Nikhil Pavan Kanaka

Columbus, OH | +1-813-705-2796 | nikhilpavan.kanaka@gmail.com | linkedin.com/in/nkanaka | github.com/npkanaka

RESEARCH INTERESTS

I am broadly interested in foundation-scale large language, speech, and multimodal models. My goal is to advance these models to further the development of general AI and enhance the effectiveness and reliability of language agents. I aim to create robust AI systems that can seamlessly integrate into various applications, significantly improving automation and accessibility.

ACADEMIC RESEARCH EXPERIENCE

Master of Science, Computer Science and Engineering, Ohio State University Jan 2023 – Dec 2024

- *Thesis In Progress*: Working on enhancing the reasoning capabilities of large language models (LLMs) by investigating parametric memory for complex reasoning tasks. Analysis of grokking in transformers and other architectures to identify performance and interpretability differences.
- Under the guidance of Dr. DK Panda, engineered JAVA Bindings for MVAPICH, a high-performance MPI library. Worked on enhancing computational workflows within NVIDIA's Spark-Rapids, GDS, and DALI, aiming to set a new benchmark for GPU-accelerated computing. (GRA Au23)
- Under the guidance of Dr. Rajiv Ramnath, designed and engineered a real-time emergency response support system for 911 aimed to enhance operator and field team efficiency. Enhanced critical information extraction capabilities by fine-tuning large language models achieving an improvement over state-of-the-art Named Entity Recognition models. Implemented a Retrieval-Augmented Generation framework that integrates vector databases with knowledge graphs for assessing medical conditions and optimizing emergency responses, unit deployment, and treatment suggestions.
- Under the guidance of Dr. Srinivasan Parthasarathy, Led the Voicequity project, addressing inherent biases in Automatic Speech Recognition systems. Developed a comprehensive testing framework with novel metamorphic transformations to simulate diverse environmental conditions. This project identified specific scenarios where ASR systems fail, proposing targeted enhancements to improve fairness, ultimately contributing to more robust and inclusive voice recognition technologies for diverse user demographics.
- Under the guidance of Dr. Christopher Brew, Led a project integrating LongFormer attention and question generation models to generate long-answer questions from academic texts. Evaluated the model using both automatic and manual methods, achieving contextual and semantic accuracy.

Bachelor of Technology, Computer Science and Engineering, GRIET, India July 2016 – May 2020

- Developed an automated system for crawling and analyzing e-commerce reviews. Utilized sophisticated sentiment analysis techniques to predict customer sentiment and designed novel summarization methods to generate concise, insightful product review summaries.
- Conducted predictive analysis on lung cancer patients to classify survival expectancy, enhancing the accuracy of prognoses. Developed an intuitive web application built on a robust classification model to predict survival expectancy categories.
- Developed a sophisticated web crawler, an intelligent internet bot designed to navigate and gather comprehensive data from websites. Engineered and implemented diverse data structures to optimize data storage and retrieval, significantly enhancing web information processing.
- Developed an application that seamlessly integrates paramedic drivers with the traffic police department, utilizing real-time data and intelligent routing algorithms to optimize transport time and ensure swift emergency response.

INDUSTRIAL EXPERIENCE

Data Science Intern – Enterprise Data Warehouse & Analytics, Ohio State University, Columbus, OH Jan 2024 – Dec 2024

- Supercharged data migration efficiency by 84.4%, slashing processing times from 6 hours to an average of 56 minutes by leveraging parallel loading strategies on the AWS Airflow (MWAA). Harnessing tailored ConnectorX source code, and precision-configuring resources for peak performance, scalability, and optimized resource allocation, ensuring seamless execution with minimal idle resources and maximum utilization.
- Built and optimized high-performing large scale backend systems and data integration pipelines (ETL/ELT) leveraging Python, Airflow, DBT, and AWS. This enhanced data warehousing and analytics, significantly streamlining data workflows and scaling processing capacity by 4.5 times.

Analytics Research Intern – Analytics Center of Excellence, OSUWMC, Columbus, OH May 2023 – Aug 2023

- Led a Data Quality initiative, designed and developed a Python-based, no-code data validation framework using greatexpectations.io, which radically streamlined workflows and transformed data validation from a laborious, manual ordeal (hours of tedious SQL query writing) into a lightning-fast, automated process, yielding a staggering 96% reduction in data validation time (from hours to mere seconds), accelerated project timelines by 30%.
- Developed and implemented natural language processing (NLP) solutions using fine-tuned large language models(Bio+Clinical BERT & LlaMa), successfully analyzing and synthesizing over 250K+ feedback entries from students, patients and doctors at the Wexner Medical Center. This approach significantly improved data-driven decision-making, directly contributing to a 16% improvement in program effectiveness and patient care satisfaction.
- Led the design and implementation of a system for automated schema migration using advanced DBMS tools and Python scripts, achieving a reduction in system downtime by 87%, and overall reliability. Collaborated closely with cross-functional teams, to ensure seamless integration and adoption.

