Kartik Choudhary

contact@kartikc.com | linkedin.com/in/kartikc727 | San Francisco Bay Area, CA

SKILLS

Tools and Languages	:	Python, C++, SQL, Java, Linux, CUDA, LaTeX, Git, Bash, XGBoost, Jupyter, TensorBoard
Big Data and Cloud	:	Spark, Hadoop, PySpark, Tableau, Grafana, Streamlit, Docker, Kubernetes, AWS, GCP, Azure
Data Science and ML	:	PyTorch, TensorFlow, ONNX, HuggingFace, Keras, MLFlow, InterpretML, MongoDB, Postgres

PROFESSIONAL EXPERIENCE

Machine Learning Engineer II, TikTok USDS   **Jul 2024 - Present**

Leading the design and optimization, and serving as the primary point of contact for the TikTok app search suggestions feature.

- Implementing a search candidate raking method using a *BCE+LTR hybrid loss function* to improve search relevance and reduce latency for the US and EU markets.
- Implemented a search query scattering strategy to increase diversity in search results for queries with unclear search intent.
- Collaborated with the architecture team to increase the number of query candidates by *8x* with minimal latency impact, improving search coverage and increasing suggestions click-through rate by *0.2%*.

Machine Learning Engineer, Reliance Jio Infocomm Ltd.   **Jul 2021 - Jul 2022**

- Coordinated a team of 6 for optimizing network coverage using *RSM* leading to 1.2M fewer daily calls with network issues.
- Attained a *15%* improvement in network throughput predictions by implementing a *Mixture of Experts* model with PyTorch.

Data Scientist, Reliance Jio Infocomm Ltd.   **Jul 2019 - Jun 2021**

- Deployed a distributed model with *Apache Spark* and *MS Azure* to perform churn prediction for *400M+* daily customers.
- Curated *over 1100* features for churn prediction to improve customer retention resulting in *$35M/mo* in additional revenue.
- Reduced data pre-processing time by *70%* during ETL by implementing a *multiprocessing pipeline* for feature extraction.

RESEARCH AND INTERNSHIPS

ML Research Intern, Meta   **Feb 2024 - Jun 2024**

- Developed evaluation metrics for judge LLMs and studied the effects of their size and architecture on their performance.
- Designed and conducted experiments to assess the impact of instruction alignment and adversarial inputs on LLMs.

Data Science Intern, Microsoft   **Jan 2024 - Feb 2024**

- Added generative model support to Responsible AI Toolbox with evaluation metrics and hierarchical importance measures.
- Created a benchmarking framework for prompt templates and did user studies to enhance the Azure AI Studio and Copilot.

Graduate Student Researcher, University of Massachusetts Amherst   **Dec 2022 - Jan 2024**

- Developed off-policy algorithms to learn personalized sepsis treatment policies from historical data of ICU patients.
- Evaluated comparative reasoning capabilities of vision-enabled LLMs like GPT-4V by creating a novel dataset. [Best Project]

Software Developer Intern, Claym Media and Tech Pvt. Ltd.   **Apr 2018 - Jul 2018**

Research Intern, Voxomos Systems Ltd.   **Apr 2017 - Jul 2017**

PUBLICATIONS

Judging the Judges: Evaluating Alignment and Vulnerabilities in LLMs-as-Judges, Under review, ARR   **2024**

A study of the properties of LLMs as judges in comparison to humans and automated evaluation methods. This work provides a comprehensive analysis of popular LLMs such as *GPT-4* and *Llama 3*, and potential pitfalls in their use as judges. [First author]

ICU-Sepsis: A Benchmark MDP Built from Real Medical Data, Reinforcement Learning Conference   **2024**

Modeled sepsis management as an indefinite-horizon Markov decision process and utilized the EHRs of over 17k ICU patients to estimate the parameters of the MDP, which can be used to accelerate applied RL research in sepsis treatment. [First author]

PROJECTS

TaskLLM: Multi-Agent Planning and Execution   **2024**

A multi-agent system with a *task-oriented language model* that can generate and execute plans in a simulated environment.

AI Writing Copilot   **2024**

A writing assistant app built using *LangChain* and *FastAPI* with customizable levels of feedback and suggestions using *RAG*.

EDUCATION

University of Massachusetts, Amherst, MA   **May 2024**

MS in Computer Science

Indian Institute of Technology Delhi, India   **Jun 2019**

B.Tech in Electrical Engineering

HONORS AND ACTIVITIES

- Graduate Teaching Assistant, CS 687 and CS 550, UMass Amherst   2023-24
- Highest annual performance rating (95th percentile) for 3 years in a row, Reliance Jio   2020-22