# Avinash Amballa

| | | | | | | |

## Education

**University of Massachusetts Amherst** *USA*

MS Computer Science *Aug 2023 - July 2025*

- CGPA: **/4.0**
- Courses: Reinforcement Learning, Responsible AI, Methods of applied statistics

**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**
- Courses: Data Structures, Algorithms, DBMS, Computer Architecture, Reinforcement Learning, Pattern Recognition and Machine learning, Image processing, Representation Learning, Calculus, Differential Equations, Regression Analysis, Combinatorics and Graph theory, Matrix Analysis, Random Processes, Complex Variables, Internet of things, Signal Processing, Digital Modulation Techniques, Information science

## Work Experience

**Bosch Global Software Technologies** *Banglore, India*

SENIOR ENGINEER, Bosch AIShield (42.5 hrs/week) *Aug 2021 – July 2023*

- Engaged in responsible AI research (developing AI algorithms in a manner that aligns with ethical principles and values), focused on AI security (securing AI models from threats and vulnerabilities)
- Vulnerability analysis, defense analysis, explainability, robustness, fairness, and drift of AI models (Image Classification, Segmentation, Object Detection, Time Series Analysis, and Language models) against adversarial threats, poisoning, extraction, and inference attacks.
- Played a significant role in the early stages of AI security research related to LLM (Large Language Model) development focused on Jail breaking and prompt injections attacks, which paved the way for the creation of a AIShield GuArdIan.
- Contributed to product development at Bosch AIShield (building micro services, handling pipelines, logging), with a key role in managing customer partnerships (Whylabs, ClearML) and internal partnership's across healthcare, automotive and BFSI industries.
- Transferred the outcomes of this research into tangible deliverables, spanning product enhancements, the publication of technical papers, and the acquisition of patents

## Publications

**[1] Targeted attacks on Time Series Forecasting** *under review*

Yuvaraj Govindarajulu, **Avinash Amballa**, Pavan Kulkarni, Manojkumar Parmar *in ACML 2023*

**[2] Discrete Control in Real-World Driving Environments using Deep Reinforcement Learning** *arxiv preprint*

**Avinash Amballa**, Advaith P, PRADIP SASMAL, Sumohana Channappayya *2211.15920*

## Patents

**[1] A Method to detect AI poisoning attacks from the Data and/or Model** *IN Patent: 202241068482*

**Avinash Amballa**, Yuvaraj Govindarajulu, Manojkumar Parmar *docket number: 404446*

**[2] A Method of Targeted Attack on Timeseries Models to alter the DIRECTION of the Model Output (A method of assessing vulnerability of an AI system and a framework thereof)** *IN Patent: 202241065028*

Yuvaraj Govindarajulu, **Avinash Amballa**, Manojkumar Parmar *docket number: 403873*

**[3] A Method of Targeted Attack on Timeseries Models to alter the MAGNITUDE of the Model Output (A method of assessing vulnerability of an AI system and a framework thereof )** *IN Patent: 202241065034*

Yuvaraj Govindarajulu, **Avinash Amballa**, Manojkumar Parmar *docket number: 403874*

**[4] A Method of Sponze attack on Deep Learning Models to increase the inference time ((A method of assessing vulnerability of an AI system and a framework thereof )** *IN Patent: in progress*

**Avinash Amballa**, Yuvaraj Govindarajulu, Manojkumar Parmar *docket number: in progress*

## Skills

| | |
|---|---|
| **Coding** | C, C++, Python, R, Arduino, MATLAB, Latex |
| **AI/ML** | Tensorflow, PyTorch, Scikit-learn, OpenCV, openAI gym |
| **Web Dev** | HTML, CSS, JavaScript, jQuery, flask, Node.js, Express.js |
| **Misc.** | PostgreSQL, Azure, Git, Docker, AKS, Unity, Elasticsearch, Nginx, Open Source, Reinforcement Learning by David Silver, NPTEL Deep Learning, Coursera's Google Data Analytics, Coursera's Deep Learning Specialization |

