

# Pawan Jayakumar

[github](#) [Website](#) [email](#)

## EDUCATION

---

### University of California San Diego

*Master of Science in Computer Science*

Sept 2024 - Dec 2025

*GPA: 4.0/4.0*

### University of Virginia

*Bachelor of Science in Computer Science*

Aug 2020 - May 2024

*GPA: 3.83/4.0*

## COURSEWORK

Software Engineering, Data Structures and Algorithm Design, Operating Systems, Machine Learning, Parallel Processing, Reinforcement Learning, Hardware Accelerators, Robotics, Probability, Linear Algebra

## EXPERIENCE

---

### Pytorch | *Open Source Software Engineer*

May 2024 - Present

- Actively engaged in the development of [Torchao](#), a library for performing architecture optimization for AI model inference and training by opening issues including reproducible code snippets with bugs, performing code reviews, and updating documentation
- Developed bit-packing algorithms to reduce memory cost of sub-byte quantized network weights and output anywhere between 2-4x
- Created a general sub-byte integer data type for low bit quantization using tensor sub-classing and bitpacking.
- LLama2 (large language model) quantized with int4 achieved 15% speed up and 3.5x memory savings over FP16
- Currently implementing Activation-aware Weight Quantization (AWQ) which is used by over 3400 models on huggingface

### Capital One | *Software Engineering Intern*

June 2023 - Aug 2023

- Designed and deployed a full stack cloud application using React, GraphQL, and AWS dynamo DB, which is used by over 15,000 monthly associates
- Optimized the team's local development environment, saving 100+ hours of development time

### Capital One | *Software Engineering Intern*

Jun 2022 - Aug 2022

- Designed and deployed a full stack cloud application to track and display changes in vulnerability reports to Capital One associates using Angular, and a variety of AWS database management services
- Negotiated with product team, presented design choices which would improve customer experience, performed code reviews, and pro-actively asked for feedback

### University of Virginia | *Teaching Assistant*

Aug 2022 - Dec 2022

- Led 100+ students in laboratory sessions and office hours by conducting code reviews and peer mentoring

## PROJECTS

---

### Policy Evaluation Benchmark

Feb 2024 - August 2024

- Constructed a testing harness for policy evaluation algorithms such as [ROS](#) and [BPS](#)
- Utilized [Slurm](#) job scheduler and [Weights and Biases](#) to parallelize model training and inference on a distributed system

### Slider

Mar 2022 - Mar 2023

- Co-developed and published an [award winning](#) puzzle game called [Slider](#) which has over 4000 unique players

### Meta Data Analytics Case Competition Finalist

Nov 2023

- Developed a product deployment strategy by researching Netflix media consumption data using machine learning and data science techniques such as Principle Component Analysis

## SKILLS

---

**Languages:** Python, C/C++, CUDA, SQL, C#, JavaScript, HTML/CSS

**Tools:** Git, Jenkins DevOps, Docker, Unix, Node, AWS Lambda, AWS DynamoDB, Copilot

**Frameworks:** PyTorch, React, Tensorflow, Angular, Rest, GraphQL, Tailwind