Software Engineer III – Ninjacart, Bengaluru, India Jun 2022 – Dec 2022

- Spearheaded backend development for Ninjacart's B2B platform by leveraging Java Spring Boot and NodeJS for effective server-side execution. Streamlined service integration with Apicurio Registry to ensure API consistency and enhance microservice interoperability. Employed MongoDB for its dynamic schema capabilities, facilitating adaptable data storage, and Kafka for its high-performance data streaming, supporting real-time operational demands. The architecture was underpinned by the ELK stack, enabling detailed system monitoring and data-driven decision-making.
- Orchestrated a complex project that expanded our platform's functionality through meticulous API integrations with key partners such as CargoFL, ePayLater, ADAT, GIMBook, and TallyWiz, significantly boosting platform efficiency, operational capacity and partner satisfaction. This expansion resulted in a 28% surge in transaction volume within just two weeks, significantly outperforming initial efficiency and operational capacity targets.
- Managed the swift development, testing, and effective deployment of a comprehensive Point-Of-Sale system integration project, seamlessly coordinating with multiple external partners to optimize interoperability, resulting in a remarkable 38% increase in transaction efficiency across integrated systems.

Software Development Engineer II − Head Digital Works, Hyderabad, India Feb 2020 - Jun 2022

- Engineered 25+ high-impact backend features for a23.com and cricket.com, driving business growth within a fast-paced agile startup environment.
- Designed and Developed end-to-end backend systems, accommodating 80M+ total users and 600K+ active users, deployed on both on-premises servers and cloud environments (AWS), with a focus on scalability to handle large user bases and high traffic, high availability to ensure minimal downtime and maximum uptime, and real-time processing to enable fast and efficient data processing across various technology stacks and software products.
- Established reliable communication pathways employing RESTful APIs with Java Spring and MVC, complemented by RabbitMQ for message queuing. Deployed in containerized environments on AWS with Jenkins for CI/CD pipelines, ensuring rapid and reliable deployment of new features and updates.
- Implemented an analytics engine to analyze over 20M+ player profiles and behaviors, identifying fraud and enhancing game integrity through dynamic bans. Boosted engagement by 24% by introducing personalized tutorials and strategic coupon incentives, fostering a more immersive experience.
- Transformed organizational analytics by leading 3+ critical research projects in event-driven systems and predictive analytics, which were rapidly integrated by leadership, and had a substantial positive impact on the business. Led the recruitment of over 4+ interns specifically for these projects.

EDUCATION

Master of Science in Computer Science and Engineering − Ohio State University − CGPA: 3.48 Jan 2023 – Dec 2024
Coursework: Advanced Computer Architecture, High Performance Deep Learning, Foundations of Speech and Language Processing, Advanced Algorithms, NLP, AI

Bachelor of Technology in Computer Science and Engineering − GRIET − CGPA: 4 May 2020
Coursework: Problem-Solving, Object-Oriented Design, Data Structures & Algorithms, Databases, Operating Systems, Networking, Machine Learning

SKILLS

- **Languages** : Java, Python, C, C++, JavaScript, Typescript, SQL, Scala, R, C#, Bash
- **Web Technologies** : Spring Boot, Spring MVC, Node.js, Express.js, React.js, Angular.js, Flask, HTML5, CSS, JSON, XML, YAML
- **Tools & Frameworks** : Git, GitLab, Jira, Jenkins, Maven, Gradle, Docker, Kubernetes, JUnit, Selenium, Postman, Netty, RabbitMQ
- **Databases** : MySQL, PostgreSQL, Cassandra, MongoDB, DynamoDB, Elasticsearch, Neo4j, Redis, Prometheus
- **Big Data Technologies** : Apache Spark, MLlib, Airflow, Hadoop, Dask, Hive, Kafka, Flink, Nifi, AWS Glue, AWS Lambda, AWS DMS
- **Cloud Platforms** : Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)
- **Technical Proficiencies** : Backend Development, AI, Microservices, REST API, CI/CD, DevOps, Machine Learning, NLP, GenAI, Agile, SDLC

LEADERSHIP

- Founding member of the Advanced Academic Center at GRIET, a hub for academic innovation. Organized hackathons through X-Kernel.
- Served as Head Administrator at canuckgaming.net, managing a large player base, overseeing transactions, and ensuring a secure gaming environment.