

## ACADEMIC DETAILS

Year	Degree	Institute	CGPA/Percentage
2018-2022	B.Tech in Computer Science and Engineering	Indian Institute of Technology Delhi	8.226/10
2018	Class XII, CBSE	Keshav Saraswati Vidyamandir, Patna	89.6%
2016	Class X, CBSE	DAV Public School, Patna	10/10

## SCHOLASTIC ACHIEVEMENTS

- Received the annual **Arjuna Award** for outstanding performance during my first year at ThirdAI.
- Secured **All India Rank 646** in Joint Entrance Exam Advanced - 2018 among 225,000 candidates
- Awarded **KVPY Fellowship** by Dept. of Science & Technology, Govt. of India
- Recipient of **National Talent Search Examination(NTSE)** Scholarship by NCERT, New Delhi
- Qualified **Regional Mathematics Olympiad(RMO)** twice in 2015-16 and 2016-17 organized by HBCSE

## PUBLICATIONS

- N. Meisburger, V. Lakshman, B. Geordie, J. Engels, D. Torres Ramos, **P. Pranav**, B. Coleman, B. Meisburger, S. Gupta, Y. Adunukota, S. Jain, T. Medini, A. Shrivastava, "BOLT: An Automated Deep Learning Framework for Training and Deploying Large-Scale Search and Recommendation Models on Commodity CPU Hardware," *Accepted at ACM International Conference on Information and Knowledge Management*, 2023. [\[Link\]](#)
- A. Shrivastava, V. Lakshman, T. Medini, N. Meisburger, J. Engels, D. Torres Ramos, B. Geordie, **P. Pranav**, S. Gupta, Y. Adunukota, S. Jain, "From Research to Production: Towards Scalable and Sustainable Neural Recommendation Models on Commodity CPU Hardware," *Accepted at ACM Conference on Recommender Systems*, 2023. [\[Link\]](#)

## WORK EXPERIENCE

Senior Software Engineer, ThirdAI  
AI Engineer, ThirdAI

July 2024 - Present  
May 2022 - June 2024

- Led Infrastructure Development for Enterprise Search Platform:** Designed and implemented a high-availability infrastructure using Nomad and Consul, orchestrating backend, frontend, and Traefik services. Scaled to a 25-node cluster with automated health checks to ensure continuous uptime. Successfully deployed to multiple customers, demonstrating reliability and scalability in production.
- Implemented Distributed Training Framework for ThirdAI's BoLT:** Led the development of the distributed training framework for ThirdAI's BoLT, enabling large-scale training of models like DLRM on CPUs. This foundational work, highlighted in [this article](#), significantly enhanced the company's ability to train billion-parameter neural networks on commodity hardware, setting the stage for subsequent innovations.
- Core Development of Enterprise Search Platform:** Developed an Ansible-based installation solution for single-command deployment across clusters, reducing setup complexity and enhancing scalability. Integrated Keycloak for secure Single Sign-On (SSO), improving access control and user management. Maintained and optimized backend services to ensure reliability, performance, and scalability in production environments.
- Led Development of ThirdAI's LLM and Bolt2.5B:** Spearheaded the end-to-end development of ThirdAI's LLMs, from initial architecture to large-scale training and fine-tuning. This included the successful training of **Bolt2.5B**, a generative model on par with GPT-2 but uniquely optimized for CPUs. The model's deployment and then worked on a Llama-7B scale model were highlighted [here](#), marking a significant milestone in the organizations AI capabilities.
- Maintainer of ThirdAI's Internal ML Library:** Actively contributed to ThirdAI's internal machine learning library, collaborating closely with clients to troubleshoot complex engineering issues. This role included managing maintaining software quality and ensuring the robustness of the ML frameworks through rigorous testing and optimization.
- Mentoring and Onboarding of New Team Members:** Mentored and guided four new hires, helping them transition smoothly into the development process. This involved providing technical mentorship, resolving their development challenges, and collaborating with them to create scalable solutions.

