Pranav Jangir

pranavjangir.com | pj2251@nyu.edu | linkedin.com/in/pranavjangir | github.com/pranavjangir | Google Scholar

Experience

PlusAI **Sept. 2024 – Present**

Senior Robotics Engineer

- **Perception Performance**
  - Optimized the perception component's performance on dual NVIDIA Jetson Orin systems, reducing aggregate p99 latencies across the optimized engines by approximately **one-third**.
  - Built runtime telemetry and used fleet data to find and fix memory leaks that caused OOM failures.
- **ROS 2 Migration**
  - Lead the ROS 1 to ROS 2 Humble migration of the autonomy stack. Made the ROS 2 RMW layer work with the in-house networking, shared-memory, and memory allocator stack.
  - Compared ROS Middleware (RMW) configurations and used composable nodes to run GPU nodes in one process, reuse CUDA contexts, and improve performance over the ROS 1 stack.
  - Built end-to-end validation with replay and simulation regression tests and automated Nsight Systems profiling reports in CI.
- **Release Engineering**
  - Created and formalized the release pipeline used for all customer-facing releases on a weekly or biweekly cadence.
  - Reduced build time and disk usage for the full multi-platform build, and built GitHub automation to generate changelogs from internal pull-request templates.

Google **Aug. 2020 – Mar. 2022**

Software Engineer

- Improved ad quality and coverage for high-volume queries in low-resource languages (such as Hindi and Polish).
- Engineered a concurrent C++ ad-serving pipeline to process the newly added cross-language search-query signals, retrieve more ads, and remove irrelevant ads while limiting the increase in p99 end-to-end latency to less than 3 ms.
- Collaborated with the Google Research team to fine-tune query-translation models designed specifically to increase ad revenue.
- Ran numerous A/B tests and ad-quality studies to evaluate new query-translation models against production models.
- Increased annual revenue by **$300 million**.

Tower Research Capital **May 2019 – July 2019**

Software Engineering Intern

- Built a semi-automated machine-learning pipeline to extract transaction data from PDF files.
- Built a clustering tool that used self-organizing maps to group similar PDF lines and generate soft labels for training data.
- Identified transactions with **88% precision** and **96% recall**, reducing PDF processing time from 3 days to approximately **2 hours**.

Selected Publications[†]

Papers

1. R. Cole, A. Gupta, and **P. Jangir**, "Parsimonious Predictions for Strategyproof Scheduling," in Advances in Neural Information Processing Systems 38 (NeurIPS 2025), pp. 110564–110584, 2025. *doi*.
2. R. Cole and **P. Jangir**, "Distributed Interview Selection for Stable Matching in Large Random Markets," arXiv:2506.19345, 2025. Poster at the 26th ACM Conference on Economics and Computation (EC '25), Stanford University. *doi*.
3. **P. Jangir**, N. Koti, V. B. Kukkala, A. Patra, and B. R. Gopal, "Match Quest: Fast and Secure Pattern Matching," in Proceedings on Privacy Enhancing Technologies (PoPETs), vol. 2025, no. 4, pp. 308–328, 2025. *doi*.
4. **P. Jangir**, N. Koti, V. B. Kukkala, A. Patra, B. R. Gopal, and S. Sangal, "Vogue: Faster Computation of Private Heavy Hitters," in Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22), pp. 3371–3373, 2022. *doi*.

Under Submission

1. R. Cole and **P. Jangir**, "Facility Location with Public Locations and Private Doubly-Peaked Costs," arXiv:2505.18114, 2025. *doi*.

[†]The custom for theoretical computer science and mathematics is to list the authors in alphabetical order by last name.

Presentations

1. **Presentation at NeurIPS'25 (San Diego, CA):** Presented our work titled *"Parsimonious Predictions for Strategyproof Scheduling."*
2. **Poster Presentation at ACM EC'25 (Stanford University, CA):** Presented our work titled *"Distributed Interview Selection for Stable Matching in Large Random Markets."*
3. **Poster Presentation at ACM CCS'22 (Virtual):** Presented our work titled *"Vogue: Faster Computation of Private Heavy Hitters"* in the poster presentation workshop.

Education

NYU, Courant Institute of Mathematical Sciences **Sep. 2022 – May 2024**
Master of Science in Computer Science *Manhattan, NY*
Thesis: Mechanism Design with Predictions *Adviser: Richard Cole*

Indian Institute of Technology, Guwahati **Jul. 2016 – May 2020**
Bachelor of Technology in Mathematics and Computing *Guwahati, Assam*
Thesis: A Study on Partition Functions *Adviser: Rupam Barman*

Academic Research Projects

Algorithmic Game Theory **Jan. 2023 – Nov. 2025**
Prof. Richard Cole, Professor, Computer Science *NYU Courant*

- Studied an online scheduling mechanism for jobs on unrelated machines when the machines were strategic agents. The mechanism used predictions that were easier to predict than those in prior work and could also support strategic jobs. [Paper]
- Studied a novel facility-location variation with doubly peaked preferences and developed approximate truthful mechanisms without money. [Paper]

Fast Filter Subset Scan for Anomalous Pattern Detection **Mar. 2023 – May 2024**
Prof. Daniel Neill, Associate Professor, Computer Science, Machine Learning for Good Lab *NYU Courant*

- Designed and implemented methods for fast, scalable detection of anomalous patterns such as an emerging disease outbreak or smuggling activity in large multivariate datasets.
- Developed a subset-scanning technique that used prior information to scan large datasets efficiently and avoided exhaustive searches that could exceed time or memory limits.

