

Balázs László

AI Safety Generalist | Technical & Governance & Advocacy

Blackpool, UK (CEEALAR) | lbazsi09@gmail.com | +36 70 365 6656

<https://www.linkedin.com/in/balazs-laszlo19/> | <https://github.com/lbazsi>



Projects

Technical AIS project sprint, BlueDot Impact, Remote

May 2026 - June 2026

- Investigated whether low-level linguistic features are represented, sparsely separable, and suitable targets for causal interventions in language-model activations.
- Established the experimental foundation for follow-up research on linguistic structure, behavioral effects, and constitutional language.

Research Sprint, Apart Research, Remote

Sept 2025 - Sept 2025

- Developed **Python** prototypes for a toolkit designed to reduce weaknesses in **CBRN AI benchmarks**.
- Earned **second place** globally while gaining experience in collaborative technical research, experimental evaluation, and research communication.

International research competition, TUDOK, HU

Dec 2024 - March 2025

- Developed a research proposal on **reducing bias** in large language models through structured databases informed by EU AI Act requirements.
- Received **two international second-place awards** for the AI governance proposal and the MLM participation study, demonstrating interdisciplinary expertise across technology, policy, and sociology.

National university research competition, BGE TDK and OTDK, HU

June 2024 - Nov 2024

- Co-authored a mixed-methods study examining student participation in multi-level marketing networks, combining quantitative analysis with qualitative research.
- Advanced to the national round and received a **special award** strengthening academic writing and presentation skills.

Robotics World Championship, First Global Challenge 2024, Athens

June 2024 - Sept 2024

- Co-built and programmed a competition robot using Python and sensor-based control, prioritizing rapid task execution, mechanical stability, and iterative troubleshooting.
- **Ranked 34th among 193** international teams while developing engineering teamwork, resilience under pressure, and cross-cultural collaboration skills.

Experience, research & volunteering

Participant, Technical ML4Good Bootcamp, Greens Norton - UK

Sept 2026 - Sept 2026

- Completed an intensive technical bootcamp focused on machine learning, AI safety, and applied research methods.
- Worked on hands-on technical exercises and collaborative projects while developing practical skills in Python, ML experimentation, and research workflows.

Mentor, Sentient Futures Incubator Program, remote

Aug 2026 - Nov 2026

- Mentoring fellows on improving synthetic data for animal-aligned value generalization by identifying potential risks and effects of linguistic differences on downstream model behavior.
- Supporting fellows in shaping their own projects and research directions throughout the program.

AI Integrations, LINAK - Operations Excellence department and R&D, Guderup - DK

April 2026 - Sept 2026

- Developed technical and operational plans to integrate AI into company workflows.
- Led workshops and presentations while contributing to technical and operational discussions and planning.

- Research Fellow**, Sentient Futures Incubator Program, remote **March 2026 – May 2026**
- Designing a welfare-first monitoring & mitigation framework for zoonotic disease risk that uses animal-welfare indicators to support earlier, humane interventions.
 - Started building Disease Free Futures, which will be a non-profit focusing on wild animal disease mitigation.
- AI Researcher**, Algovverse AI Research Program, remote **Oct 2025 – Present**
- Working on steganography detection research from a **mechanistic interpretability** perspective primarily in Python.
 - Collaborating in an international research environment on experimental design, model analysis, and technical research workflows.
- Leader**, Next Generation Hungary (NGH) - Erasmus+ (ID:E10372683) **Sept 2024 – Present**
- Participated in international youth initiatives addressing climate change, and youth empowerment.
 - Contributed to projects organized in Portugal, Greece and Lithuania, focusing on emerging topics.
- Secretary General**, JEF Hungary, HU **July 2024 – July 2025**
- Directed international youth projects, conferences, and policy discussions while securing and administering project grants.
 - Represented Hungary in the organisation's Federal Committee, developing international leadership, financial management, and stakeholder-coordination experience.
- Programming Educator - Python**, Logischool, Tatabánya - HU **Sept 2022 – Jan 2023**
- Taught Python to secondary-school students through interactive exercises and project-based lessons.
 - Adapted technical explanations to different learning styles, strengthening leadership and mentoring.

Education

- University of Cambridge**, Computer Science Tripos **January 2026**
- Offered admission to the Computer Science Tripos; unable to take up the place due to insufficient funding.
- University of Southern Denmark**, BSc in Mechatronics **Sept 2025 - Sept 2026**
- Administration - ARC Rocketry Club: Developing an inventory system via building data-driven automation pipelines using **Python and Microsoft tools**
 - Founder, President - SIRAIIS (Student Initiative for AI Safety Research): Established a university group focusing on **AI Safety research**

Courses

- AGI Strategy Course**, BlueDot Impact, Remote **Aug 2025 - Oct 2025**
- Explored global governance, safety, and long-term planning for advanced AI systems.
 - Developed early expertise in **technical AI governance** at the intersection of technology and public policy.
- AI Safety Technical Fellowship**, AI Safety Hungary, Budapest - HU **Aug 2023 - Sept 2023**
- Explored foundations of **AI alignment and interpretability** through lectures and collaborative research discussions.
 - Strengthened theoretical knowledge of **machine learning** and research methodologies.
- Mathias Corvinus Collegium Talent program**, MCC, HU **Jan 2022 - June 2025**
- Selected for a multi-disciplinary excellence program providing training in both **STEM and social sciences**.
 - Attended intensive seminars and conferences with academics, entrepreneurs, and policymakers.

Skills

Programming languages: Python, R, C, C++, LaTeX, SQL	Languages: Hungarian	native (C2)
Technologies: Excel, Word, Canva, WordPress	English	fluent (C2)
Tools: Git, CAD (NX), MATLAB, Power BI	German	fluent (C1)
	Danish	intermediate (B1)