Athul Raj

athul_raj@outlook.com | linkedin.com/in/athul-raj-nine9 | github.com/rajathul | Espoo, Finland

Research Interests

Large Language Models for reasoning and knowledge-grounded systems, efficient fine-tuning and evaluation of domain-specific models, and the integration of LLMs into embodied and robotic systems.

Education

Aalto University — Espoo, Finland
Joint MSc in Autonomous Systems; Minor in Entrepreneurship and Innovation — *Aug. 2024 – Sep. 2025*

KTH Royal Institute of Technology — Stockholm, Sweden
Joint MSc in Autonomous Systems; Minor in Entrepreneurship and Innovation — *Aug. 2023 – Aug. 2024*

National Institute of Technology Silchar — Silchar, India
Bachelor of Technology in Electronics and Communication Engineering — *July 2018 – May 2022*

Relevant Coursework

Machine Learning, Introduction to Robotics, Distributed Artificial Intelligence and Intelligent Agents, Reinforcement Learning, Computer Vision, Basics of Sensor Fusion.

Research Experience

Master Thesis Intern — Mar 2025 – Sep 2025
IRIS AI

- Designed and executed controlled experiments to evaluate how domain-specific fine-tuning improves RAG factual accuracy, semantic match, and retrieval grounding
- Implemented **LoRA** and **full SFT** pipelines on **Llama 3.1 8B Instruct** and compared them under identical data
- Developed a data-augmentation workflow using the **OpenAI API** to generate high-quality reasoning traces for reasoning-based fine-tuning experiments.
- Implemented a comprehensive evaluation using **lexical metrics** (ROUGE), **semantic metrics** (WISDM), and **LLM-as-a-judge metrics** (AlpacaEval, custom Prometheus-based metrics).
- Scaled experiments on **multi-GPU HPC clusters using DeepSpeed and Accelerate**

Research Publications

- **Short-term Electricity Price Forecasting: A Study in the Iberian Market**
  Naiya Khalid, Isabel Seruca, Athul Raj
  Submitted to CENTERIS / ProjMAN / HCist 2025
- **A Review on Self-Supervised Learning**
  Athul Raj, Srinjoy Dutta
  International Journal of New Technology and Research, January 2023

Work Experience

AI Intern (Part-time) — Feb 2025 – Present
VanaciPrime — *Remote*

- Led a comparative study of **rule-based, open-source LLM, and proprietary LLM** approaches for document information extraction, achieving accuracy improvements from 35% to **70%** (open-source) and **95%** (Gemini) on real-world client data.
- Designed and implemented an **Agentic LLM-based research and retrieval system** integrating **ArXiv**, internal knowledge bases, and automated web crawling to support structured evidence gathering and report generation.
- Built a **time-series forecasting pipeline** for electricity prices, including data ingestion, feature engineering, and model training with scikit-learn.
- Developed a **document understanding pipeline** for identity and contract processing, combining **PaddleOCR** with **MRZ parsing** to extract structured fields from ID documents and populate downstream legal contracts.

Software Developer July 2022 – July 2023
Samsung Data Systems *Gurugram, India*

- Developed backend solutions using Spring and SAP Hybris for e-commerce applications.
- Collaborated with cross-functional teams using Jira and Confluence.

Research and Development Intern June 2021 – July 2021
Unique Maritime Group *Sharjah, UAE*

- Integrated hardware and software for an autonomous survey vessel; developed a real-time battery monitoring system on Arduino.

Projects

Autonomous Mobile Robot with 5-DOF Arm | *Python, PyTorch, ROS2, OpenCV*

- Developed an autonomous navigation and object manipulation pipeline using ROS2 nodes for perception, motion planning, and control.
- Trained a **MobileNetV2** object detection model in PyTorch and integrated it with a RealSense depth camera for object detection and mapping.
- Implemented the **state machine** of the robot to coordinate perception, manipulation, and navigation modules, enabling robust multi-module operation.

Reinforcement Learning Agent using DQN | *Python, PyTorch, Gymnasium*

- Implemented a Deep Q-Learning (DQN) agent in PyTorch to solve the **LunarLander-v2** environment with stable convergence.
- Experimented with ϵ-greedy exploration, replay buffer sampling, and target-network updates to stabilize training.
- Compared training performance across different reward shaping strategies and visualized learning curves using Matplotlib.

Technical Skills

Languages: Python, C++, MATLAB
Frameworks & Tools: PyTorch, Hugging Face, LangChain, ROS2, scikit-learn, OpenCV, DeepSpeed, Accelerate
AI/ML Expertise: Fine-tuning LLMs, LLaMA/Mistral/Qwen, RAG, vector DBs, OpenAI/Gemini APIs, multi-GPU training
Other: GitHub, GitHub Actions, Docker, Linux, VS Code, Jupyter, Pandas, NumPy, Matplotlib

References

Dr. Ivan Vankov — ivan@iris.ai
IRIS AI

Dinu Granaci — dinugranaci@vanaciprime.com
VanaciPrime

Chetan A. Chitnis — chitnis@uniquegroup.com
Unique Maritime Group