# 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** **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 the arithmetic sampling (sampling strategy that samples diverse sequences in parallel from **Large Language Models** ) with 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, scikit-learn, Azure, AWS, Docker, Git

- Spearheaded research in responsible AI, focusing on vulnerability assessment, robustness, explainability, fairness, causality and drift detection across 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**.
- Contributed to the initial stages of securing **LLMs**, focusing on analyzing and mitigating jailbreaking attacks (prompt engineering), which laid the groundwork for developing the AIShield Guardian application.
- Established partnerships with key players in the healthcare, financial, and MLOps sectors including Databricks, and Whylabs to enhance the security and reliability of their AI models, yielding a **revenue surge of around 10%**.
- Transitioned the research insights into product features by developing microservices, pipelines, and logging infrastructure across Azure & AWS, accounting for **30% of the overall workload**.

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

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

- Enhanced the web translation application by migrating existing pipelines based on XML and JSON to a fine tuned encoder-decoder Transformer model (multi-head self attention & cross attention) on the existing XML and JSON data.
- Implemented scalable REST APIs with **Flask**, and integrated with the frontend web interface built on **React** to demonstrate the 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

**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 coupled with Huber Loss, while conducting rigorous training on the EUROC dataset.
- Achieved superior performance (low validation and test loss) with attention-based models (Transformers), surpassing the capabilities of existing work on Dilated CNN's through hyperparameter optimization.

**Explaining Adversarial Robustness** (IITH) Jan 2021 - Apr 2021

- Employed SHAP, Grad-CAM, FAM techniques to produce insightful visual explanations for adversarial samples.
- Analyzed the behaviors of learned Convolution filters to understand the model's interpretability and robustness.
- Conducted in-depth research into the frequency domain analysis of adversarial examples employing Fourier transforms and filters for MNIST, CIFAR-10, Fashion MNIST datasets.
- Explaining adversarial examples in frequency and complex space via complex valued neural networks is in progress.

**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 and decoder LSTM architecture. Trained the model on MSVD dataset.
- Achieved a higher BLEU score compared to a baseline model with custom CNN and LSTM. This indicates that our model has better alignment between generated and reference captions, reflecting improved model performance.
- Ongoing work on predicting missing video frames through image in-painting, self-supervised learning techniques.

**AlphaConnect-4** (IITH) Jan 2020 - Apr 2020

- Inspired by deep mind's AlphaGo, implemented competitive multi-agent Reinforcement Learning on connect-4.
- Utilized a combination of Monte Carlo Tree Search (MCTS) for opponent modeling and Actor Critic for agent reinforcement. This scenario resembles a zero-sum mini-max game. Designed the connect-4 environment on python.
- Plotting the agent's performance (mean reward and std over training iterations) shows an increasing learning curve.
- Applied transfer learning to enable the agent's performance in connect-5 game, all 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

## TEACHING

- Research Assistant under Prof. Sumohana S Channappayya, 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 2022
- Appreciation for my work on Digital Pencil at the Inter IIT Tech Meet 2018
- Ranked 12$^{th}$ nationwide in the KL University 2017

## SERVICE

- Core Member of UMass Data Science Club 2023-24
- Core Member 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