Kunal Jain

kunal.jain@research.iiit.ac.in |+91 9315529391 | 🔿 kjain1810 | in kjain1810

PUBLICATIONS

- [1] Jain, S., **Kunal Jain**, Ravindran, A., and Purini, S. Seer: A framework for optimizing traffic camera placement and deep learning inference at the edge for vehicle path reconstruction. *9th Symposium on Edge Computing (SEC)*, 2024.
- [2] Khan^{*}, R. I. S., **Kunal Jain^{*}**, Shen, H., Mallick, A., Parayil, A., and et. al. Fairserve: Ensuring fair IIm serving amid diverse applications. *Under review* (2025).
- [3] Kunal Jain, Adapa, K. S., Grover, K., Sarvadevabhatla, R. K., and Purini, S. A cloudfog architecture for video analytics on large scale camera networks using semantic scene analysis. *Cluster, Cloud and Internet Computing (CCGrid) 2023.*
- [4] Kunal Jain, K.J., P., and Bodas, T. Bayesian optimization for function compositions with applications to dynamic pricing. 17th Conference on Learning and Intelligent Optimization (LION17), 2023.
- [5] **Kunal Jain**, Parayil, A., Mallick, A., and et. al. Intelligent router for Ilm workloads: Improving performance through workload-aware load balancing. *Under review* (2025).

RESEARCH

MICROSOFT RESEARCH | INTERN (JAN'24-JUN'24) | RESEARCH FELLOW (JUN'24 -)

- Worked on intelligent **routing methods** for large number of workloads between **multiple model instances**. Reduced the latency of request completion by identifying harmful workload mixes and training an **RL agent** to take optimal routing decisions
- Investigating ways to drop requests and **throttle users in a fair manner** when the number of requests exceeds available hardware capacity
- Working on acquisition and allocation of GPUs when serving 100s of fine-tuned LLM models in order to improve service availability and reduce hardware cost
- Working on new methods to host multiple fine-tuned SLMs on a single GPU cluster

COMPUTER SYSTEMS, IIIT-H | UNDERGRADUATE RESEARCHER | AUG'21 – JUN'23

- Built a scalable video processing **distributed system** for **large scale camera networks** which **deploy computer vision** pipelines with minimal overhead. Leveraging the system to develop a **distributed algorithm** for tracking vehicles in large cities
- Worked on cross domain applications of **Bayesian Optimisation** in areas such as **revenue management** and **supply chain optimization**. Explored various optimization scenarios like extremely noisy evaluations, composite objectives, switching cost, etc
- Worked on **Federated Learning** algorithms from a **game theoretic perspective** in order to understand the behaviour of participants in a competitive marketplace

EXPERIENCE

IIIT HYDERABAD | TEACHING ASSISTANT | NOV'21 - AUG'22, JAN'23 - MAY'23

• Conducted labs, tutorials and office hours and prepared and evaluated examinations for the following courses: (i) Computer Programming (~350 students), (ii) Data Structures and Algorithms (~320 students) and (iii) Advanced Algorithms (~50 students).

UBER | SOFTWARE ENGINEERING INTERN | SEP'23 - OCT'23 | HYDERABAD

• Worked with the **internal Risk team** on providing **explanations of actions** (based on specified rules) taken by the platform on users such as banning them. Accumulated the results on a single end-point and performed a **progressive backwards search** on (action, feature) pairings to **identify root causes**

DREAMVU | HARDWARE ACCELERATION INTERN | SEP²¹ – NOV²¹ | Hyderabad

• Implemented **GPU based accelerations** for computer vision algorithms using OpenCV which improved efficiency by **300%**. Used image processing tools such as **GStreamer** to **increase throughput** of image captured from cameras from 1 FPS to 28 FPS

EDUCATION

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY, HYDERABAD

B.Tech + MS (by research) in Computer Science 2019-2024

CGPA: 8.9/10

 DELHI PUBLIC SCHOOL, ROHINI

 SENIOR SECONDARY SCHOOL
 2017-2019

 X: 10/10 CGPA | XII: 94%
 XII: 94%

SKILLS

LANGUAGES AND SCRIPTING

C/C++ • Rust • Java • Python • GoLang • Bash

• Racket • LATEX MISCELLANEOUS

Linux • Git • Docker • Tensorflow • PyTorch • OpenCV • BoTorch • OpenMP • MySQL • MongoDB • Spark • Hadoop

ACHIEVEMENTS

• Swabs2Lab, CNI, IISc Bangalore - Winner of the optimisation track among 50 teams

• ICPC Regionalist (Kanpur region) - Ranked 15th out of 1200 teams

• Google HashCode - Ranked 28th out of 3000+ Indian teams

- Asia Pacific Informatics Olympiad, 2018 -Honorable mention while representing India
- INOI 2018, 2019 (National finalist) One of
- 30 finalist out of 1500 participants in India

• Dean's Merit List - Top 5% academic performance in college during the year of 2019-20

- JEE Mains & Adv. 2019 Top 0.2% percentile
- Dean's Research Award For extraordinary research work during undergrad

SERVICES

• Reviewer: ML and Compression Workshop@NeurIPS 2024

COURSEWORK

COMPUTER SCIENCE

Statistical Methods in Al • Advanced Algorithms • Game Theory • Compilers • Data Systems • Applied Optimization • OS and Networks • Distributed Systems • Blockchains MATHEMATICS

Discrete Structures • Real Analysis • Linear Algebra • Probability and Statistics • Dynamical Processes in Complex Networks