Avinash Amballa

Webpage: https://amballaavinash.github.io **Mobile:** +16095054919
Linkedin: https://www.linkedin.com/in/avinashamballa **Mail:** amballaavinash@gmail.com

Education

University of Massachusetts Amherst *USA*

MS Computer Science *Aug 2023 - July 2025*

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

Indian Institute of Technology Hyderabad (IIT-H) *India*

Bachelor of Technology in Electrical Engineering with Minor in Computer Science and Engineering *Jul 2017 - June 2021*

- CGPA: **8.8/10.0**
- Relevant coursework:Algorithms, DBMS, Pattern Recognition, Machine learning, Image processing, Representation Learning

Work Experience

Google *USA*

Graduate Student Researcher *Feb 2024 –*

- Sampling diverse sequences in parallel from large language models via Arithmetic Sampling.

Bosch (BGSW) *Banglore, India*

SENIOR Engineer (Research Scientist) *Aug 2021 – July 2023*

- Spearheaded research in *responsible AI*, focusing on comprehensive vulnerability assessment, robustness, explainability, fairness, and drift detection across diverse domains, including computer vision, time series, speech and language models.
- Pioneered groundbreaking research in AI privacy and security, developing novel attack and defense strategies against various threat models, encompassing *adversarial attacks, poisoning, model extraction, and inference attacks.*
- Played a pivotal role in the early phases of securing large language models *(LLMs)*, specializing in countering jailbreaking and prompt injection attacks, laying the foundation for the creation of the AIShield Guardian application.
- Cultivated strategic partnerships with industry leaders in healthcare, automotive, and financial services, including Whylabs and ClearML, fostering collaborative innovation.
- Architected microservices, pipelines, and logging systems for Bosch AIShield product, showcasing expertise in designing robust and scalable systems.

GE Digital *Banglore, India*

Software Development intern *May 2020 – July 2020*

- Enhanced GE's web translation application by migrating existing pipeline based on XML and JSON to a modern deep learning architecture based on sequence-to-sequence models including *encoder-decoder with attention, and transformers (BERT)*.
- Built and deployed scalable REST APIs with Flask and integrated seamlessly with frontend web interfaces to enable low-latency translation.

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 Timeseries 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 Timeseries Models to alter the MAGNITUDE of the Output *IN Patent App.*

Yuvaraj Govindarajulu, **Avinash Amballa**, Manojkumar Parmar *202241065034*

[4] A Method of Sponze attack on Deep Learning Models to increase the inference time *IN Patent App.*

Avinash Amballa, Yuvaraj Govindarajulu, Manojkumar Parmar *in progress*

Research Projects (Research Assistant)

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 environment.
- Utilized a combination of *Monte Carlo Tree Search (MCTS)* for opponent modeling and *Actor Critic* for agent reinforcement (single agent and single opponent). Designed the connect-4 game environment as well.
- Achieved impressive results by training the agent on low-dimensional board games and successfully applied transfer learning techniques to enable the agent's performance in higher-dimensional environments, all with minimal additional training.

Gyro Correction in IMU sensors

Prof. K Sri Rama Murty (IIT-H), DRDO India (Defence Research and Development Organisation) *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 with attention-based models (Transformers), surpassing the capabilities of existing Dilated CNN methods in this domain.

Explaining Adversarial Robustness

Prof. Aditya T Siripuram (IIT-H) *Jan 2021 - Apr 2021*

- Visual Explanations: Employed variants of *Grad-CAM* and *GRAD-FAM* techniques to produce insightful visual explanations for adversarial samples. Analyzed the behaviors of Convolutional layers to enhance model interpretability and robustness.
- Frequency Domain Analysis: Conducted in-depth research into the frequency domain analysis of adversarial examples employing *Fourier transforms* and *filters* for MSIST and CIFAR-10 datasets
- Complex-Valued Neural Networks: Currently involved in ongoing research focused on explaining adversarial examples within a frequency and complex space using *complex valued neural networks*.

ViCaP: VIdeo Captioning And Prediction

Prof. Aditya T Siripuram (IIT-H) *Sep 2020 - Dec 2020*

- Implemented a *vision-language* video captioning method utilizing convolutional encoder with a attention based decoder
- Engineered a three-step search algorithm, employing Optical Flow techniques, to predict missing frames within video sequences. Additionally, exploited conditional *Generative Adversarial Networks (GANs)* for further frame prediction accuracy.
- Currently expanding capabilities in predicting missing frames within videos by exploring *self-supervised learning*.

Articles

[1] Reinforcement learning algorithms: An Overview *github.com/AmballaAvinash*

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

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

Skills

Coding	C, C++, Python, Java, R
AI/ML	Tensorflow, PyTorch, Keras, Scikit-learn, OpenCV, pandas, openAI gym, aif360
Web Dev	HTML, CSS, JavaScript, jQuery, flask, Node.js, Express.js
Misc.	PostgreSQL, Azure, Git, Docker, Elasticsearch, Nginx, Unity

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** 12^{th} **nationwide** in the KL University exam and received a prize worth 75k INR

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"