VENKATA SESH TEJ MATTA

State College, Pennsylvania, 16801
814-769-1326 vmm5481@psu.edu linkedin.com github.com

Education

Pennsylvania State University- University Park Campus **Jan. 2024 – May. 2025**
Master of Science in Computer Science and Engineering *GPA : 3.5/4*

Indian Institute of Technology- Madras (IIT- Madras) **Jun. 2018 – Jun. 2022**
Bachelor of Technology in Chemical Engineering *GPA : 7.68/10*
Minor degree in Systems Engineering *GPA : 8.3/10*

Experience

Pennsylvania State University — Part Time — Teaching Assistant **Aug. 2024 – Ongoing**

- Oversaw all aspects of Programming Language Concepts course, including assignment creation, exam supervision, project ideation, and student support in Python, Scheme, OOP, alongside coordinating course logistics and communications.

NLP lab — Part Time — Research Assistant **Jul. 2024 – Ongoing**

- **Simulation and Data Analysis in Python using PHYRE :** Developed and implemented Python scripts to simulate physical interactions using the PHYRE environment extracting object parameters such as position, velocity. Analyzed and extracted object parameters, ensuring data aligned with the ESPRIT dataset structure for machine learning tasks.

Jefferson lab — Summer Research Internship — AI Research Assistant **May. 2024 – Ongoing**

- Applied Generative Adversarial Networks (GANs) and Normalizing Flows to experimental data from laboratory studies and achieved 3D reconstructions of proton distributions within nucleus, contributing to advancements in modern physics and astrophysics research.
- Developed a Python-based data analysis pipeline utilizing JAM3D libraries to evaluate Transverse Momentum Dependent (TMD) parton distribution functions (PDFs). Implemented advanced 3D contour plotting techniques using Matplotlib to visualize complex TMD functions across different kinematic variables.

Graphene AI Health Tech Pvt.Ltd— Full Time — Machine Learning Engineer Jun. 2022 – Aug. 2023

- **Knowledge graph** : Conceptualized and Developed a comprehensive knowledge graph, leveraging Wiki-based Entity Extraction and Neo4J graph database, with a primary focus on optimizing data extraction queries. Successfully implemented knowledge graph to enhance data extraction processes, resulting in a 30% increase in the amount of relevant data extracted. Overcame challenges associated with handling large-scale graphs, demonstrating effectiveness in optimizing performance at scale while including ontology design and data persistence optimizations.
- **Relevance Filter pipeline** : Engineered and developed an automated fine-tuning pipeline for Graphene's in-house category-specific relevance filters with RoBERTa and DistilBERT as base models. Leveraged GPT-3.5 Turbo for oversampling small human-annotated training sets. Deployed over 15 category-specific relevance filters to production in 50% less time than previous manual efforts, each achieving an F1 score above 90%. Created an intuitive UI as an integral component of an automated framework, improving user experience & ease of interaction.

Projects

Natural Language Processing Project, Pennsylvania State University **Jan. 2024 – May. 2024**

- Created a model to classify dialogue segment breaks in situated language interactions using the GUIDE dataset. Utilized NLP models including T5, MIST and RoBERTa to analyze linguistic cues and predict segment transitions in dialogue.
- Demonstrated T5-large model outperformed others, emphasizing its capability in complex dialogue structure interpretation. Managed full project lifecycle from conceptualization to deployment, enhancing practical understanding of machine learning and NLP applications.

Bayesian Online Change point Detection Algorithm, IIT -Madras **Jul. 2020 – Nov. 2020**

- Implemented a recursive Online Bayesian estimator in MATLAB, developed by Adams and MacKay, for detecting change points in streaming data via sequential analysis. Tested implementation on well-drilling NMR data using a normal-gamma prior and successfully Detected positions of change points and obtained estimates for mean and variance.

Publications

Loan Distribution and Prediction System using ML—IJARESM **Jan. 2024**

- Co-authored "Loan Distribution and Prediction System using ML," published in the International Journal of All Research Education and Scientific Methods (IJARESM), January 2024, ISSN: 2455-6211.
- Developed a predictive model leveraging machine learning to assess and mitigate risks in loan approvals

Recognize Captcha using Neural Networks —IJSREM **Nov. 2023**

- Co-authored "Recognize Captcha using Neural Networks," International Journal of Scientific Research in Engineering and Management, demonstrating CAPTCHA-solving via deep learning (Nov 2023, ISSN: 2582-3930).

Technical Skills

Languages: Python, C, C++, MATLAB, SQL
Tools: PyTorch, TensorFlow, vLLM, HuggingFace, LangChain, OpenAI, Scikit-Learn, SpaCy, Docker
Databases: MS-SQL, Neo4j,PostgreSQL