Secure Machine Learning Algorithms **Dec. 2021 – Aug. 2022**
Dr. Arpita Patra, Associate Professor, Department of CSA *Indian Institute of Science*

- Designed an algorithm that solves the Private Heavy Hitters problem and beat the state of the art Poplar in running time by upto 6x. Poster accepted at ACM CCS'22 and full paper at IEEE TDSC. [Paper] [GitHub]
- Built CUDA-style C++ tensor kernels for privacy-preserving neural network inference, including GEMM, convolution, max pooling, and ReLU, with SIMD batching; implemented distributed zero-knowledge proof (ZKP) verification for multiplication triples. [GitHub]

Fractional Partition Functions **Aug. 2019 – April 2020**
Prof. Rupam Barman, Professor, Department of Mathematics *IIT Guwahati*

- Studied algebraic and combinatorial properties of fractional partition functions and their relation to the famous Ramanujan congruences.
- As a part of my bachelor's thesis, wrote a survey of the many congruences that can be proved using fractional partition functions and other open problems colloquially known as "Mao's conjectures". [Thesis]

Other Projects

Reinforcement Learning with Adversarial Agents | GitHub | Blog Post **Feb. 2024 – May 2024**

- Designed an algorithm that used behavior cloning and imitation learning to learn a policy from adversarial online expert advice.
- Used contrastive learning and variance-based regularization so the policy ignored harmful advice.

Benchmarking OpenMP vs Rayon | GitHub | Report **Sep. 2023 – Dec. 2023**

- Implemented matrix-multiplication algorithms to compare parallel execution in C with OpenMP and Rust with Rayon.
- Monitored page faults, cache references, and cache misses to analyze performance differences.

Future video frames prediction | GitHub | Paper | Presentation **Jan. 2023 – May 2023**

- Leveraged self-supervision techniques with the SimVP model for future segmentation prediction in videos. Achieved a 44% Jaccard similarity, improving over the baseline of 23%.
- Innovated a New Object Suppression (NOS) decision tree-based technique to rectify prediction errors.

Verkle Trees benchmarking | GitHub | Report **Sep. 2022 – Dec. 2022**

- Implemented Verkle, Patricia-Merkle, and Binary-Merkle in C++. Compared the proof sizes, proof generation time, and proof verification time for varying key sizes for all the three trees.
- Compared two vector commitment schemes KZG and IPA in our Verkle Trees across different widths.

Real time Bus Routing Algorithm | GitHub **Oct. 2019 – Dec. 2019**

- Developed a heuristic algorithm for the NP-Hard School Bus Routing Problem (SBRP) in dynamic edge weight graphs.
- Implemented metaheuristic Tabu search algorithm along with genetic algorithm to ensure that the solution converges quickly.

PintOS **Sep. 2018 – Nov. 2018**

- Implemented a multi-level feedback queue scheduler, system calls for user programs and virtual memory management module for an instructional operating system framework - PintOS

Technical Skills

Languages: C++14/17, C, Python, Bash, Groovy, SQL
Performance Engineering: CPU/GPU performance optimization, Data-Oriented Design, NUMA tuning, SIMD (AVX-512 / NEON), OpenMP
GPU / Machine Learning: CUDA, TensorRT, Nsight Systems, Nsight Compute, PyTorch, TensorFlow, OpenCV, ONNX, FP16 inference, sensor fusion
Robotics and Middleware: ROS 1, ROS 2 Humble, Fast DDS, DDS/RTPS, shared-memory IPC, CDR, NVIDIA Jetson Orin (aarch64)
Debugging and Profiling: GDB, perf, Google Benchmark, Valgrind
Build and Infrastructure: Linux, GCC, CMake, Make, Docker, Jenkins, CI/CD, GitHub Actions, Debian packaging, ccache, Slurm, AWS, Google Cloud Platform (GCP), GoogleTest (GTest), Git

Teaching and Mentoring

Roaring Cubs Collective (RCC): RCC is a STEM research mentorship program by Columbia University's students. I mentored three NYC high school students in conducting Machine Learning research.
NYU: Teaching assistant for undergraduate courses on Theory of Computation, Security and Cryptography and Mathematical Statistics.
Google: Mentored a summer intern in the Ads Research team and helped them land a full-time offer.

Awards and Achievements

Thesis Award: Awarded the best Master's thesis award in the Computer Science Department at NYU.
William A. Chittenden Fellowship: Competitive fellowship awarded to only one incoming PhD student every year.
Jerry S. Dobrovolny Fellowship: Fellowship supporting graduate study at UIUC.
UMass Amherst CICS Scholarship: Scholarship supporting graduate study at UMass Amherst.
Oxford Indira Gandhi Graduate Scholarship: Selected for the scholarship to support proposed DPhil research in computational fairness.
Google: Awarded 2 spot bonuses & 3 peer bonuses for leading the project and helping engineers onboard.
Google Kick Start: Ranked **36**$^{\text{th}}$ globally (out of ≈4k participants) in Google Kick Start Round F 2019 and **58**$^{\text{th}}$ globally in round G 2019.