Avinash Amballa

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

Education

University of Massachusetts Amherst *Amherst, USA*
MASTER OF SCIENCE IN COMPUTER SCIENCE *Aug 2023 - July 2025*

- CGPA: **4.0/4.0**
- Relevant coursework: Reinforcement Learning, Responsible AI, Advanced Natural Language Processing, Intelligent Visual Computing, Applied Statistics

Indian Institute of Technology Hyderabad (IIT-H) *Hyderabad, India*
BACHELOR OF TECHNOLOGY IN ELECTRICAL ENGINEERING WITH MINOR IN COMPUTER SCIENCE *Jul 2017 - June 2021*

- CGPA:**8.8/10.0**
- Relevant coursework: Data Structures, Algorithms, DBMS, Pattern Recognition & Machine learning, Image processing, Representation Learning, Probabilistic Graphical Models, Convex Optimization, Regression Analysis, Information theory

Work Experience

Google *USA*
GRADUATE STUDENT RESEARCHER *Feb 2024 – May 2024*

- Sampling diverse sequences in parallel from **Large Language Models (LLMs)** via Arithmetic Sampling.
- Experimenting arithmetic sampling with self-consistency and MBR decoding strategies for generating diverse candidates.
- Extending the arithmetic sampling to incorporate more diverse measures of sequence similarity in sampling space using ideas from box embeddings.

Bosch (BGSW) *Bangalore, India*
SENIOR RESEARCH SCIENTIST *Aug 2021 – July 2023*

- Spearheaded research in responsible AI, focusing on **vulnerability assessment**, **robustness, explainability, fairness, causality** and **drift** detection across various domains like vision, time series, speech and language models.
- Led AI security research, developing novel attack and defense strategies against various threat models, encompassing adversarial, poisoning, model extraction, and inference attacks. Resulted in **1 published paper** and **4 filed patents**.
- Significantly 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 **strategic alliances** with key players in the healthcare, financial, and MLOps sectors including Whylabs and ClearML to enhance the security and reliability of their AI models, yielding a **revenue surge of around 10%**.
- Transitioned the research insights into product features (Bosch AIShield), overseeing the development of microservices, pipelines, and logging infrastructure across Azure & AWS, accounting for **30% of the overall workload**.

GE Digital *Bangalore, India*
SOFTWARE DEVELOPMENT INTERN *May 2020 – July 2020*

- Enhanced the web translation application by migrating existing pipelines based on XML and JSON to a **encoder-decoder machine translation** model by incorporating **multi-head self attention and cross attention**.
- Developed and implemented scalable **REST APIs** with Flask, which were seamlessly integrated with the frontend web interface to demonstrate the web translation functionality.

Publications & Preprints

[1] Targeted attacks on Time Series Forecasting *arxiv preprint*
Yuvaraj Govindarajulu, **AVINASH AMBALLA,**, Pavan Kulkarni, Manojkumar Parmar *2301.11544*

[2] Discrete Control in Real-World Driving Environments using Deep Reinforcement Learning *arxiv preprint*
AVINASH AMBALLA, Advaith P, PRADIP SASMAL, Sumohana Channappayya *2211.15920*

[3] Automated Model Selection for Tabular Data *arxiv preprint*
AVINASH AMBALLA, Anmol Mekala, Gayathri Akkinapalli, Manas Madine, Priya Yarrabolu, Przemyslaw A. Grabowicz *2401.00961*

Patents

[1] A Method to detect AI poisoning attacks from the Data and/or Model *IN Patent App.*
AVINASH AMBALLA, Yuvaraj Govindarajulu, Manojkumar Parmar *202241068482*

[2] A Method of Targeted Attack on Time Series Models to alter the DIRECTION of the Output *IN Patent App.*
Yuvaraj Govindarajulu, **AVINASH AMBALLA**, Manojkumar Parmar *202241065028*

[3] A Method of Targeted Attack on Time Series Models to alter the MAGNITUDE of the Output *IN Patent App.*
Yuvaraj Govindarajulu, **AVINASH AMBALLA**, Manojkumar Parmar *202241065034*

[4] A Method of Sponge attack on Deep Learning Models to increase the inference time *IN Patent App.*
AVINASH AMBALLA, Yuvaraj Govindarajulu, Manojkumar Parmar *202441006640*

Research Projects (Research Assistant)

Gyro Correction in IMU sensors

PROF. K SRI RAMA MURTHY (IIT-H), 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 mechanisms,** and **Transformer Encoder** coupled with Huber Loss, while conducting rigorous training on the EUROC dataset.
- Through **hyperparameter optimization**, achieved superior performance (low validation and test loss) with attention-based models (Transformers), surpassing the capabilities of existing work on Dilated CNN's.

Explaining Adversarial Robustness

PROF. ADITYA T SIRIPURAM (IIT-H) *Jan 2021 - Apr 2021*

- Employed variants of **SHAP, Grad-CAM** and **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

PROF. ADITYA T SIRIPURAM (IIT-H) *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

PROF. VINEETH N BALASUBRAMANIAN (IIT-H) *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.

Articles

[1] ChatGPT - The future of Conversational AI *medium.com/@amballaavinash*

[2] Reinforcement Learning algorithms: An Overview *github.com/AmballaAvinash*

[3] Graph Compression by BFS: An Overview *github.com/AmballaAvinash*

Skills

Languages	Python, C, C++, Java, R
AI/ML	PyTorch, TensorFlow, Keras, Hugging Face, scikit-learn, numpy, OpenCV, openAI gym, aif360,NLTK
Web Dev	HTML, CSS, JavaScript, React, jQuery, Node.js, Express.js, flask
Misc.	Data visualization, Big data analytics, Azure, AWS, Docker, Git, Elasticsearch, PostgreSQL, Nginx

Teaching

2020 **Research Assistant** under Prof. Sumohana S Channappayya and Prof. Aditya Siripuram at IIT-H

2019 **Teaching Assistant** for the course Digital Signal Processing under Prof. K Sri Rama Murty at IIT-H

Achievements

2022	**Promising Startup and Global Info Sec award** for Bosch AIShield at Bosch FitFest
2022	**Runner-Up** Tinkerer's Lab Competition on AI at IITH
2018	**Appreciation for my work on Digital Pencil** at the prestigious Inter IIT Tech Meet - 2018
2017	**Ranked** 12th **nationwide** in the KL University exam

Service

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