Naseela Pervez

LinkedIn| naseelasidiqi@gmail.com | GitHub | Los Angeles | 574-387-9779

Education

University of Southern California Jan 2022-Dec 2023
Master of Science, Computer Science(Data Science) GPA: 3.54/4.0

SRM Institute of Science and Technology (India) Jul 2017 - Aug 2021
Bachelor of Technology, Information Technology GPA: 8.9/10.0

Technical Skills

- **Languages**: Python, C++, R, SQL.
- **Frameworks**: PyTorch, Tensorflow, Numpy, Pandas, Statsmodels, Scikit-Learn, PySpark, Matplotlib, Seaborn, HuggingFace.
- **Tools & Cloud Platforms**: Jupyter, Git, VS-Code, OpenAI, Open Hermes, VertexAI, Gemini, Mistral AI, Google Cloud Platform.

Experience

USC Information Science Institute (ISI) Los Angeles, CA
Research Assistant Feb 2024-Present

- Conducted a comprehensive analysis of personality traits and gender bias in scientific abstracts generated by Large Language Models (LLMs), comparing human-written and AI-generated texts using the LIWC tool.
- Developed and implemented evaluation frameworks to assess the alignment of lexical and psychological features in texts, contributing to insights on gender disparities and biases in AI-generated scientific literature.
- Leverage network science to explore the **glass ceiling effect** in patent citation networks and undercover the effects of assignee organization prestige and gender gaps of patent assignments (ongoing).

***Management of Innovation, Entrepreneurial Research, and Venture Analysis (MINERVA)* - USC** Los Angeles, CA
Research Staff Feb 2024-Present

- Designed a novel clustering scheme to generate domain-specific labels for a set of scientific abstracts using diverse keywords and incorporated background information specific to the document employing Mistral Large(7B).
- Collaborated in a team of 6 to implement novel unsupervised multi- label classification algorithm focusing on importance of background information needed for labeling scientific documents.

USC Libraries Los Angeles, CA
Data Analysis & Visualization Student Assistant May 2022-Dec 2023

- Performed detailed analysis of library vendors and identified overlaps in journal vendors resulting in per-usage cost reduction of 12%.
- Communicate and inform strategic decisions about programs, services, and resource allocation from library usage metrics.

SRM Institute of Science and Technology Chennai, India
ML/AI Research Assistant Sep 2020-Dec 2021

- Developed a high-throughput data pipeline using Tweepy and Kafka to gather 1M COVID-19 tweets focused on mental health, and fine-tuned a BERT model achieving 87% classification accuracy.

Publications

- Pervez, N., & Titus, A. J. (2024). Inclusivity in Large Language Models: Personality Traits and Gender Bias in Scientific Abstracts. (Accepted at NLAI 2024).
- Sakhrani, S., Pervez, N., Kumar A., Morstatter, F, Graddy-Reed & A., Belz, A.(2024) Artificial Intuition: Efficient Classification of Scientific Abstracts (Accepted at SDProc at ACL 2024).
- Pervez, N., & Titus, A. J. (2024). Integrating MLSecOps in the Biotechnology Industry 5.0.IntechOpen. doi:10.5772/intechopen.114972)

Affiliations

- Scheduled to attend **ACL 2024**: Presenting "Artificial Intuition: Efficient Classification of Scientific Abstracts" at SDProc workshop.
- Mentor for the BUGS Jr. program at MINERVA USC, guiding high school students in their research in Natural Language Processing (NLP) projects, specifically network science and document classification techniques.
- Active member of Rewriting The Code (RTC), and Society of Women Engineers (SWE) promoting the inclusion of women in STEM.

Course Projects

Does Explainable AI Agree With Psychologists? (Machine Learning at USC)

- Modeled classification algorithms on LIWC and scored features(Big5, IRI, and MFT) of texts classified as high or low classes.
- Verified lack of alignment in LIWC and personality traits of text considering 50% of words yielded important by LIME and SHAP.

Restaurant Recommender System (Data Mining at USC)

- Designed a hybrid recommendation system based on item-based collaborative filtering and XGBoost regression to predict restaurant ratings for a user resulting in RMSE of 0.97 from 1.09 baseline.

Little Go (Foundation of Artificial Intelligence at USC)

- Developed robust 5*5 Go game player using minimax search algorithm with alpha-beta pruning achieving 90% win rates.