

Sai Suresh Macharla Vasu

Master's Student, Saarland University

saisuresh1999@gmail.com — sama00014@stud.uni-saarland.de — +49 1573 824 1524 —

linkedin.com/in/sai-suresh-macharla-vasu — saisuresh1999.github.io

RESEARCH INTERESTS

AI safety and fairness in LLMs; Persona Modeling & Social Simulation.

EDUCATION

Saarland University, Saarbrücken, Germany

Oct 2023 – Present

Master of Science in Data Science and Artificial Intelligence (DSAI)

Current Grade: 1.6

Anna University, Chennai, India

Aug 2017 – Aug 2021

Bachelor of Engineering in Computer Science and Engineering

Final Grade: 8.92 CGPA (\approx 1.8 GPA on the German grading scale)

ACADEMIC EXPERIENCE

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbrücken, Germany

Student Research Assistant – Interactive Machine Learning Group

Jul 2024 – Nov 2025

- **AuditCopilot: Fraud Detection with LLMs** — achieved interpretable anomaly detection surpassing traditional audit tests.
- Worked on **tax evasion detection** using LLMs to surpass conventional machine learning baselines on journal-entry datasets.
- Designed **model faithfulness evaluations** via counterfactual generation and post-hoc explanation techniques for financial anomaly detection.

Saarland University, Saarbrücken, Germany

Teaching Assistant / Head Tutor – Machine Learning Courses

Nov 2024 – Feb 2026

- **Elements of Machine Learning (WS 25/26)**: Head Tutor under Prof. Dr. Isabel Valera — managing and coordinating a team of tutors, overseeing grading and tutorial logistics.
- **Machine Learning (SS 25)**: Tutor under Prof. Dr. Peter Ochs — led sessions on optimization-based learning methods and the mathematical foundations of ML.
- **Elements of Machine Learning (WS 24/25)**: Tutor under Prof. Dr. Isabel Valera — led sessions that emphasized conceptual understanding and provided an intuitive explanation of ML fundamentals.

PUBLICATIONS

- **Macharla Vasu, S. S.**, Sheth, I., Wang, H.-P., Binkyte, R., Fritz, M. (2025). Justice in Judgment: Unveiling (Hidden) Bias in LLM-assisted Peer Reviews. **NeurIPS 2025 Workshop on LLM Evaluation**. arXiv:2509.13400
- Kadir, M. A., **Macharla Vasu, S. S.**, Nair, S. S., Sonntag, D. (2025). AuditCopilot: Leveraging LLMs for Fraud Detection in Double-Entry Bookkeeping. **NeurIPS 2025 Workshop on Generative AI in Finance**.
- Kempermann, M. L., **Macharla Vasu, S. S.**, Raveenthiran, M., Weber, I. (2026). Challenges of Evaluating LLM Safety for User Welfare. Submitted to the **International Association for Safe & Ethical AI (IASEAI'26)**, Under review.

AWARDS

Saarland Scholarship for International Students

Saarbrücken, Germany

StudienStiftungSaar & Saarland Ministry of Finance and Science

Summer Semester 2025

Awarded for outstanding academic performance during the M.Sc. in Data Science and Artificial Intelligence at Saarland University.

Rashtrapati Scout (Pfadfinder) Award

New Delhi, India

The Bharat Scouts and Guides, presented by the Former President of India, Mr. Pranab Mukherjee

Dec 2016

National-level distinction recognizing excellence in leadership, community service, and citizenship.