## Articles

**[1] ChatGPT: The future of Conversational AI**

## Projects

### AlphaConnect-4

Prof Vineeth N Balasubramanian *Jan 2020 - Apr 2020*

- Inspired by deep mind's AlphaGo, implemented competitive Multi-agent RL (single agent and single opponent) on Connect-4 game env
- Employed Monte Carlo Tree Search (MCTS) on opponent and Policy Gradients on the agent. Designed the game environment as well
- Trained the agent on low-dimension boards and used transfer learning to make the agent play in higher dimensions with minimal training

### Gyro Correction in IMU sensors

Prof. K Sri Rama Murty, DRDO India (Defence Research and Development Organisation) *Apr 2021 - Jul 2021*

- IMU sensors are noisy and biased due to axis misalignment. Hence, developed a model for predicting the gyro correction in IMU sensors
- Trained with various architectures such as DB-LSTM, LSTM with attention, and Transformer Encoder with Huber Loss on the EUROC dataset
- With hyperparameter tuning, attention models achieve better performance with respect to existing work on Dilated CNN

### Explaining Adversarial Examples & Robustness

Prof. Aditya T Siripuram *Jan 2021 - Apr 2021*

- Using variants of Grad-CAM and GRAD-FAM, creating visual explanations on adversarial samples and analyzing the Conv layers
- Studied Frequency domain analysis of adversarial examples through Fourier transforms and filters and evaluating Adversarial robustness
- Research on explaining adversarial examples in a frequency and complex space using complex-valued neural networks is in progress

### VICAP: VIdeo Captioning And Prediction

Prof. Aditya T Siripuram *Sep 2020 - Dec 2020*

- Implemented the video captioning (vision-language) method by processing the video data using CNN with DB-LSTM encoder-decoder
- Predicted the missing frames in the video using a three-step search algorithm (Optical flow), conditional GANS
- Extended this work on predicting missing frames in a video using self-supervised learning (in progress)

### Metaverse

Bosch Hackathon *Jun 2022 - Aug 2022*

- Created a Metaverse Persona using UneeQ framework, and hosted this to local host and added customization to the webpage
- Integrated google dialogue flow backend through webhook URL. Fed custom intents to make the persona respond accordingly to the user
- In addition, integrated person identification, gender detection, and emotion recognition to the persona

### Digital Pencil

Inter IIT Tech meet *May 2018 - Jul 2018*

- This device translates hand Gestures into digital Characters. Created labeled data set using pyGARL, Arduino pro micro, accelerometer
- Trained the model using linear SVM, kernel SVM, and ANN. Results show that kernel SVM and ANN outperform linear SVM in accuracy

### Open Face

Self project *Jun 2021 - Aug 2021*

- Implemented One-Shot (Few-Shot) Learning Facial Recognition using Siamese Network (Embedding Learning) on AT&T faces data
- Trained and analyzed the performance of Prototypical Networks and Relation networks. Deployed the models in the open-vino framework

## Awards and Achievements

2022 **Promising Startup award** Bosch AIShield at Bosch FitFest

2022 **Global Info Sec award** Bosch AIShield

2018-19 **Presented my works at Inter IIT** Tech Meet - 2018 at IIT Bombay and Tech Meet - 2019 at IIT Roorkee

2017 **Secured $12^{th}$ rank nationwide** in the KL University exam and received a prize worth 75k INR

## Positions of Responsibility

2020 **research Assistant** under Prof. Channapayya and Siripuram at IIT-H

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

2018-19 **Core member** of Elektronica(Electronics, AI, Signal Processing Club) and Cepheid(Astronomy, Astrophysics Club) at IIT-H

2018-19 **Security Coordinator** at IIT-H tech and cultural fest "ElanNvision"

## Other Interests

Travel, Cricket, Photography & editing, Astrophysics, and Quantum Mechanics