# Yashwanthi **Anand**

PhD Candidate | AI Researcher

an.yashwanthi@gmail.com | yashwanthi.xyz | yashianand | yashwanthi-anand

## Summary

Ph.D. researcher focusing on safe and reliable autonomous systems, reinforcement learning, and AI evaluation, using Bayesian inference, human feedback modeling, failure detection, and adaptive learning for autonomous agents in robotics and simulation environments, with tools such as PyTorch, ROS2, and Isaac Gym.

## Skills

| | |
|---|---|
| **Frameworks** | PyTorch, TensorFlow |
| **Simulation Environments** | Gymnasium, PyBullet, NVIDIA Isaac Gym, MuJoCo |
| **Tools** | LiteLLM API, Weights & Biases, D3.js |
| **Programming** | Python, C++, ROS2, Git |

## Education

### Ph.D. in Computer Science, Oregon State University

*Corvallis, OR*

Advisor: Dr. Sandhya Saisubramanian

*2022 – Present*

- Research focuses on reinforcement learning, Bayesian inference, and scalable failure detection to ensure safe and reliable autonomous systems.
- Developed a differential testing framework for diagnosing failures in autonomous systems, enabling black-box evaluation of agent reliability.
- Developed a framework that integrates human feedback into decision-making to improve model alignment and minimize risks.

### M.S. in Computer Science, Oregon State University

*Corvallis, OR*

Advisors: Dr. Thinh Nguyen, Dr. Jinsub Kim

*2019 – 2021*

- Thesis: Resource-Aware Distributed Data Sanitization for Privacy Preserving Machine Learning.
- Developed algorithm for minimizing privacy leakage under bandwidth constraints across federated sensor networks.

## Publications

### Uncovering Systemic and Environment Errors in Autonomous Systems Using Differential Testing

*AAAI FSS*

**Yashwanthi Anand***, Rahil P Mehta*, Manish Motwani, Sandhya Saisubramanian

*2025*

- Introduced a black-box testing framework to distinguish between agent-side model errors and environment-induced task infeasibility.
- Demonstrated improved state coverage and error attribution across continuous and discrete domains with single and multi-agents.

### Multi-Objective Planning with Contextual Lexicographic Reward Preferences

*AAMAS*

Pulkit Rustagi, **Yashwanthi Anand**, Sandhya Saisubramanian

*2025*

- Proposed a learning-based method to infer context-dependent priority orderings over multiple objectives in sequential decision-making.
- Validated approach in simulated domains and hardware experiments using TurtleBot navigation tasks.

### Adaptive Feedback Selection for Avoiding Negative Side Effects

*RLC RLBRew Workshop*

**Yashwanthi Anand**, Sandhya Saisubramanian

*2024*

- Proposed a human feedback preference-aware query selection algorithm optimizing information gain and user perception of workload and competence.
- Evaluated the approach across simulated Gymnasium domains and on a Kinova Gen3 7DoF robotic arm.

## Selected Projects

### Visual Exploration of Large-Scale Image Dataset for Machine Learning with Treemaps

*IEEE Viz*

Donald Bertucci, Md Montaser Hamid, **Yashwanthi Anand**, Anita Ruangrotsakun, ..., Minsuk Kahng

*2022*

A scalable and interactive way to explore image datasets used in machine learning, that helps explore images using a zoomable treemap.

### Score Distribution Graph-Based Visualization

Click here for a demo

*Feb 2022*

A browser-based visualization tool to analyze an ML model's score distributions.

### Tumor Suppressor Gene Prediction using GCNs

Click here for project details

*Mar 2020*

A semi-supervised model using Graph Convolutional Networks achieving 75% accuracy on biological graph datasets.

# Experience

### Graduate Research Assistant, Oregon State University
*Corvallis, OR*

Advisor: Dr. Sandhya Saisubramanian | Intelligent and Reliable Autonomous Systems Lab
*Sep 2021 – Present*

- Designed RLHF pipelines that integrate human feedback preferences and adaptive querying, to improve safety in autonomous systems.
- Conducted human-subject experiments using a Kinova Gen3 7DoF robotic arm to validate safe learning from human feedback.
- Developed a black-box differential testing framework to localize failures in autonomous systems caused by agent's model defects or environment task infeasibility.
- Built GPU-enabled evaluation tool for robustness and coverage analysis across RL domains.

### Graduate Teaching Assistant
*Corvallis, OR*

Oregon State University
*2019 – 2021, 2024*

- Instructed and mentored over 100 students in Computer Networks and Reinforcement Learning courses.
- Co-led lectures and lab sessions.

### Software Intern
*Tamil Nadu, India*

ikval Softwares LLP
*Dec 2018 – Jan 2019*

- Prepared a phoneme-labeled Tamil dataset to support the development of a text-to-speech engine.

# Awards

2025 **Travel Award**, AAAI Fall Symposium Series
2025 **Accepted to CRA-WP Grad Cohort for Women**, Computing Research Association
2024 **Scholarly Presentation Award**, Oregon State University

# Activities

### Student Reviewer
*2021–*

AAAI, CHI, IUI

### RL Reading Group Organizer
*2024*

AI Graduate Student Association, Oregon State University

Facilitated student-led research seminars, reading groups, and invited speaker sessions on Reinforcement Learning.

### Treasurer and Faculty Relations Officer
*2022 – 2023*

EECS Graduate Student Association, Oregon State University

Managed event budgets and organized faculty-student engagement initiatives.