

Vinayaka Hegde

Portfolio | +1(979)721-2288 | vinayakah@tamu.edu | [linkedin.com/in/vinayaka2000](https://www.linkedin.com/in/vinayaka2000)

Education

- Texas A&M University** College Station, Texas
Master's in Computer Science Aug 2024 – May 2026
Courses: Software Engineering, Advanced Algorithms, Artificial Intelligence
- PES University, Bangalore, India** Aug 2017 – May 2021
Bachelor of Technology in Computer Science & Engineering GPA: **3.78/4.0**
Courses: Cloud Computing, Big Data, Operating Systems, Computer Architecture, Deep Learning
Awards: Dr CNR Rao Scholarship Award (Top 5% among 900+ students)

Skills

- Programming Languages:** Python, Javascript, Java, Ruby, C++, HTML, CSS, SQL
- Frameworks & Libraries:** Flask, React.js, PyTorch, NLTK, pandas, numpy, scikit-learn
- Cloud & Tools:** Apache Hadoop, Git, Docker, AWS, Postman, Heroku, Github Actions

Work Experience

- PES Labs** Bangalore, India
Software Engineer: ML Team, Project: Deploying ML workloads on Nvidia Jetsons Jan 2023 - July 2024
 - Implemented object detection models such as Mobilenetv3, Resnet50, YOLOv8, and deployed them using **Docker**
 - Achieved **up to 3x** reduction in inference latency by leveraging the **TensorRT** framework and offloading selective model layers to the Deep Learning Accelerator for optimized execution
 - Reduced **model complexity by 12%** by leveraging PyTorch's qnnpack engine to lower the precision of weights via **INT8 quantization**
 - Developed a tool using PyTorch's runtime dispatcher to compute the Floating Point Operations (FLOPs) of models, resulting in a **47% reduction** in the FLOPs after applying the above optimization methods
- Leadsquared** Bangalore, India
Software Development Engineer I, Integrations Team July 2021 - Dec 2022
 - Migrated backend **RESTful APIs** from on-premises to **AWS EC2**, reducing infrastructure costs by **Rs. 92K**, and **improving performance by about ~ 30%**
 - Optimized long-running and bulk-update tasks by transitioning to **Python**-based batch jobs, **reducing execution time by 23%**
 - Orchestrated **Adobe E-Sign API integration** for Mercer, creating a customized workflow that automated document signing
- Hewlett Packard Enterprise** Bangalore, India
Software Developer Intern, Project: Backend for a cloud-based application Jan 2020 - July 2020
 - Implemented a **scalable**, fault-tolerant Database-as-a-Service(DBaaS) for the backend of a cloud-based application
 - Zookeeper ensured **high availability and fault tolerance**, and RabbitMQ streamlined communication across nodes
 - Utilized nginx as reverse-proxy, and configured an application load-balancer, achieving **15% increase in throughput**

Selected Projects

- SQL Engine for vast datasets** (MapReduce, Hadoop Distributed File System, Apache)
 - Engineered a map-reduce based SQL engine, similar to Hive, capable of query parsing, selection, and aggregation
 - The mapper filters rows based on the WHERE clause, and reducer aggregates results using an aggregate function
- Enhanced Neural Machine Translation using Attention** (Autoencoders, LSTM, RNNs)
 - Optimized a translation model, achieving **97% accuracy** with attention, improving upon the baseline of 94%
 - Employed LSTM-based sequence-to-sequence architecture with dropout regularization
- MERN Stack Blogging App** (Node.js, React.js, Express, MongoDB, Github Actions, JWT)
 - Developed a **Node.js/Express** backend for a blogging app, integrating **JWT authentication** and **MongoDB data models**, with automated deployment via **GitHub Actions**
 - Architected the frontend using **React components**, and designed custom hooks that resulted in **28% reduction** in page load times

Selected Publications

- [1] *Performance Characterization of Containerized DNN Training and Inference on Edge Accelerators*, **HiPC 2023**
- [2] *Towards Efficient Scheduling of Concurrent DNN Training and Inferencing on Accelerated Edges*, **CCGrid 2023**