Saad Mohammad Abrar

+1 (205) 240-1972 | Email: saad6522@gmail.com | LinkedIn | GitHub | Website | Google Scholar | Santa Barbara, CA

Machine learning research scientist with a Ph.D. in Computer Science and 6+ years of experience evaluating forecasting, geospatial, time-series, and foundation-model systems on large-scale real-world data. Experienced in designing rigorous model comparisons, statistical validation frameworks, robustness and failure analyses, and reproducible evaluation pipelines over TB-scale scientific, mobility, and geospatial datasets. Strong background in Python, PyTorch, statistical modeling, and spatiotemporal ML, with 20+ peer-reviewed publications. Research focuses on determining where models work, where they fail, and whether observed improvements are scientifically meaningful.

Education

University of Maryland, College Park — Aug 2020 – May 2026
Ph.D. in Computer Science — CGPA: 3.90/4.00 — *College Park, MD*

- **Thesis:** *Data-Driven Decision-Making for Societal Systems: Auditing Mobility, Contextual Data, and Models for High-Stakes Applications*
- **Relevant Coursework:** Causal Inference and Evaluation Methods, Big Data Infrastructure, Advanced Numerical Optimization, Foundations of Deep Learning, Automated Learning and Data Analysis.

University of Maryland, College Park — Aug 2020 – May 2023
M.S. in Computer Science — CGPA: 3.96/4.00 — *College Park, MD*

Experience

Postdoctoral Researcher — July 2026 – Present
University of California, Santa Barbara — *Santa Barbara, CA*

- Investigating how **urban physical barriers** influence social integration using large-scale **human mobility and geospatial data**, combining spatial analysis, causal inference, neighborhood accessibility measures, and cross-community mobility modeling.
- Developing an end-to-end geospatial document-analysis pipeline by fine-tuning compact **Qwen-VL models** and combining **YOLO-OBB + TrOCR** for detection, localization, and transcription of handwritten labels from historical 1930s FHA maps.

Graduate Research Assistant — Aug 2020 – Jun 2026
Urban Computing Lab, University of Maryland — *College Park, MD*

- Led evaluation-driven machine learning research from **problem formulation and experimental design** through data engineering, model development, statistical validation, failure analysis, and communication of results across large-scale geospatial and behavioral datasets.
- Developed and evaluated **spatiotemporal forecasting models** using STGNNs, STGCNs, GRUs, Temporal Fusion Transformers, gradient-boosted models, and statistical baselines, analyzing predictive performance across heterogeneous geographic and temporal settings.
- Designed rigorous **model-evaluation frameworks** covering predictive accuracy, robustness, demographic bias, statistical reliability, and performance trade-offs; investigated when strong aggregate metrics masked systematic failure modes across populations and places.
- Built reproducible **Python-based evaluation and benchmarking pipelines** using PyTorch, Hugging Face, vLLM, AWS Bedrock, and GCP, supporting large-scale experimentation across **21 vision-language models and 8 open-weight LLMs**.
- Developed statistical evaluation workflows using **hypothesis testing, multiple-comparison correction, regression, monotonic-trend analysis, ranking metrics, and subgroup analyses** to distinguish robust model behavior from benchmark artifacts and noisy improvements.
- Built scalable **data engineering and feature pipelines** integrating TB-scale mobility traces, POI metadata, Census demographics, social-media signals, public-health records, and ground-truth financial transaction data for large-scale geospatial modeling and model validation.
- Developed automated **prompt optimization and high-throughput inference workflows** using DSPy/GEPA to accelerate recurring model evaluations and systematically characterize contamination, reasoning failures, and sensitivity to experimental design.
- Developing an **LLM-based population simulation framework** that combines structured contextual data, retrieval-augmented in-context learning, and agent-based generation with evaluation workflows for measuring how closely simulated mobility patterns reproduce observed behavior.

Research Assistant — Jan 2020 – Jul 2020
Database Group, North Carolina State University — *Raleigh, NC*

- Developed and rigorously evaluated graph representation-learning and candidate-ranking models for drug repurposing using negative sampling, margin-based ranking loss, five-fold cross-validation, MRR, and Hits@K; achieved **0.80 MRR** versus an approximately 0.70 DTINet baseline.

Selected Publications *Full list available at Google Scholar.*

S.M. Abrar, N. Awasthi, D. Smolyak, N. Sigalo, and V. Frias-Martinez. *Auditing the Fairness of the US COVID-19 Forecast Hub's Case Prediction Models.* **PLOS ONE, 2025.**

S.M. Abrar, N. Awasthi, D. Smolyak, and V. Frias-Martinez. *Analysis of Performance Improvements and Bias Associated with the Use of Human Mobility Data in COVID-19 Case Prediction Models.* **ACM Journal on Computing and Sustainable Societies, 2023.**

N. Sigalo, N. Awasthi, **S.M. Abrar**, and V. Frias-Martinez. *Using COVID-19 Vaccine Attitudes on Twitter to Improve Vaccine Uptake Forecast Models in the United States.* **JMIR Infodemiology, 2023.**

