Aaron Broukhim

Email: aabroukh@ucsd.edu **Mobile:** (310) 867-3564
Website: https://aabroukh.github.io **LinkedIn:** linkedin.com/in/aabroukh

Education

Sep 2022 to June 2026

University of California San Diego
Ph.D. in Computer Science & Engineering
- Grad-level Coursework: Deep Learning, Unsupervised Learning, Deep Generative Models, Probabilistic Reasoning, Learning Algorithms, Recommender Systems, & AI Agents
- Research Interests: Preference-Based Learning, Conversational Audio Models, Social Signal Processing, Autonomous Vehicles, Digital Mental Health Interventions
- Thesis Topic: Preference-Based Learning for Coversational Audio Models
- Advisor: Nadir Weibel

Sep 2019 to June 2021

University of California San Diego
B.S. in Cognitive Science: Machine Learning & Neural Computation
Minor in Computer Science & Engineering
- Undergrad-level Coursework: Supervised, Unsupervised, Reinforcement, Deep Learning, Genetic Algorithms, ML for Brain Computer Interfaces, & Deep Learning for Music

Research Experience

Sep 2022 to June 2026

University of California San Diego
Graduate Student Researcher
- Optimized a Mixture of Experts (MoE) model, experimenting with weight averaging to enhance performance while reducing computational costs.
- Developed a Retrieval-Augmented Generation (RAG) pipeline to improve a large language model's understanding of Dialectical Behavior Therapy (DBT) for real-time mental health interventions.
- Investigated misinformation perception, identifying user traits correlated with distrust in misinformation tags through survey-based studies.
- Developed an intuitive web tool to analyze biases in AI models, improving accessibility in content moderation for non-technical users.
- Researched community-based labeling for content moderation, comparing its effectiveness against traditional top-down approaches.
- Analyzed user trust in autonomous vehicle explanations, evaluating the impact of accurate and inaccurate AI-driven explanations on user confidence.

Core Competencies: Machine Learning, Large Language Models (LLMs), Mixture of Experts (MoE), Retrieval-Augmented Generation (RAG), Model Optimization, Content Moderation AI, Trust in AI, Human-Centered AI, AI for Mental Health, Quantitative Research

July 2021 to Aug 2022

University of California San Diego
Undergraduate Student Researcher
- Web Scraped Facebook using Selenium to make inferences on a user's connections
- Designed logistic regression models capable of detecting hate speech on Twitter & used word embeddings to bin dataset into different types of hate speech
- Designed and implemented a UI in React with a Node and Flask backend to help non-tech savvy users identify faulty ML systems
- Created SQL databases on an AWS server and conducted queries to support UI backend
- Built Balltree with various similarity metrics to show similar tweets that may be mislabeled

Core Competencies: XAI, Natural Language Processing (NLP), Web Scraping, Hate Speech Detection, Word Embeddings, Full-Stack Development, Database Management

Other Professional Experience

Jan 2022 to Aug 2022

Virufy
Associate Audio Machine Learning Engineer
• Peer review ML research involving cough audio classification in Tensorflow and SK-Learn
• Finetuned existing models via transfer learning to conduct audio classification
• Utilize Sagemaker to explore new feature and architecture combinations with novel data
Core Competencies: Audio Feature Analysis, PyTorch, Peer Review, Sagemaker

Selected Publications (Google Scholar)

Accepted or Published

2. Kaufman et al., WARNING This Contains Misinformation: The Effect of Cognitive Factors, Beliefs, and Personality on Misinformation Warning Tag Attitudes, **CSCW 2025**
1. Kaufman et al., What Did My Car Say? Impact of Autonomous Vehicle Explanation Errors and Driving Context On Comfort, Reliance, Satisfaction, and Driving Confidence, **CHI 2025**

In-prep or Review

2. Broukhim et al., How Preference Data Differs in Text Compared to Audio, NeurIPS 2025
1. Chien et al., Community Content Moderation, CSCW 2025

Projects

Selected Projects

Vehicle Motion Forecasting
• Explored various model architectures (MLP, LSTM, CNN, Transformer) to predict motion of a car
• Selected and normalized relevant features (position, velocity, lanes) from the Argoverse dataset

Snake Reinforcement Learning
• Utilized N-Step Temporal Difference and SARSA methods to play snake and compared performance
• Designed a custom Open-AI gym environment and deep Q-Learning agent in keras that reached level 20 consistently (max 40) with a small state space of 10 and action space of 3

Generalizable Deep Speech Anonymization
• Leveraged Generative Adversarial Networks for Denoising Diffusion Implicit Models to learn a distribution over language agnostic speaker embeddings
• Used a modified loss to bias the generative model towards an anonymized distribution

GLOBEM Supervised and Exploratory Analysis
• Use generalized linear model to select features from GLOBEM dataset for PHQ-4 and PANAS prediction
• Conduct gridsearch with random forrest classifier and k-fold validation
• Analysis of best model's features using SHAP values

Miscellaneous

Volunteer

DISCO Lead
UC San Diego Visit Day Lead
Founder of the Graduate Student Volleyball Party (GSVP)
3x GRADWIC Mentor

Reviewer | CHI '25

Awards | NLM T15 Training Grant Fellowship

Tech Skills

Programming Languages: Python, SQL, Java, CSS, HTML, C++, C, Javascript
Frameworks: Pandas, NumPy, SK-Learn, TensorFlow, Keras, PyTorch, Selenium, React, Node, Flask
Misc: Git, HuggingFace, Sagemaker, MySQL, AWS, Illustrator, Lightroom, Photography, Firebase

Languages | English, Spanish, Farsi, Hebrew