

Work Experience

Senior Clinical Data Scientist	Aiberry	May 2023–Present
<ul style="list-style-type: none">• Unlocked ~30% more top-of-funnel ARR by fine-tuning custom LLMs for extracting clinical insights from screenings• Halved our largest compute cost while sustaining accuracy by overhauling our feature engineering process in Python• Catalyzed >\$3M in new VC funding leads by driving the publication of the company's first clinical validation study• Delivered highly requested features by leading R&D into LLMs and speeding up inference on AWS Lambda >700%• Increased available screenings 500% by owning ML engineering from raw data to deploying models in production• Generated >10 warm sales leads by giving talks about AI and LLMs at prominent institutions like AWS		
Senior Data Scientist	The Looma Project	2022–2023
<ul style="list-style-type: none">• Enabled real-time analytics worth >\$1M in full-funnel value by putting a LightGBM ML model in production• Created a new revenue channel for a mission-critical product by implementing computer vision models in Python• Reduced latency in reporting performance metrics during pilots 400% by creating automated, scalable reports• Sped up a customer-facing data API 600% by optimizing SQL calls to BigQuery• Engaged 50% of employees on data science case study results by developing an interactive, participatory system		
Principal Data Scientist	Lab for Scalable Mental Health	2020–2022
<ul style="list-style-type: none">• Decreased user depression 19% by architecting A/B tests in a linear regression framework with multiple imputation• Saved 20 hours of manual work per week by creating a suicidality screener for text data using boosted tree models• Engineered end-to-end pipeline for 100% of the organization's data with a mandate to only use open-source tools		
Data Scientist	Lab for Scalable Mental Health	2019–2020
<ul style="list-style-type: none">• Reduced churn 53% in digital health products by A/B testing the tradeoffs between effectiveness and churn• Achieved >90% sign-up rate for reproducible workflow tools after presentations at national conferences		

Technologies and Languages

- Languages: Python (numpy, pandas, scikit-learn, PyTorch), R (tidyverse, ggplot2), Javascript, Rust
- Technologies: SQL, GCP, BigQuery, AWS, S3, Lambda, Sagemaker, Git, GitHub, Docker, Command line
- Analytic Techniques: Causal inference (DAGs, PSM, Double ML), A/B testing, Multilevel models, Machine learning

Education and Certifications

- **Ph.D. Clinical Psychology**, University of Texas at Austin, Austin, TX. **2014–2020**
- **M.A. Clinical Psychology**, American University, Washington, DC. **2011–2013**
- **B.A. Psychology**, American University, Washington, DC. **2008–2011**

Projects

- [Podcast Episode Quality Predictor](#) - A production model that predicts past week episode quality, 0.80 F1 Score
- [mlcheck Command Line Tool](#) - Rust-based command line tool that checks Python and R scripts for ML best practices
- [Data Engineering and Validation](#) - Open-source textbook used in several organizations' [training materials](#)

Other Information

- Published >25 technical papers, [cited > 3,700 times](#)
- Co-wrote the [#1 new release](#) in the teen mental health category on Amazon (later translated to Mandarin)
- Researched and fact-checked an episode of [If Books Could Kill](#), a top podcast according to [Vulture](#)