Koushik Reddy Parukola

Bloomington, IN (703)-789-1618 parukolakoushik@gmail.com Portfolio LinkedIn GitHub

SUMMARY

- Extensive experience in machine learning, deep learning, and generative AI, focusing on NLP and fine-tuning large language models (**LLMs**) such as **BERT**, **LLaMA**, and **GPT**. Skilled in designing scalable ML pipelines and optimizing data preprocessing for high model performance.
- Proficient in **NLP pipeline** development and research at the intersection of quantum computing and ML, with contributions to **Quantum Natural Language Processing (QNLP)** to advance computational complexities.
- Background in specialized deep learning models for drug data analysis and knowledge graph pipelines for medical and disease-related data, leveraging **high-performance computing** for biomedical applications.
- Skilled in quantum tools like **Qiskit** and **PennyLane** for Quantum Machine Learning **(QML)**, including QNNs and quantum classifiers, with expertise in quantum data encoding and circuit optimization.
- Strong interdisciplinary approach, integrating **quantum systems** into classical **NLP** frameworks and applying both classical and quantum computing to solve complex computational challenges.

TECHNICAL SKILLS

Programming Languages: Python, Java, R, C, C#, SQL/NoSQL, and [Qiskit, PennyLane, Cirq, Lambeq (Quantum)]

Quantum Computing: Quantum Neural Networks (QNNs), Variational Quantum Algorithms (VQA), Amplitude, Basis and Angle Encoding, shors Algorithm, Quantum Fourier Transform & Phase Estimation, IBMQ, Rigetti, PennyLane.

Artificial Intelligence (ML, DL and NLP) : Supervised(Classification, Regression, SVM, Decision Trees, KNN, Naive Bayes, XGBoost, Random Forests), Unsupervised (Clustering, Pattern Search, K-Means, PCA, SVD, LDA), LSTMs, CNNs, RNNs, Vector Embeddings/Databases, Knowledge Graphs, Text classification, Sentiment Analysis & Topic Modeling, Named entity Recognition, LangChain and RAG models, Transformers models, LoRA, LLMs.

RESEARCH EXPERIENCE

Quantum - NLP Lab @ Indiana University, *Apr 2024 - Present | Bloomington, Indiana*

Research Assistant

- Conducted an extensive comparison of word class representations using classical word embedding models (**Word2Vec**, **GloVe**) and large language models (**GPT**, **Claude**, **LLaMA**, **Mistral**).
- Reduced high-dimensional word embeddings (e.g., 3000+ in **GPT**) through dimensionality reduction techniques like **Fisher's Discriminant**, **PCA**, and **T-SNE** to enable effective clustering and performed clustering analysis with intra- and inter-cluster distance evaluation to explore the semantic and syntactic relationships within and across word classes.
- Extended research by analyzing layer-wise word embeddings in **transformer models** to understand their capture of contextual and semantic nuances across layers. Investigated layer-specific representation transformations, revealing insights into the progressive encoding of context and meaning within **transformer networks**.
- Developed **quantum classifiers** that encode word embeddings into **quantum circuits**, enabling the comparison of semantic similarity in the quantum versus classical space, driving innovations in **quantum natural language processing (QNLP)**.
- Utilized **amplitude encoding** and **angle encoding** to map high-dimensional word embeddings into quantum circuits, addressing scalability challenges by optimizing encoding algorithms to reduce circuit depth and improve efficiency.
- Improved **quantum model accuracy** by implementing error correction techniques and quantum circuit transpilation, reducing noise and optimizing circuits for **IBM Quantum** / **PennyLane** hardware and simulators.
- Developed feature extraction techniques using two-gate **ansatz circuits** in quantum classifiers, optimizing the parameterized quantum circuits to effectively extract relevant features from encoded word embeddings, enhancing model performance and reducing circuit depth in **quantum NLP** applications.

PROJECTS

Regression with Variational Quantum Circuits , *Quantum Computing* *Sept 2024 - Oct 2024 | Bloomington, IN*

- Developed a quantum-classical hybrid model for multidimensional regression using Variational Quantum Circuits (VQC), applying quantum computing principles to solve complex regression tasks.
- Classical data was encoded into quantum states using angle embeddings technique, enabling quantum transformations that processed input features via qubit rotations.
- The quantum circuit leveraged strongly entangling layers, allowing qubit interactions to model intricate relationships between data features, boosting the circuit's complexity and flexibility.
- Training was accelerated through JAX's just-in-time compilation, with circuit parameters optimized via Optax's Adam optimizer, ensuring precise and efficient updates during each step.