## INTERNSHIP

Summer Research Intern, University of Edinburgh  
Under Prof. Antonio Barbalace, School of Informatics

April 2020 - July 2020

- Implemented a messaging layer for multikernel communication using IVSHMEM and deployed it on QEMU.

- Utilized arm and x86 kernels and Popcorn Linux virtual machine images for deployment and testing.
- Studied interrupt transportation in a multi-kernel environment on QEMU to reduce latency.

## Software Solution Intern, Samsung Electronics

June 2021 - July 2021

### Graphics Lab

- Designed an Augmented Reality-based PacMan game in Unreal Engine using ARCore.
- Developed an algorithm for calculating the working plane using ARCore Traced Planes, added movement for agents and ghosts in the game, and worked on world-building.

## COURSE PROJECTS

---

### Portfolio Management System

Prof. Abhilash Jindal, Jan 2022 - Mar 2022

- Made an web-application, which connects the popular crypto exchanges, stocks, download and store data for popular stocks
- Added support for visualizing historical data, their current portfolio, getting top quotes and perform analytical operations on the data
- Used data processing frameworks like Kafka, Spark, Redis and Cassandra and Node.js for backend

### Simulating vehicles movements using Bayes and Kalman filter

Prof. Rohan Paul, Feb 2021 - Mar 2021

- Implemented HMMs for filtering, smoothing and predicting the motion of an aerial vehicle in a simulated environment
- Implemented Kalman filter for estimating aerial motion using error-prone sensor data
- Used Hungarian Algorithm as data association strategy for filtering sensor data for multiple aerial vehicles observed simultaneously

### Learning Optimal policy using MDP and Q-Learning

Prof. Rohan Paul, Apr 2021 - May 2021

- Implemented Value Iteration and Q-Learning algorithm for a mobile robot movement to estimate optimal policy for its movement towards a goal state
- Observed dependence of discount factor, iterations on Value iteration and exploration parameter on Q-Learning algorithms

### User-Level Thread Library

Prof. Kolin Paul, Mar 2021 - Apr 2021

- Implemented a user-level thread library, which provides all the important APIs provided by pthread library of C
- Implemented locks and conditional variables for threads using compare and swap
- Tested the working of library by implementing matrix-multiplication and producer-consumer programs on top of it

### Metro Route Display

Prof. Anshul Kumar, Sep 2019 - Oct 2019

- Displayed bus route between two user-specified locations on a CRT display with the help of a FPGA board
- Used Dijkstra algorithm to find the shortest path between two locations of the city

## TECHNICAL SKILLS

---

**Languages:** Python, C++, C, Go, Java, OCaml, Prolog, Bash Scripting

**Software:** Docker, FastAPI, Keycloak, PostgreSQL, SQLite, AWS, Azure, Google Cloud, Traefik, HashiCorp Nomad, HashiCorp Consul, HashiCorp Vault

**Frameworks/Libraries:** PyTorch, Ansible, Ray, GitHub, Hugging Face, OpenAI, Gloo

## RELEVANT COURSES

---

- **Computer Science:** Special Topics in AI: Planning and Estimation for Autonomous Systems, Introduction to AI, Machine Learning, Operating System, Cloud computing technology fundamentals, Parallel Programming and Distributed Systems, Introduction to Database Management Systems, Deep learning For mechanics, Analysis & Design of Algorithms, Computer Networks, Data Structures & Algorithms, Discrete Mathematics, Digital Logic & System Design, Programming Languages, Theory of Computation, Computer Architecture, Design Practices
- **Mathematics and Electrical Engineering:** Probability & Stochastic Processes, Graph Theory, Signals and Systems , Introduction to Electrical Engineering, Linear Algebra, Calculus

## EXTRA CURRICULAR ACTIVITIES

---

- Member of the **Algorithm and Coding Club**, IIT Delhi
- Taught children of orphanage under the program **Apna Parivar** of NSS IIT Delhi