

Ajay Kumar

www.ajay-kumar.com | linkedin.com/akumar | github.com/akumar | scholar.google.com/akumar

Summary

PhD student (Computer Science, expected May 2027, 4.0 GPA) with 5.5+ years in data science, distributed ML systems, big data processing, and LLMs. Achieved 2× multi-GPU pipeline optimization, developed knowledge graphs for genomics, and led NLP on large-scale social data. Built 8k+ member tech community.

Skills

Programming Languages : Python, R, SQL; **Generative AI:** LLMs, Agents, Multi-Agents; **Big Data:** Apache Hadoop, Spark, Hive; **Machine learning:** TensorFlow, PyTorch, Scikit-learn, Neural Networks, Deep Learning, Keras, forecasting, Knowledge Graphs and Extraction; **Data Analysis/Visualization:** Pandas, NumPy, SciPy, Matplotlib, Seaborn, Statistical modeling; **AWS Services:** AWS-SageMaker, Django for AWS, AWS Cloud practitioner **Web Development:** HTML, CSS, JavaScript, Django, Flask; **Soft skills:** Problem-solving, Team Collaboration, Technical Writing, Project Management.

Experience

Graduate Research Assistant, University of Missouri – Columbia, MO, USA Aug 2022 - Present

- Optimized a scalable distributed **multi-GPU algorithm** by **2X** with NVIDIA's parabricks pipeline for COVID19 human genome variant calling using Google's JobShop Scheduling algorithm (Published at **ACM-BCB Conference 2025** [click here](#)).
- Engineered a scalable tool (using, **Apache Spark, Hadoop, Hive, Pig, multi-cloud**) for Democratizing Variant Calling on Human Genomes Using **Commodity Clusters (FABRIC, CLOUDLAB, AWS)** (Demonstrated at **CIKM-2024**, [click here](#)).
- Developed a **VariantKG** tool for analyzing human variants (big data of **650-GB size**) using **Knowledge Graphs and Graph Machine Learning** (published in **Frontiers in Big Data-2023**, [click here](#))
- Designed a **multi-model machine learning (ML)** tool, **G2PDeep-v2** (**Python, ReachJS, Tensorflow, Keras and SQL,**) for healthcare dataset (submitted to **Journal of Genome Biology**) **Link: www.G2PDeep.org**.
- Performed **Sentiment Analysis** with **big data tools** of Covid19 Vaccine Twitter dataset with **fine-tuned BERT and RoBERTa Large Language Models (NLP/LLMs)**. **Link: click here**

Research Scientist, International Rice Research Institute – South Asia, IN Aug 2020 - 2022

- Developed a **multi-model deep learning (DL)** architecture, named **DeepMap** for regression/classification outperformed traditional **machine learning models** by **13-20%** (submitted research in **Genome Biology**) **Link: www.pypi.org/DeepMap**.
- Lead a strategic partnership between **Amazon Web Services (AWS)** and **International Rice Research Institute (IRRI)**, resulting in a **50% increase** in **computational resources** for genomic/bioinformatics researchers.
- Lead an international tech training for **100+ scientists/researchers** in advanced bioinformatics techniques, **Python basic to automation, and ML/DL model development to deployment**, resulting **40% improvement** in research efficiency across multiple institutions.
- Conducted machine learning analyses on agricultural datasets (**Flask, AWS, Pandas, Matplotlib, Seaborn, Neural-Networks**), producing **4-research papers** and attaining **200K+ views** on related Quora answers.

Software Engineer, Intern, Centre for Analysis And Systems Studies, SPA – New Delhi, IN June 2019 – August 2019

- Developed a government website for the Urban STeP project at the School of Planning and Architecture (SPA), New Delhi, using PHP, HTML, and JavaScript; attracted **15,000** visitors and increased local area planning participation by 30%.
- Designed user interfaces and implemented an efficient photo-tagging system for the “Biodiversity of City” Android application (Android Studio/Kotlin, Firebase); boosted user engagement by 35%, accelerated biodiversity data collection by 50%, and improved security by 25% through PyTest-identified fixes and enhanced data protection.

Key Achievements

- Honored with **\$2,500** ACM UPE Executive Council Award for leading international chapter operations as Vice President.
- Secured **\$5,500** competitive travel fellowship from FABRIC to present research and build global collaborations.
- Grew and actively manage a high-engagement **8,000+** member technical community driving industry-relevant discussions and knowledge sharing through **LinkedIn, GitHub, YouTube**.
- Published/co-published **7+ papers** in prestigious international venues, contributing to advancements in computing, distributed cloud computing and artificial intelligence (available at **google scholar**).

Education

University of Missouri-Columbia, USA, Ph.D. Computer Science Aug 2022 - Dec 2026

- **Thesis Title:** Optimized scientific workflows on the FABRIC national research testbed, advancing large-scale distributed computing capabilities (**Thesis Advisor:** Dr. Praveen Rao)
- **GPA: 4.0/4.0** (Previously completed **M.S. and, B.S.** in Computer Science)