Avinash Amballa

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

Education

University of Massachusetts Amherst *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) *Hyderabad, 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 – May 2024*

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

Bosch (BGSW) *Bangalore, 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, such as computer 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 large language models (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%.
- Designed and implemented microservices, pipelines, and logging infrastructure for Bosch AIShield product, demonstrating proficiency in developing resilient and scalable systems.

GE Digital *Bangalore, India*

Software Development intern *May 2020 – July 2020*

- Enhanced the web translation application by migrating existing pipeline based on XML and JSON to a machine translation architecture based on sequence-to-sequence models including *encoder-decoder with attention, and transformers (BERT)*.
- Developed and implemented scalable RESTful APIs using 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 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)

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 (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*

- Inspired by work "Adversarial Examples Are Not Bugs, They Are Features", explaining adversarial examples is recommended.
- Employed variants of *SHAP*, *Grad-CAM* and *FAM* techniques to produce insightful visual explanations for adversarial samples. Analyzed the behaviors of Convolutional layers 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 using *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 by incorporating *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 environment.
- 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 game environment on python.
- Plotting the agent's performance (mean reward and std over training iterations) shows increasing learning curve.
- Applied transfer learning to enable the agent's performance in connect-5 game, all with minimal additional training.

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

Languages	Python, Java, C, C++, R
AI/ML	PyTorch, TensorFlow, 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

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"