Adrian Sauter

Contact No.: +49 157 33715232 **| Email:** adriansauter07.AS@gmail.com | [**LinkedIn**](#) | [**Website**](#)

Master's student in Artificial Intelligence at the University of Amsterdam, focused on aligned, explainable, and generally safe AI. I combine industry experience, technical research, interdisciplinary thinking, and clear communication to bridge complex models with accessible insights. My work centres on explainable AI and responsible ML systems that build trust and deliver practical impact.

EDUCATION

Universiteit van Amsterdam   **Sep 2023 – Present**

M. Sc. in Artificial Intelligence

- **GPA:** 9.05 (best: 10, worst: 1; grades 8-10 are considered excellent; more information [here](#))
- **Thesis:** Contextual Sensitivity of Moral Beliefs in Large Language Models
- **Best project award** in the course "Interpretability and Explainability in AI"
- **Relevant Coursework:** Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, Reinforcement Learning, Fairness and Transparency in AI, Game Theory, Interpretability and Explainability in AI, Human-Centered AI

Eberhard Karls Universität Tübingen   **Oct 2019 – July 2023**

B. Sc. in Cognitive Science

- **GPA:** 1.41 (best: 1, worst: 4)
- **Thesis:** Video Background Extraction with a Masked-Autoencoder-Based Neural Network
- **Relevant Coursework:** Programming and Machine Learning, Psychology, Calculus and Linear Algebra, Statistics and Probability, Linguistics, Philosophy, Neurobiology

WORK EXPERIENCE

KPN Amsterdam   **Feb 2025 – Aug 2025**

Research & Development Intern for XAI

- Researching Explainable AI for Customer Call Classification through Concept-Bottleneck Models using LLMs.

University of Amsterdam   **Sep 2024 – Present**

Teaching Assistant

- Led tutorials and seminars for graduate courses in Computer Vision 1, NLP 1, and Interpretability and Explainability in AI, instructing and supervising 30-200 students. Graded assignments and exams, devised exam questions, supervised projects.

ELLIS Unit Amsterdam   **Apr 2024 – Aug 2024**

Project Assistant

- Contributed to the planning and communication of projects that are supported and organized by ELLIS, created content on AI and the ELLIS Unit Amsterdam.

Neuro-Cognitive Modeling Group, Universität Tübingen   **Mar 2023 – Aug 2023**

Research Assistant

- Set up online psychological experiments, provided technical consultancy for professors associated with the lab, graded assignments of the Cognitive Modelling graduate course.

Leibnitz-Institute for Knowledge Media Research, Tübingen   **Dec 2021 – Oct 2022**

Research Assistant

- Supervised in-person psychological experiments and helped analyzing the obtained data.

PUBLICATIONS / PREPRINTS

- **Sauter, A.,** Pemmasani Prabakaran, R.S., Miletić, M., & Ott, R. (2024). "Studying How to Efficiently and Effectively Guide Models with Explanations" - A Reproducibility Study. **Transactions on Machine Learning Research**.
- Traub, M., Becker, F., **Sauter, A**., Otte, S., & Butz, M. (2024). Loci-segmented: improving scene segmentation learning. In **International Conference on Artificial Neural Networks** (pp. 45–61).
- **Sauter, A.,** Zuidema, W., & Kloots, M. D. H. (2025). The Curious Case of Visual Grounding: Different Effects for Speech-and Text-based Language Encoders. arXiv preprint arXiv:2509.15837. **Submitted to ICASSP 2026.**