Rajni Pawar

+1(312) 863-9852 | rajnippawar20@gmail.com | linkedin.com/in/raznaee | My Website | github.com/rajnipawar|

Education

- **Illinois Institute of Technology** — Chicago, IL
  Doctorate of Philosophy in Computer Science — GPA: 3.83 — *Aug, 2024 - Present*
  Relevant Coursework: Advanced OS, Algorithms, ML, Computer Networks
- **University of Mumbai** — Mumbai, India
  Bachelor of Engineering - Electronics and Telecommunication Engineering | CGPA: 9.07/10 — *June, 2021*

Research Experience

- **Gnosis Research Center, Illinois Institute of Technology** — Chicago, IL, USA
  Graduate Research Assistant - Dr. Xian-He Sun — *Jan 2025 - Present*

Accelerating GPU I/O

- GPUs dominate data workloads, yet lack optimized I/O libraries.
- Contributing to project **IOWarp**, a data management platform that aims to **accelerate data-intensive workflows**.
- Researching **NVIDIA GPUDirect Storage** (GDS) to explore GPU-centric I/O stacks on cloud GPUs (AWS, Chameleon, GCP) and optimize NVMe–GPU DMA and NUMA-aware data paths.
- GDS shows **60% lower latency** than traditional CPU based data movement.

Enhanced Scientific Exploration with AI

- Scientific workflows are cumbersome, involving numerous applications and manual coordination to achieve insight.
- **Fine-tuned LLMs** like granite, ph-4-mini-reasoning, etc, to support **deployment tool under IOWarp, Jarvis**.
- Leveraged **Unsloth, LoRA and GPUs** for faster computation of training/inference achieving **95% accuracy**.
- Built a **Model Context Protocol(MCP)** server integrating AI tools with the **National Data Platform**, enabling researchers to access federated scientific datasets seamlessly, **accelerating AI-driven research workflows by 40%.**

An Active Tokenizing I/O Stack

- Scientific programs generate massive data for AI workloads, where synchronous tokenization adds significant overhead.
- Developed **tokenization-integrated I/O stack** storing **ready-to-use tokens for LLMs & RAG** workloads.
- Accelerated AI pipelines by 57.1% eliminating preprocessing & validating scalability on **HPC with MPI & SLURM**

Work Experience

- **Tata Consultancy Services** — Mumbai, India
  System Engineer — *June 2021 - July 2024*
  - Implemented **high-performance computing workflows** for complex data analysis and model training/inference using **C++** and **CUDA**, optimizing for **GPU-accelerated processing**, and scheduling jobs with **SLURM** and **LSF**, **reducing computational time by 50%**.
  - Engineered an enterprise-grade website, developing **50+ scalable AEM components** to support **dynamic content management**. **Collaborated with UI/UX designers** to create responsive front-end interfaces using **HTML, CSS, and Javascript**, resulting in a **30% increase in user metrics.**
  - Integrated AEM components using **Java**, leveraging **Jenkins CI/CD** for automated deployments, with **Bitbucket** and **JIRA** for project tracking to enhance functionality and support **5M+ users.**

Other Projects

- **SimpleChat - Distributed Messaging Application**: Developed a multi-node real-time messaging system using C++ and Qt6 with TCP socket programming, implementing ring network topology for automated message routing between interconnected nodes and modern GUI with conversation management.
- **DevPlate: Stylized Devanagari Script License Plate Conversion to Standard Script using Deep Learning**: Engineered a recognition system of license plate numbers featuring an Indian script and converting them into standard scripts to aid law enforcement authorities using the **KNN algorithm, OpenCV, and EasyOCR** with an **accuracy of 85%**.

Achievements & Publications

- Presented research poster at the **SSDBM 2025** Conference on **Scaling Data Tokenization for AI Systems**.
- M. Kolhekar, S. Kurle, R. Pawar, J. S. Kumar and R. Verma, "**Stylized Devanagari Script License Plate Conversion to Standard Script using Deep Learning**," (**ICAST**), 2023, pp. 143-148, doi: 10.1109/ICAST59062.2023.10454944
- **AWS Certified Solutions Architect - Associate**
- **Microsoft Certified: Azure Fundamentals**
- **Won TCS Ideathon** in 2021-2022 for the Smart Package Manager project.

Skills

- **Programming**: **Python, MPI, OpenMP, CUDA, C/C++**, Java, SQL, Bash, HTL
- **Systems**: **Distributed, HPC, SLURM, Microservices,** Algorithms, Linux, Sharding, Networks
- **AI/ML**: PyTorch, Tensorflow, Pandas, SKlearn, LLMs, LangChain, RAG, LLM fine-tuning, MCPs
- **Cloud/Devops**: **Google Cloud/Colab, Chameleon Cloud**, AWS, Azure, Docker, Kubernetes, CI/CD, Agile, Git, Jenkins