

Shweta U. Narkar, Ph.D.

+1 (646)-239-5325 | shweta.3695@gmail.com | [linkedin.com/in/shwetanarkar](https://www.linkedin.com/in/shwetanarkar) | [Google Scholar](#)

PROFESSIONAL SUMMARY

PhD-trained machine learning and data scientist with experience developing ML models, building data systems, and evaluating predictive workflows for scientific applications. Skilled in Python-based ML pipelines, dataset design, and model evaluation across research and industry environments. Industry experience at IBM Research and current work on ML systems for speech analysis and model benchmarking.

EXPERIENCE

AI/ML & Data Scientist (Volunteer)

Feb 2026 – Present

Mirro.ai

Remote

- Run speech-analysis AI services locally and interact with model APIs through Swagger to test speech emotion model outputs.
- Execute audio inference workflows and analyze JSON outputs from speech emotion recognition models.
- Examine Python model code and service architecture to understand model inference pipelines and data flow.
- Investigate benchmarking datasets and evaluation metrics for speech emotion recognition models to support model performance analysis.

Postdoctoral Research Associate (Computational Science)

Aug 2025 – Present

Rensselaer Polytechnic Institute

Troy, NY

- Collaborate with experimental teams to transform laboratory workflows into structured, ML-ready datasets for downstream modeling and analysis.
- Design and curate scientific datasets enabling reproducible data pipelines and predictive modeling.
- Mentor graduate and undergraduate researchers on data preparation, exploratory analysis, and modeling workflows.

Machine Learning Research Intern

May 2020 – Sept 2020

IBM Research

Yorktown Heights, NY

- Designed and implemented Model LineUpper, a system for multi-level comparison of machine learning models enabling expert evaluation beyond aggregate performance metrics.
- Led a controlled user study with 14 professional data scientists evaluating 16 machine learning models (Decision Trees, Random Forest, Logistic Regression, LightGBM), identifying expert model comparison strategies.
- Applied explainable AI techniques (SHAP-based feature importance) to analyze model behavior and support instance-level error diagnosis.
- Co-authored peer-reviewed research presented at the 2021 Intelligent User Interfaces Conference.

Research Assistant (PhD Researcher – Applied ML & Data Analysis)

May 2021 – May 2025

Rensselaer Polytechnic Institute

Troy, NY

- Modeled experimental outcomes using machine learning (Random Forest, Gradient Boosting) on a curated dataset, achieving $R^2 = 0.734$ and informing the design of 24 follow-up experiments.
- Built a reusable data ingestion framework (Express.js, MongoDB) capturing 190+ experiments, enabling scalable dataset growth for downstream modeling and analysis.
- Developed the first ontology for prebiotic chemistry, enabling structured annotation and reuse of 200+ experiments across multiple research areas.
- Collaborated with experimental scientists to analyze datasets and develop visualizations supporting hypothesis generation.

Research Assistant (Scientific Data Systems)

Jun 2018 – Apr 2021

Rensselaer Polytechnic Institute

Troy, NY

- Analyzed and visualized microbial locality datasets spanning 1000+ enzymes, supporting biochemical and geobiological research.
- Curated and standardized 10+ archival scientific datasets for the NASA-funded Earth First Origins project, improving data accessibility for downstream analysis.
- Contributed to ontology-based knowledge systems supporting structured scientific data integration across research initiatives.

SELECTED PUBLICATIONS

- Narkar et al. *Model LineUpper: Supporting Interactive Model Comparison at Multiple Levels for AutoML*. Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI), 2021.
- Cluster Analysis of Presolar Silicon Carbide Grains: Evaluation of Their Classification and Astrophysical Implications. *Astrophysical Journal Letters*, 2021.
- Invited Talk: Leveraging Computational and Semantic Technologies in Understanding Prebiotic Chemistry. Geological Society of America, 2023.
- Talk: Enabling Semantic Technologies and eScience for Exploring the Prebiotic Chemistry Experimental Landscape. AbSciCon, 2024.

EDUCATION

Rensselaer Polytechnic Institute	Troy, NY
Ph.D. in Multidisciplinary Science (Computer Science + Astrobiology)	May 2025
Rensselaer Polytechnic Institute	Troy, NY
M.S. in Data Science & Analytics	May 2019
University of Pune	India
B.E. in Computer Engineering	Aug 2017

TECHNICAL SKILLS

Programming: Python, R

Machine Learning & Data Science: scikit-learn, pandas, NumPy, statistical modeling, model evaluation, SHAP

Data Systems & Tools: MongoDB, PostgreSQL, Git, Docker, Azure DevOps

Visualization: Matplotlib, seaborn, ggplot2, D3.js

Familiar: Hugging Face Transformers

LEADERSHIP & SERVICE

- Student Officer and Interim Secretary-Treasurer, Geoinformatics & Data Science Division, Geological Society of America.
- Teaching Assistant, Rensselaer Polytechnic Institute (CS I; Econometrics in Big Data with R).
- Mentor, Rensselaer Research Experience program supporting undergraduate research training.
- Recipient, RPI Founders Award of Excellence.