Vineet Sharma

vineetds@mit.edu | vinnysharma.dev | github.com/VinnySha | linkedin.com/in/vineetds

Education

Massachusetts Institute of Technology — *BS, Computer Science and Engineering* Expected: May 2028

- **Relevant Coursework:** Algorithms, Software Construction, Systems, Deep Learning, Linear Algebra, Probability
- **Extracurricular:** HackMIT Organizing Team, MIT Entrepreneurship Club, FTC Robotics

Experience

HackMIT Mar 2026 – Present

DevOps Head | Full-Stack Systems, Cloud Infra/Deployment (AWS), Team Leadership Cambridge, MA

- Leading development for internal AND external tools of **HackMIT (1,000+ hacker-event)**.
- Managing and mentoring a team of **15 developers**, setting technical direction, code standards, and milestones.
- Designing and delpoying systems across **8 projects**, including registration, team formation, and project submission.

dSideAI (CSAIL-Affiliated Startup) May 2025 – Sept 2025

Software Engineer Intern | React, Node/Express, MongoDB, Docker Remote

- Built and shipped fintech MVP in **12 wks**, owning core backend APIs and frontend flows.
- Implemented **trade compliance + risk guardrails** using **5 news APIs**, reducing high-risk recommendations.
- Designed portfolio generation + backtesting pipeline (LLM-assisted analysis), achieving **+5%** vs benchmarks.
- Presented product to enterprise stakeholders (large banks, Big Tech corp dev/M&A), driving weekly iteration.

MIT EECS – 6.1010 Fundamentals of Programming Jan 2025 – May 2025

Lab Assistant | Python Cambridge, MA

- Mentored **10+ students/week** during lab hours on Python fundamentals, debugging, and problem decomposition.
- Supported staff by refining **12 weekly assignments**, catching edge cases and improving autograder correctness.

Stacked Circuits (MIT Sandbox Funded Startup) Jan 2025 – Apr 2025

Software Engineer Intern | React, Node.js, PostgreSQL, Figma Cambridge, MA

- Built full-stack e-commerce prototype for PCB manufacturing, including Gerber uploads and order workflows.
- Produced **10 Figma iterations** + implemented final flows, improving checkout clarity and reducing friction.
- Evaluated build-vs-buy decisions across platforms and APIs to minimize cost while preserving scalability.

Projects

AppReader (HackMIT Application Review Platform) | *React, Flask, PostgreSQL, JWT, Tailwind*

- Built a full-stack application review platform to process **1000s of applications** across **35 concurrent reviewers**.
- Owned backend and database design, replacing a legacy schema with scalable PostgreSQL + JSONB architecture.
- Designed 15-min TTL reviewer claims using **PostgreSQL advisory locks** to prevent double-assignment errors.
- Built transactional CSV ingestion with flexible column mapping for evolving application schemas.

Compute-Aware Hybrid Attention Architecture Search | *PyTorch, CUDA, LLM Distillation*

- Designed a layer-wise method to selectively replace softmax attention with linear attention in LLMs, reducing attention complexity from $O(n^2)$ to $O(n)$.
- Trained linear attention replacements across **28 transformer layers** using knowledge distillation on ∼**40M** tokens, ranking layers by normalized distillation loss.
- Built hybrids achieving **5.2**× token throughput and **27**× faster time-to-first-token versus full-attention baselines.

CardioSense, Hacking Medicine 2025 (3rd Place, $1,000 Prize) | *Python, React, Next.js, OpenAI API*

- Built communication layer for remote cardiac monitoring, enabling clinicians to track patient vitals in real time.
- Integrated wearable data with GPT-powered summaries, providing alerts and care recommendations automatically.

Technical Skills

Python, Java, C, Typescript, React, Node.js, Express, RISC-V Assembly, PostgreSQL, MongoDB, PyTorch, Git, Docker