

Manon Lilott Kempermann

Dresden, Germany • manonkempermann.eu • manon@kempermann.email

Personal Statement

I see myself as a Global Citizen. For that, I want to contribute to society locally and globally to drive community and improve present and future life on earth. I am particularly concerned about global catastrophic risk from AI and believe that only substantial efforts in AI safety and governance will enable the benefits we hope to get through this technology. I therefore aim to effectively grow and advance these fields within my capacities.

University and School Education

Saarland University, Germany: BSc Data Science and AI 2023 - (2026)

- Minor in Philosophy
- Average Grade: 1.3 (1.0 highest in German scale, ~ 3.8 GPA in US, ~ first-class undergraduate degree with honours in UK)
- Relevant Coursework:
 - **Mechanistic Interpretability:** *An Interpretability View of The Impact of Fine-Tuning on the Refusal Mechanism in Large Language Models* ([report available](#))
 - **Research Seminar:** *Fine-Tuning: From Benign to Malign – How Different Fine-Tuning Methods Degrade Safety Guardrails in LLMs During Benign Fine-Tuning* ([report available](#))
 - **Programming:** Java (2048 Game, Route Planner, Compiler translating subset of C into MIPS), C (Image Edge Detector, Page Rank), MIPS Assembly (Rock, Paper, Scissors with Random Number Generator)

Kilgraston School, UK: Scottish Qualification Certificate (Highers) 2021 - 2022

- Subjects taken with final grade: Mathematics (A), Human Biology (A), Chemistry (A), Latin (A), German (A), Business Management (A)

Technical University Dresden, Germany: Young student

- Psychology 2019 – 2021
- Civil Engineering 2020 – 2021

Scholarships and Honors

2025	<i>Young Researcher and Abbe-Grantee</i> (Heidelberg Laureate Forum, Carl-Zeiss Foundation): for participating in the 12th Heidelberg Laureate Forum for Mathematics and Computer Science
2025 – ongoing	<i>Scholarship</i> (Cusanuswerk, German Federal Ministry of Research, Technology and Space): for outstanding academic performance and societal engagement (top ~1% German students)
2023 – ongoing	<i>Honours Programme</i> (Saarland University, Faculty of CS & Mathematics): top 5% students
2021 – 2022	<i>Scholarship</i> (Dresden Trust): for attending Kilgraston School

Work Experience

- 01/2026 – ongoing *Program Director (Safe AI Germany, SAIGE)*: coordinating the SAIGE Incubator Program
- 12/2025 – ongoing *Pathfinder Mentor (Kairos)*: mentoring AI safety group organisers within the Pathfinder Fellowship
- 08/2025 – ongoing *Founder and Lead Organiser (AI Safety Saarland)*: organizing public lectures (300+ attendees), AI Safety Introductory Fellowship (40 fellows), mentored research incubator (30 fellows, 15 mentors), and general group management (10 executive members)
- 05/2024 – 02/2026 *Research Assistant (Interdisciplinary Institute for Societal Computing)*: evaluating LLMs on context-sensitive user risks

AI Safety Training

- 01/2026 *Technical AI Safety (BlueDot Impact)*: online course
- 08/2025 – 12/2025 *Pathfinder Fellow (Kairos)*: fellowship providing mentorship, funding and training for AI safety group organisers
- 10/2025 – 11/2025 *AGI Strategy (BlueDot Impact)*: online course
- 02/2025 – 05/2025 *AI Safety, Ethics and Society (Center for AI Safety)*: online course
- 03/2025 *ML4Good*: 10-day AI safety bootcamp
- 03/2025 *Global Challenges Project*: workshop in Oxford

Publications

Brahmani Nutakki, **Manon Kempermann**, Ingmar Weber: You See It, They Don't: An Exploratory Study of User-to-User Variation in Instagram Comments, ICWSM '26, to appear.

Manon Kempermann, Sai Suresh Macharla Vasu, Mahalakshmi Raveenthiran, Theo Farrell, Ingmar Weber: Challenges of Evaluating LLM Safety for User Welfare. IASEAI 2026. [Preprint available](#).

Jianlong Zhu, **Manon Kempermann**, Vikram Kamath Cannanure, Alexander Hartland, Rosa M. Navarrete, Giuseppe Carteny, Daniela Braun, and Ingmar Weber. 2025. Learn, Explore and Reflect by Chatting: Understanding the Value of an LLM-Based Voting Advice Application Chatbot. In Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25). Association for Computing Machinery, New York, NY, USA, Article 58, 1–19. [Paper available](#).

Volunteering

- 04/2025 – ongoing *Founder (KI mit Demokratie)*: talks and workshops on AI and AI safety for general public and schools
- 10/2022 – 08/2023 *weltwärts Volunteer (Pakashka Sacha, Ecuador)*: supported marginalised indigenous youth in accessing tertiary education; taught English at local schools; interim-led the organisation for 3 months; established a daughter project including leading the construction of a volunteer house in an indigenous village to support primary schools in the area
- 11/2019 – 07/2022 *Violin Teacher (Musaik e.V., Dresden)*: free music education programme using music for social integration, partnered with the Sächsische Staatskapelle Dresden