

YASH SARDA

Email: ysarda9@gmail.com | Github: github.com/ysarda
Linkedin: linkedin.com/in/ysarda | Phone: +1 (512) 825 6578

TECHNICAL SKILLS

Machine Learning & AI: PyTorch, Tensorflow, RLLib, Transformers, Reinforcement Learning, NLP

Programming Languages: Python, C#, MATLAB, Fortran 90, C++

Cloud & Devops: AWS (ECS, Inferentia, Lambda), Kubernetes, Docker, Grafana, Prometheus

Databases & Tools: CUDA, NumPy, OpenCV, ANSYS Fluent, Solidworks

Languages: English (Native), Hindi (Native), Mandarin (Conversational), Spanish (Basic)

WORK EXPERIENCE

Software Engineer | Portalis AI

July '23 – Oct '25

- Created conversational LLM interfaces for Gemini AI, Together AI, and Ohai AI.
- Redesigned Whisper speech-to-text architecture for deployment on AWS Inferentia, reducing STT price by 62%
- Deployed load testing lambda for peak and sustained performance review of deployed GPU service
- Designed async CUDA microservice for processing text embedding, speech-to-text, and semantic ranking requests
- Deployed various containerized architectures, including the CUDA inference service, to AWS ECS
- Added metrics, logging, and traces to all requests and tokens, allowing C-suite to predict product costing schemes
- Created visibility dashboards for production and development environments with Grafana and Prometheus

Machine Learning Engineer II | Shield AI/Heron Systems

Mar '21 – Aug '22

- Designed and implemented multi-head order-invariant transformer for metric projection of StarCraft 2 battles
- Led integration of existing code library and machine learning models with DOD combat simulator NGTS
- Optimized NGTS classifier with over 92% accuracy for aerial engagement win probability and attrition
- Trained multiple reinforcement learning agents for over 1 million timesteps using PyTorch and RLLib
- Co-led intern project for reinforcement learning in game environments, delegating tasks and resources
- Designed and created 10 multi-layer stackable terminals for customizable agent environments

Advanced Air Mobility Intern | NASA Langley Research Center

Aug '20 – Dec '20

- Designed and ran computational fluid dynamics simulations of deflected wing slipstream on Langley K-cluster
- Modified existing scripts to iteratively solve for wing geometry and export as mesh file

EDUCATION

Georgia Institute of Technology | MS, Computer Science

Aug '24 - Present

GPA: 4.00

The University of Texas at Austin | BS, Aerospace Engineering

Aug '17 - Dec '20

GPA: 3.53

Certificate: Computational Science, Engineering, and Mathematics

PROJECTS/RESEARCH

Undergraduate RA | Computational Astronautics Science and Tech, UT Austin

Oct '17 – Feb '21

- Designed and tested a recurrent convolutional neural network for classification and object detection
- Integrated a kinematic back-propagation algorithm with object detection for localization and tracking
- Created a hand-assembled database of “falls” based on NASA records and eyewitness reports