Srikar Kashyap Pulipaka

Columbus, IN | +1 (812)-778-5178 | srikar.kashyap@gmail.com | linkedin.com/in/srikarkashyap | github.com/srikarkashyap | Google Scholar

Education

Indiana University, Bloomington **Aug 2022 - May 2024**
Master of Science in Computer Science *GPA 3.87 / 4.00*
Coursework: Applied Algorithms, Computer Networks, Applied Machine Learning, Deep Learning Systems, Computer Vision, Natural Language Processing (NLP), Graph Analytics.

Experience

Software Engineer **Jun 2024 - Current**
Cummins, Inc *Columbus, IN*

- Built a proof of concept for an e-commerce platform using Django, featuring user authentication, subscription management, and Stripe-powered payment gateways. Developed APIs with Django REST Framework and deployed on AWS EC2 with a MySQL database **supporting up to 10,000 concurrent users in testing and completing the project in under 4 weeks, reducing estimated development time by 30%**.
- Led a 6-member team in maintaining and enhancing a legacy Apache Cocoon web application built on Java. Delivered key fixes to long-standing production defects and implemented new features, **resulting in a 40% improvement in application performance and generating over $3M in additional revenue in three months.**
- Developed AI agents using Azure OpenAI APIs to automate software documentation generation and code snippet generation for the above project, **reducing team onboarding time by 50% (from 6 weeks to 3 weeks) and increasing documented code coverage from 25% to 90%.** Implemented a Retrieval Augmented Generation (RAG) system on top of the documentation, enabling user query resolution and reducing time to find relevant information by 30%.

Research Assistant and Engineer **Aug 2023 - May 2024**
Indiana University Department of Linguistics and Kelley School of Business *Bloomington, IN*

- Created a Python Streamlit-based tool for evaluating outputs of large language models (LLMs), including authentication mechanisms and MySQL-backed data storage. **It has supported over 20 researchers at IU in their work over the past 6 months.**
- **Built a large-scale data collection pipeline to gather 300,000+ HuggingFace models and linked GitHub codebases.** Analyzed the code repositories for security vulnerabilities using Semgrep and Bandit scanners, presented findings at sponsor meetings in San Francisco, and integrated results into the AI Risk Database.
- Competed in the PAN 2024 Text Classification Challenge, fine-tuning LLMs to predict conspiracy theory content in social media posts. **Achieved an F1 score of 0.83, winning the competition, and published findings at the CLEF - PAN 2024 conference**.

IT Intern **May 2023 - Aug 2023**
Cummins, Inc *Columbus, IN*

- Designed and developed a full-stack object storage application for internal design files, featuring link sharing, training module creation, and video embedding. Hosted the application on Amazon EC2, utilizing S3 for object storage and AWS RDS for relational data management.
- Developed a Machine Learning pipeline to predict customer satisfaction for a service organization, **achieving over 85% accuracy and enabling proactive prioritization of user requests.** Integrated data from multiple sources (SQL, CSV, Excel) and deployed the model using Amazon SageMaker, with results effectively visualized in a PowerBI dashboard.

Data Science and Software Generalist **Oct 2020 - Jul 2022**
Heritage Foods Limited *Hyderabad, India*

- Developed and deployed a real-time quality control system on Amazon EC2, **processing 50,000 events daily across 16 plants to monitor packaging weights via TCP-IP streams nationwide.**
- Utilized Amazon Simple Notification Service (SNS) for SMS and email alerts to **reduce reaction time for weight deviations from 30 minutes to 11 minutes, increasing packaging accuracy by 30% and saving $300,000 annually.**
- Automated data collection and processing by replacing manual workflows with Python scripts and scheduled API calls, **reducing dashboard data load time by 55% (from 4.5 hours to 2 hours) and saving over $100,000 annually.** Deployed daily processing jobs on Amazon EC2 with S3 storage, enabling seamless integration with SQL Server and Tableau. .

Skills

Programming Languages: Python, Java, MySQL, NoSQL, C, JavaScript
Data Platforms: MySQL, MS SQL Server, SQLite, MongoDB, Elasticsearch
Frameworks & Libraries: Django, Flask, PySpark, Transformers,, PyTorch, NLTK, Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn
DevOps and Other Tools: Git, GitHub, Jenkins, Docker, Tableau, PowerBI, HTML, CSS, Bootstrap, Postman

Projects

Question Answer Detection Using Computer Vision | *Python, Machine Learning, Deep Learning* **Jan 2023 - May 2023**

- Developed a computer vision system to detect question answers in children's responses using deep learning, **improving accuracy from 63% to over 80% through transfer learning techniques**. Also leveraged OCR for data extraction and traditional ML algorithms like SVM for detecting answer/no-answer and has-image/no-image classifications.

Self Hosted Photos App | *HTML5, CSS3, JavaScript, Flask, Python, Machine Learning, LLMs* **Oct 2024 - Dec 2024**

- Designed and developed a full-stack Python Django application for a locally hosted Photo Gallery and Editor, featuring clustering of similar photos using visual embedding algorithms, person face identification, and object-based search. Hosted via Docker with optional integration to Amazon S3 for file storage and auto-sync folder functionality using rsync.