

Rachit Sabharwal

📍 Cypress, Texas 📞 (585) 281-1928 💬 rachitest

Professional Summary

Ph.D. Candidate in Biostatistics with dual expertise in Real-World Evidence (RWE) generation and Machine Learning. Experience at Genentech, BMS, and Dow developing graph neural networks, NLP pipelines, and Bayesian uncertainty frameworks. Proven ability to translate complex biomedical data (EHR, claims, omics) into actionable insights for drug discovery and safety signal detection.

Education

The University of Texas Health Science Center at Houston

Doctor of Philosophy in Biostatistics

Houston, TX

Aug 2022 - present

- Minors: Epidemiology, Health Economics
- Certificates: Advanced Data Science

The University of California, Berkeley

Certificate in Software Development and Programming

Berkeley, CA

June 2023 - Apr 2025

The University of Texas Health Science Center at Houston

Master of Science in Biostatistics

Houston, TX

Jan 2020 - May 2022

- Certificates: Data Science

University of Rochester

Bachelor of Science in Environmental Health

Rochester, NY

Sept 2014 - May 2018

Experience

Research & Development Intern, MiLOS (Machine Learning, Optimization, & Statistics),

Engineering & Process Sciences, Core R&D

Lake Jackson, TX

The Dow Chemical Company

May 2025 - Aug 2025

- Authored internal guidance on Uncertainty Quantification (UQ), demonstrating that Bayesian frameworks improved model trustworthiness by ~20-30% compared to frequentist baselines in simulated real-world scenarios.
- Created R application allowing users to automate up to 50% of LCA (Life Cycle Assessment) analysis, estimated to be worth ~15MM annually across Dow

Research and Early Development, Development Sciences & Informatics - Informatics Intern

Genentech

San Francisco, CA

May 2021 - Jan 2022

- Used Deep Transfer Learning via PyTorch and Raytune to create a DNN to predict ADEs for DILI (Drug-induced Liver Injury)
- Created a Knowledge Graph with Neo4j and a Graph Neural Network using NetworkX and PyTorch to generate gene expression signature-likes for drugs
- Developed a framework for the tokenization of internal documents for ingestion into text-mining application

Consumer & Market Knowledge - Advanced Analytics Co-Op

Procter & Gamble

Cincinnati, OH

Jan 2021 - May 2021

- Created predictive models, analytics, and visualizations that facilitated a deep understanding of consumer and shopper behaviors
- Used parallel computing (Dask and Modin) to develop both predictive and explanatory models enabling insights into market trends and retailer behavior
- Created and serviced big data ETL pipelines utilizing the Google Cloud Platform, Python, and Apache Airflow

Data Engineering Intern

Bristol Myers Squibb

San Francisco, CA

June 2020 - Aug 2020

- Utilized Python, HTML, CSS, and Javascript in creating a multifeatured patent recommendation app to significantly improve scientists' workflow
- Developed and serviced ETL pipelines using Python and Apache Airflow for multiple datasets of varying sizes (small, medium, and large)
- Designed and maintained both relational and graph databases in PostgreSQL and Neo4j

Publications

Trust and Uncertainty Quantification in Machine Learning Models Under Measurement

Error

Aug 2025

Sabharwal R

The Dow Chemical Company, Internal White Paper

Factors associated with elevated SARS-CoV-2 immune response in children and adolescents	Aug 2024
Messiah SE, Abbas R, Bergqvist E, Swartz MD, Talebi Y, Sabharwal R , Han H, Valerio-Shewmaker MA, DeSantis SM, Yaseen A, Gandhi HA, Amavisca XF, Ross JA, Padilla LN, Gonzalez MO, Wu L, Silberman MA, Lakey D, Shuford JA, Pont SJ, Boerwinkle E 10.3389/fped.2024.1393321 (Frontiers in Pediatrics)	
Baseline characteristics of SARS-CoV-2 vaccine non-responders in a large population-based sample	May 2024
Yaseen A, DeSantis SM, Sabharwal R , Talebi Y, Swartz MD, Zhang S, Leon Novelo L, Pinzon-Gomez CL, Messiah SE, Valerio-Shewmaker M, Kohl HW 3rd, Ross J, Lakey D, Shuford JA, Pont SJ, Boerwinkle E 10.1371/journal.pone.0303420 (PLoS One)	
An Interactive Online Dashboard with COVID-19 Trends and Data Analysis in Northeast and South Texas	Apr 2024
Zhang Z, Sabharwal R , Lee M, Zhang K, McGaha P, Crum M, Bauer C, Fisher-Hoch SP, McCormick JB, Reininger BM, Thomas S, Guajardo E, Pinon D, Yaseen A research.ebsco.com/linkprocessor/plink?id=894625e1-7146-30bf-aa2c-9f5637dac41e (Texas Public Health Journal)	
Long-term immune response to SARS-CoV-2 infection and vaccination in children and adolescents	Oct 2023
Messiah SE, Talebi Y, Swartz MD, Sabharwal R , Han H, Bergqvist E, Kohl HW 3rd, Valerio-Shewmaker M, DeSantis SM, Yaseen A, Kelder SH, Ross J, Padilla LN, Gonzalez MO, Wu L, Lakey D, Shuford JA, Pont SJ, Boerwinkle E 10.1038/s41390-023-02857-y (Pediatric Research)	
Scholarly recommendation systems: a literature survey	June 2023
Zhang Z, Patra BG, Yaseen A, Zhu J, Sabharwal R , Roberts K, Cao T, Wu H 10.1007/s10115-023-01901-x (Knowledge and Information Systems)	
Biorec: A Biomedical Recommendation System for Academic Conferences and Journals	Apr 2022
Sabharwal, R UTHealth School of Public Health, MS Thesis	
Data Cleaning for eCommerce: Standardizing Data Handling Practices for eCommerce Datasets	May 2021
Sabharwal R Procter & Gamble, Internal White Paper	

Skills

Languages: English (Native/Bilingual), Hindi (Native/Bilingual), French (Intermediate)

Work Authorization: US Citizenship, Canadian Citizenship

Technical Skills

Machine Learning Tooling: Scikit-learn, TidyModels, Raytune, Optuna, Pytorch, Tensorflow, Huggingface, JAX

Programming Languages: Python, R, SAS, MATLAB, Javascript, C, Java, HTML, CSS

Databases: Research Electronic Data Capture (REDCap), RDBMS (PostgreSQL, SQLite, MySQL), NoSQL DBMS (MongoDB, Elasticsearch, Neo4J), BigQuery

Cloud and Distributed Computing: AWS (AWS HPC), GCP, Azure, Spark, Hadoop, Slurm, On-Prem HPC

DevOps: Git, GitHub, GitLab, Docker, GitHub/GitLab CI/CD, Jenkins, Kubernetes, Jira, Confluence

Workflow Orchestration: Airflow, Prefect, Cron, Luigi

Frameworks and Platforms: Streamlit, FastAPI, Django, Flask, Heroku, Replit, Great Expectations, PyTest