K. T. Zinat*, **S.M. Abrar***, S. Saha, S. Duppala, S. N. Sakhamuri, and Z. Liu. *ProcVQA: Benchmarking the Effects of Structural Properties in Mined Process Visualizations on Vision–Language Model Performance.* Findings of **EMNLP**, 2025; **CVPR** VizWiz Workshop. **Equal contribution*

S.M. Abrar, E. Kurella, A. Dadarya, N. Awasthi, K. T. Zinat, and V. Frias-Martinez. *Do LLMs Know Your Neighborhood? Evaluating LLM Priors for Neighborhood-Level Mobility Prediction and Structural Alignment.* Under review

N. Awasthi*, **S.M. Abrar***, D. Bardhoshi*, B. Buddi*, E. Kurella*, and V. Frias-Martinez. *The Coordinates Were Always There: Detecting and Quantifying Pretraining Contamination from Geospatial Data in LLMs.* **SIGSPATIAL '26**. **Equal contribution*

Selected Projects

U.S. COVID-19 Forecast Model Audit | *Forecast Evaluation, Statistical Analysis, Geospatial Modeling*

- Audited case-prediction models from the **U.S. COVID-19 Forecast Hub** to determine whether aggregate predictive performance masked systematic differences across populations and geographic contexts; published in **PLOS ONE** (2025).
- Developed reproducible model-comparison and subgroup-evaluation analyses focused on **predictive reliability and heterogeneous failure modes**, demonstrating why headline accuracy alone can provide an incomplete picture of forecast quality.

Geospatial LLM Contamination Audit | *vLLM, Hugging Face, DSPy/GEPA, Statistical Testing*

- Built an end-to-end evaluation framework using matched open and closed geospatial datasets across trajectory, POI, and origin–destination tasks to test whether apparently strong model performance reflected genuine generalization or benchmark exposure.
- Automated prompt optimization and large-scale inference across **8 open-weight LLMs**; developed GeoSim and Contamination Inflation metrics and applied Mann–Whitney U tests with Benjamini–Hochberg correction, identifying **benchmark inflation above 60%** on POI tasks.

ProcVQA: Vision-Language Model Evaluation Benchmark | *VLMs, Hugging Face, GCP, Pydantic*

- Built scalable evaluation pipelines and benchmarking protocols for **21 state-of-the-art vision-language models**, including GPT-4.1 and Llama-family models, using structured outputs and controlled multimodal reasoning tasks.
- Identified a **19–56% performance decline** from single-hop to multi-hop reasoning and characterized systematic hallucination, instruction-following, and structural-reasoning failures as visualization complexity increased.

Neighborhood-Level Mobility Model Evaluation | *Python, AWS Bedrock, GCP, Polars, XGBoost*

- Constructed point-, trajectory-, and temporal-mobility prediction benchmarks from anonymized traces covering **976K+ devices** and **8,756 census block groups** across four U.S. metropolitan areas, integrating sociodemographic and built-environment features.
- Compared LLM predictions with XGBoost, Random Forest, Decision Tree, and Histogram Gradient Boosting baselines and developed OLS- and Jonckheere–Terpstra-based structural-alignment audits to identify **geographic priors and subgroup-specific failure modes**.

Human Mobility Simulation with Deep Generative Models | *Reinforcement Learning, GANs, Transformers, CNNs*

- Evaluated MoveSim, a deep generative architecture combining a self-attention generator, CNN discriminator, and REINFORCE policy for synthetic human mobility generation.
- Benchmarked the model across sparse Foursquare and dense GeoLife trajectory datasets, characterizing how data sparsity affected trajectory quality, generative performance, and generalization.

Technical Skills

Programming & Scientific Computing: Python, NumPy, pandas, Polars, R, SQL, C++
ML & Statistical Modeling: PyTorch, TensorFlow, scikit-learn, statsmodels, XGBoost, Time-Series Forecasting, Spatiotemporal Modeling, GNNs, Transformers, Reinforcement Learning, Causal Inference
Model Evaluation: Experimental Design, Benchmarking, Robustness Evaluation, Failure Analysis, Subgroup Analysis, Ranking Metrics, Hypothesis Testing, Model Comparison
Foundation Models: Hugging Face, vLLM, DSPy/GEPA, AWS Bedrock, LLMs, VLMs, RAG, Prompt Optimization
Data & Infrastructure: AWS, GCP, Spark, Hadoop, Snowflake SQL, PostgreSQL, GitHub, GitLab

Selected Achievements & Service

Awards & Honors: Jacob K. Goldhaber Travel Grant (2025), CS Department Travel Grant (2025), CS Summer Research Fellowship (2021), Undergraduate University Merit Scholarship (2013)

Peer Review Service: KDD (2021, 2022), WSDM (2022), ACM COMPASS (2023–2024) and other multidisciplinary journals

Mentoring: Mentored **10** students (**8** undergraduate, **2** master's) on various research projects