Avinash Amballa

6095054919 | amballaavinash@gmail.com | amballaavinash.github.io | www.linkedin.com/in/avinashamballa

EDUCATION

University of Massachusetts Amherst, USA — Aug 2023 - May 2025

Master of Science in Computer Science (with Specialization in Data Science) — **CGPA:4.0/4.0**

Relevant coursework: Reinforcement Learning, Responsible Artificial Intelligence, Advanced Natural Language Processing, Intelligent Visual Computing, Applied Statistics

Indian Institute of Technology Hyderabad (IIT-H), India — Jul 2017 - June 2021

Bachelor of Technology in Electrical Engineering with minor in Computer Science — **CGPA:8.8/10.0**

Relevant coursework: Data Structures, Algorithms, DBMS, Machine learning, Representation Learning, Linear Algebra

WORK EXPERIENCE

Google, Graduate Student Researcher — Feb 2024 – May 2024

Technologies: Python, Pytorch, numpy, HuggingFace, GPU

- Experimenting arithmetic sampling (strategy to sample diverse sequences in parallel from **Large Language Models**) with Chain of Thought self-consistency and MBR decoding strategies for generating diverse candidates.
- Incorporating more diverse measures of sequence similarity in sampling space using ideas from box embeddings.

Bosch (AIShield), Senior Research Scientist — Aug 2021 – July 2023

Technologies: Python, Tensorflow, Pytorch, scikit-learn, Azure, AWS, Docker, Git, SQL, DevOps

- Spearheaded research in **responsible AI**, focusing on vulnerability assessment, robustness, interpretability, fairness, causality and drift detection across computer vision, time series, speech and language models.
- Led AI security research, developing novel attack and defense strategies for adversarial, poisoning, model extraction, and inference attacks. Resulted in **1 published paper** and **4 filed patents**.
- Played a pivotal role in the early phases of securing **LLMs** by focusing on LLM alignment and analyzing jailbreaking attacks, laying the foundation for developing AIShield Guardian application.
- Established partnerships with key players in healthcare, financial, and MLOps sectors including Databricks, and Whylabs to enhance security and reliability of AI models, yielding a **revenue surge of around 10%**.
- Transitioned research insights into product features by developing microservices, pipelines, and logging infrastructure across Azure & AWS, accounting for **30% of overall workload**.

GE Digital, Software Development Intern — May 2020 – July 2020

Technologies: Python, Tensorflow, pandas, Flask, ReactJS, JavaScript

- Enhanced **web translation** application by migrating existing pipelines based on XML and JSON to a fine tuned encoder-decoder T5 Transformer on the XML and JSON data.
- Deployed scalable REST APIs with **Flask**, and integrated with frontend web interface built on **React** to demonstrate web translation functionality.

TECHNICAL SKILLS

Languages	Python, C, C++, Java, R, SQL
AI/ML	PyTorch, TensorFlow, Keras, scikit-learn, numpy, pandas, OpenCV, openAI gym, NLTK
Web Dev	HTML, CSS, JavaScript, React, jQuery, Node.js, Express.js, flask
Misc.	Data visualization, Big data analytics, Azure, AWS, Docker, Git, PostgreSQL, Elasticsearch

PUBLICATIONS & PREPRINTS

[1] Govindarajulu, Y., **Amballa, A.,** Kulkarni, P., & Parmar, M. (2023). Targeted Attacks on Time Series Forecasting. arXiv preprint arXiv:2301.11544.

[2] Amballa, A., Sasmal, P., & Channappayya, S. (2022). Discrete Control in Real-World Driving Environments using Deep Reinforcement Learning. arXiv preprint arXiv:2211.15920.

[3] Amballa, A., Mekala, A., Akkinapalli, G., Madine, M., Yarrabolu, N. P. P., & Grabowicz, P. A. (2024). Automated Model Selection for Tabular Data. arXiv preprint arXiv:2401.00961.

ACADEMIC PROJECTS

Optimization in Reinforcement Learning (UMass) Sep 2024 - Nov 2024

- Programmed Reinforce with baseline, one step Actor Critic, Episodic Semi Gradient SARSA, episodic Semi Gradient n-step SARSA and Tabular Dyna-Q algorithms from scratch.
- Optimized and evaluated RL algorithms on Acrobat, Cartpole and deterministic Grid World environments.
- Attained superior performance on Cartpole and Acrobat using Reinforce with baseline and Actor Critic methods.

Gyro Correction in IMU sensors (IITH, DRDO India) Apr 2021 - Jul 2021

- Spearheaded the creation of a gyro correction model for IMU sensors to mitigate noise and axis misalignment issues.
- Leveraged diverse architectural approaches, including DB-LSTM**,** LSTM with attention mechanism**,** and Transformer Encoder. Trained models on EUROC dataset with Huber Loss.
- Achieved superior performance (low validation and test loss) with attention-based models (Transformers), surpassing the capabilities of current work on Dilated CNN's through hyperparameter optimization.

ViCaP: VIdeo Captioning And Prediction (IITH) Sep 2020 - Dec 2020

- Implemented a vision-language video captioning method utilizing VGG16 feature extraction network with attention based encoder-decoder LSTM model. Trained the model on MSVD dataset.
- Achieved a higher BLEU score compared to a baseline model with custom CNN and LSTM, reflecting model's alignment between generated and reference captions.
- Predicted the missing video frames through conditional generative modeling. Investigating self-supervised learning.

AlphaConnect-4 (IITH) Jan 2020 - Apr 2020

- Inspired by deep mind's AlphaGo, applied competitive multi-agent Reinforcement Learning on connect-4 game.
- Utilized a combination of Monte Carlo Tree Search (MCTS**)** for opponent modeling and Actor Critic for agent reinforcement on this zero-sum mini-max game. Designed connect-4 environment in python.
- Visualized mean reward and standard deviation across training iterations, showing an increasing learning curve.
- Fine-tuned the learnt connect-4 agent on connect-5 game to improve its performance with minimal additional training.

PATENTS

[1] A method to detect poisoning of an AI Model and a System thereof. IN Patent App. 202241068482

[2] A method of Targeted Attack on Time Series Models to alter the DIRECTION IN Patent App. 202241065028

[3] A method of Targeted Attack on Time Series Models to alter the MAGNITUDE IN Patent App. 202241065034

[4] A method of Sponge attack on Deep Learning Models to increase the inference time IN Patent App. 202441006640

SERVICE

- Core Member of UMass Data Science Club 2023-24
- Coordinator of IITH Elektronica (Electronics, AI Club) and Cepheid (Astrophysics Club) 2018-19
- Coordinator of Security at IIT-H tech and cultural fest "ElanNvision" 2018-19

TEACHING

- Research Assistant under Prof. Sumohana Channappayya and Prof. Aditya Siripuram at IIT-H 2020
- Teaching Assistant for the course Digital Signal Processing under Prof. K Sri Rama Murty at IIT-H 2019

ACHIEVEMENTS

- Promising Startup award for Bosch AIShield at Bosch FitFest 2022
- Runner-Up Tinkerer's Lab Competition on AI 2019
- Appreciation for the work on Digital Pencil at the Inter IIT Tech Meet 2018
- Ranked 12^{th} nationwide in the KL University 2017