Roozbeh Ehsani

Email: ehsan010@umn.edu ▪ Tel: (734) 510-4673 ▪ Location: MN (Open to Relocation) ▪ LinkedIn | Google Scholar | Git Hub

EDUCATION

University of Minnesota, Doctorate *Jan 2022- Dec 2025(Expected)*

- Field: Developing Generative models, Stochastic Modeling | **GPA**: 4
- Honors: Roger E.A. Arndt Fellowship, Graduate Fellowship, Frank and Julie Tsai Award

University of Minnesota, Master *Sep 2022- Dec 2024*

- Major: Computer Science and Engineering (CSCI) Department | **GPA**: 3.96

EXPERIENCE

Data Science Intern **Solventum (3M), Saint Paul, MN** **May 2025 - Aug 2025**

1. Material Reinvent ($3B project):

- Built an interactive **Dash app** to search for connectivity of company materials, visualize the material graph (**Cytoscape**); click nodes to display each material's test methods, enabled instant (**2s**) hierarchy navigation, analysis, and comparison.
- Optimized test strategy using machine learning (ML) clustering methods to segment materials by similar test-value ranges; testing one representative per cluster, **reducing the required number of tests and expenses by more than 80%**.
- Engineered an end-to-end API to automate the full workflow–from material search, graph generation, and statistical analysis (**ETL**)– containerized the API using Docker and deployed to **AWS (ECR, ECS)** for scalable and efficient execution.

2. Text Mining:

- Designed a dictionary-based schema to standardize unstructured Excel file formats, enabling machine readability (Copilot agent), and cut data parsing time from 10 min to seconds with higher accuracy.
- Developed text-mining pipelines to extract and structure key information from unstructured clinical study datasets, **delivering critical dataset to cross-functional teams** to accelerate validation of reinvented materials and test devices.

Graduate Research Scientist **University of Minnesota, Minneapolis, MN** **Jan 2022 - Present**

- Analyzed **large raw dataset** and used statistical method to detect underlying patterns and extracted key characteristics; provided clean and required dataset for downstream models (**EDA, Data Wrangling**) (Scholar1, Scholar2).
- Designed and implemented a computationally efficient stochastic–ML hybrid model to generate high-fidelity synthetic data, **reproducing more than 80%** of the first and second-order statistical moments (**Monte-Carlo simulation**)(Scholar).
- Developed a computationally efficient stochastic framework for generating high-fidelity 2D synthetic fields, eliminating costly physical simulations and accurately reproducing the energy spectra and statistical behavior of **complex systems** (Scholar).
- Developed a heuristic **Bayesian-stochastic** framework leveraging Mixture Density Networks (Deep Learning) to synthesize **time-series dataset**, reproducing high-resolution spectral energy distributions observed in experimental datasets (Git).
- Collaborated with a multidisciplinary DOE-funded team to design a sediment-bypass system for hydropower dams; led slurry-flow experiments, cleaned noisy sensor data, and developed data-driven, physics-based models to inform pump design and improve sediment transport efficiency (OSTI.GOV).

Data Science Projects

- **Online Retail:** Analyzed the Online Retail dataset by computing **Recency**, **Frequency**, and **Monetary** (**RFM**) metrics for each customer, performing EDA and visualizations, and segmenting customers into actionable groups (Champion, Loyal, Lost, etc.). Presented segment-level insights in **Tableau** to support data-driven marketing initiatives, including **Customer Lifetime Value** (**CLV**) modeling, churn analysis, recommendation systems, promotion design, and retention strategies (Git).
- Designed, conducted, and evaluated an **A/B test** for a sales-training program by calculating sample size through power and MDE analysis, estimating test duration, and performing confidence-interval evaluation; enabling the quantification of **KPI lift** and determining whether the training program delivered a statistically significant and cost-effective improvement (Git).
- **Advanced Machine Learning:** Developed a reinforcement learning–based controller to dynamically switch optimization strategies while training an **LSTM** model, demonstrating faster convergence and robust training performance on benchmark **NLP datasets** (Git).
- **Computer Vision:** Built a 3D environment reconstruction pipeline along with **semantic information (U-net)** to generate environment maps, supporting applications in navigation, simulation, and AR/VR systems (Git).
- **Recommendation Systems:** Implemented and compared multiple recommendation system models, with the item-based approach achieving the best performance (RMSE = 1.0) on the MovieLens-100K dataset (Git).
- **Database Management Systems (DBMS):** Implemented and benchmarked the performance of B-tree and Hash indices for range and exact-match queries, highlighting key considerations for **system design** and product performance optimization.

SKILLS

- **Analytical:** Generative Stochastic Models, Statistical Analysis, Data Mining, Machine/Deep Learning, DBMS, A/B Testing
- **Technical:** Python (scikit-learn, Pandas, PyTorch), SQL, EDA, ETL, AWS, Docker, Git, CI/CD, MLOps, UI, Power BI