV**).

UC San Diego / Salk Institute — *La Jolla, CA*

Researcher | PhD Student — *September 2017 - May 2023*

- Executed cross-institute research in sensory neuroscience resulting in four publications. Designed experiments, mentored research assistants, collected behavioral and neuroimaging data, and designed computational models to link brain and behavior.
- Implemented neural architecture for smell detection to improve performance of O(1) memory **similarity search algorithm** (Bloom filter) on both natural (odor response) and benchmark (eg. MNIST) data by **55% over existing SOA** model (*PNAS*).
- Designed experiment to examine neural representations (fMRI) of sensory context and compare with behavior. Built **generative model** to simulate observers estimated parameters by **maximum likelihood estimation** (*git repo*, *PLOS Biology*, *commentary*).
- **Fine-tuned Image to Text transformer** on clinical pose data to aid in diagnosis of bipolar disorder. Increased **CIDEr** score from 0.06 (zero-shot) to 0.84 on validation set (*Cosyne 2023b*).

National Institute of Neurological Disorders and Stroke (NINDS) — *Bethesda, MD*

Clinical Researcher | Post-bac — *September 2015 – May 2017*

- Coordinated with clinical teams to collect data from Epilepsy patients. Managed synchronization, transfer, and storage (**ETL**) of ECoG data from recording systems to research database. Data science lead on **multi-center study** to improve working memory.
- Discovered feature (sample entropy) that predicted memory performance across individuals (r=0.51). (*jNeuroscience*).
- Optimized classifier through **feature engineering** & **hyper-parameter tuning** to predict semantic memory. Increased cross-validated **AUC-ROC** from 0.63→ 0.68 and **deployed model in real-time** neural feedback experiments (*NatureCommunications*).
- Built GUI used by neurologists to control neural stimulator for experimental and clinical interventions. Improved clinical workflow saving **10+ minutes/clinical encounter** and increased data quality and machine readability (*BrainStim*)

Education

UC San Diego — *La Jolla, CA*

Ph.D. in Neurosciences, computational specialization — *May, 2023*