- The model's performance, demonstrated by the R2 score, showcased its ability to generalize over the quadratic target function and handle multidimensional data accurately. Generated 3D surface plots using Matplotlib, providing a visual comparison of the predicted function against the true target function, facilitating deeper analysis of the model's performance.
- Combined quantum computing techniques with **classical optimization** to explore the practical applications of quantum machine learning for real-world regression problems, showcasing the potential for high-dimensional data modeling.

Advanced Bio-Medical Knowledge Graphs - LLM Integration, *NLP* *Jan 2024 - May 2024 | Bloomington, IN*

- Investigated the integration of Large Language Models (LLMs) to enhance medical knowledge graph accuracy, focusing on reducing hallucinatory outputs crucial for clinical reliability. Focussing on patient-disease-diagnosis-drug relationship.
- Created an automated Named Entity Recognition (NER) system using advanced models like BioBERT, GPT-4, and other fine-tuned models from Hugging Face Transformers, achieving up to 89.1% accuracy in extracting medical entities, which facilitated the construction of knowledge graphs from complex medical texts.
- Transformed NER outputs into subject-predicate-object triplets and seamlessly integrated these into knowledge graphs, enhancing the graphs' structural richness and functional utility for clinical decision-making.
- Empirical analysis and model testing demonstrated that the methodologies applied significantly enhance the precision and reliability of data handling in medical AI applications, with a notable improvement in entity and relation extraction accuracy, achieving up to 83.1% with Clude-sonnet, 85.8% with GPT 4 88.2% with Claude-3 Opus.

Cancer Drug Synergy Prediction, *Bio-Informatics - Academic Research* *Jan 2023 - July 2023 | Bloomington, IN*

- Engineered a sophisticated NN model for predicting cancer drug synergy, determining the combined effect of two or more drugs surpassing the sum of their individual effects in enhancing the precision and effectiveness of cancer treatment regimens.
- Trained with the NCI ALMANAC dataset, augmented with synthetic data comprising 130,182 data points, with 60 human cancer cell lines derived from 9 different tissues, achieving an outstanding AUC score of 90.81%, Using techniques like Feed Forward Neural Networks, Autoencoders and ERT's etc, surpassing traditional models like SVM's.
- Conducted in-depth analysis on drug fingerprint data, drug targets, and gene expression, utilizing ReLU and ADAM for model optimization using datasets like AstraZeneca's, contributing significantly for understanding cancer cell line behaviors.

WORK EXPERIENCE

Luddy school of Informatics computing and Engineering, *Aug 2023 - Dec 2023 | Bloomington, Indiana*
Associate Instructor for CSCI-C 292:

- Worked as an Associate Instructor alongside Professor Chabane Maidi on Virtual/Augmented Reality and Programming course (CSCI-C 292) with C# and the Unity engine.
- Researched AI algorithms like MCTS for estimating outcome likelihood and an optimized A* pathfinder, reducing computation time by 15% and steps from 100 to 20, enhancing efficiency in VR and games.
- Developed pathfinding and decision-making for NPCs using finite state machines, rule-based AI, and heuristic algorithms to improve realism in game environments.
- Automated NPC responses with NLP, enabling them to understand and respond conversationally using generative AI, enhancing player interactions and immersion.

EDUCATION

Indiana University , *Masters in Data Science, CGPA: 3.6* *Aug 2022 - May 2024 | Bloomington, Indiana*
Relevant coursework: Elements of Artificial Intelligence, Computational Linguistics, Neural Language Models, Applied Machine Learning, ML in Bio-Informatics, Data Mining, Statistical and Exploratory Data Analysis and Virtual Reality.

Natural Language Processing Lab at Indiana University - Knowledge Grpahs and Disease/Drug Ontology.
Quantum QNLP Lab at Indian University - Encoding classical bits to quantum bits(Qubits).

New Horizon College of Engineering *July 2017 - Aug 2021 | Bangalore, India*
Bachelor of Engineering in Computer Science and Engineering, CGPA: 8.76
Relevant coursework: Data Mining, Database Management, Data Structures, Computer Organization, Operating Systems with Scheduling Algorithms, Big Data Analytics, Introduction to Artificial Intelligence, Machine Learning with Python.
Vice President - App Development club @ NHCE

Publications & Certifications

- IBM Quantum Challenge - Completed IBM Quantum Challenge 2024 on Entanglement, Transpilers, Circuit Knitting, VQEs, W state Ansatz, Parameterized Circuits. Completed in top 11%.
- Traffic Sign Classification - 'Innovative Journal of Innovative Science and Research Technology' Volume 6 - Issue 6 - June 21.
- A paper on Classical Word embeddings to Quantum Circuits techniques and similarities - NAACL 2025 - Under Review
- Advanced Knowledge Graphs Integrating with LLMs - NERs to Relation Extraction and Language Models - In Progress.
- NLP with Transformers - Advanced Text Processing and Large model Training and Fine Tuning with Hugging Face transformer. - Certificate.