Ediz Ucar

+44 7918 652 632 — edizucar83@gmail.com — github.com/edizucar — linkedin.com/in/ediz-ucar — Cambridge, UK

Summary

Software Engineer at Entrust, with a strong interest in ML, AI safety, and existential risks.
Languages: Python, Go, C, C++, PyTorch, MATLAB, Einops, TransformerLens, SAELens
Tools & Methodologies: Docker, Git, Unix, Scrum, CI/CD (Bamboo, Bitbucket, Artifactory)

Professional Experience

Software Engineer **October 2023 – Present**
Entrust, Cambridge, UK

- Developed software platform for Hardware Security Modules, focusing on monitoring and maintenance.
- Worked within Scrum teams on industry-standard software using Golang, version control, and CI/CD pipelines.
- Enhanced system performance through continuous integration and automated deployment.

Software Engineering Intern **Jul 2022 – Sep 2022**
Entrust, Cambridge, UK

- Implemented cryptographic standards (PKCS11) in C.

Software Engineering Intern **Jul 2021 – Sep 2021**
Loopworks, Brighton, UK

- Developed chatbot functionalities using Microsoft's NLP APIs.

Education

B.A. in Computer Science **Sep 2020 – Jul 2023**
University of Cambridge - Second Class Division I

AI Research Accelerator Program **Mar 2023 – Apr 2023**
Cambridge AI Safety Hub - Intensive 2 week coding program modeled on the ARENA program. Created various models from the ground up (CNN, RNN, Transformer, DQN, PPO). Performed various Mechanistic Interpretability techniques using TransformerLens.

Futures Fellowship **Jan 2024 – Mar 2024**
Cambridge Existential Risks Initiative - Explored existential risks including AI and nuclear risk, governance, and philosophical implications. Replicated Power et al.'s grokking research.

A-Levels **Sep 2018 – Jul 2020**
BHASVIC Sixth Form College, Brighton, UK - Mathematics (A*), Further Mathematics (A*), Computer Science (A*), Physics (A*)

Projects

Evolution of SAE Features Across Layers in LLMs **Jun 2024 – Sept 2024**
Repo Demo Paper
I was a lead contributor to a paper accepted into NeurIPS ATTRIB which developed a method for creating causal graphs for GPT-2 Sparse Autoencoder features which gave semantic feature clustering. Conducted under mentorship of Stefan Heimersheim of Apollo Research.

Adversarially Trained Transformers **Jan 2024 – Jul 2024**
Repo
Led a research project on adversarial training to eliminate unwanted representations in transformer models' latent space.

Instagram Transparency Visualization **Sept 2022 – May 2023**
Repo Demo Paper
Built a visualization tool to enhance Instagram's data transparency for parents, informed by user research. Supervised by Prof. Alan Blackwell and Bianca Schor of the University of Cambridge, Computer Laboratory.

Explanations as a Mechanism for Algorithmic Transparency **Oct 2022 – Dec 2022**
Paper
Conducted a user study into the effect of explanations as a mechanism for improving the algorithmic transparency of social media services such as Facebook.

Achievements

Cambridge AI Hub Hackathon - First Place **Winter 2024**
Won the hackathon by developing a new training regiment for Sparse Autoencoders of LLMs by adding a new loss term based on KL divergence within token space.

HackX Hackathon - Runner Up **Autumn 2022**
Repo
Developed existential risk solutions, achieving runner-up among fifteen teams.

GetSeenVentures GenAI Hackathon - Top 10 **Mar 2024**
Repo
Created 'Skill Sailor,' an application to assist refugees in identifying career paths.

Languages & Interests

Languages: English (Native), Turkish (Proficient), Spanish (Lower Intermediate)
Interests: AI Safety (Research at Cambridge AI Safety Hub), Cinema (Sci-Fi Enthusiast, former Film Society President), Bouldering & Tennis