

Justin Shenk, Ph.D.

AI Safety Researcher

✉ shenk.justin@gmail.com • 🌐 justinshenk.com • 📧 justinshenk • 📧 in/justinshenk • ✖ x.com/justinshenk • 🏆 Scholar

Professional Experience

- Independent AI Safety Researcher** – Berlin / London 2024 – Present
- Design frontier-model control evaluations and co-author agent benchmarks with Redwood Research (via Equistamp); LinuxArena accepted at NeurIPS 2026 (E&D Track); presented at ICML 2026 AIWILD (Oral, Best Paper).
 - Lead a mechanistic interpretability project on time representation and intertemporal steering in Qwen models (ICML 2026 MechInterp workshop).
 - Supervise 20+ early-career researchers across AI Safety Camp, SPAR, and CORDA Democracy; 4 mentee-first-author ICML 2026 workshop papers, 4 placed in AI safety roles or fellowships.
 - Facilitate BlueDot Impact's AGI Strategy, Technical AI Safety, and Frontier AI Governance courses.
- Interim CTO** – Simerse Apr. 2023 – Jul. 2023
- Led a technical team and designed CI/CD pipelines for cloud ML training and deployment on Google Cloud Platform (GCP).
 - Built research infrastructure for transformer-based models and orchestrated distributed training workflows.
 - Developed zero-shot recognition of utilities with transformer-based vision models.
- CTO, Co-Founder** – VisioLab, Berlin, Germany Jan. 2020 – Nov. 2022
- Built and led a team of 8+ engineers on a EUR 1M budget; supervised production APIs handling 1M+ transactions per month.
 - Directed development of a one-shot Computer Vision learning algorithm achieving 99.99% accuracy.
- AI Research Engineer** – Peltarion, Stockholm, Sweden Oct. 2018 – Oct. 2019
- Initiated and managed a year-long video facial analysis research project with Nordic partners.
 - Researched and optimized neural network architectures for production deployment, delivering performance benchmarks.
- AI & IoT Software Innovator** – Intel (Contractor) 2017 – 2018
- Built Machine Learning applications for the Intel Neural Compute Stick.
 - Demoed edge AI solutions at Intel booths across NeurIPS, ICML, and CVPR.

Teaching & Community Leadership

- Facilitator** – BlueDot Impact: AGI Strategy, Technical AI Safety, and Frontier AI Governance courses 2025 – Present
- Chapter Lead** – Founded the AI Salon Berlin chapter organizing monthly meetups on AI impact 2025 – Present
- Startup Mentor & Founder Coach** – Techstars, Applied Data Incubator 2021 – 2024
- Mentor** – Google Summer of Code for OpenCV 2021 – 2022
- Teacher & Tutor** – Edge Computing and Deep Learning; Intro to ML and TensorFlow at University of Osnabrück 2017 – 2021
- Co-Founder** – Open History Project; awarded the German Prize for Journalistic Engagement 2014 – 2017
- Open-Source Contributor** – Delve (neural network interpretability), Traja (behavior analysis), FER (facial expression recognition)

Education

- Ph.D. in Computational Neuroscience** 2022
Donders Institute, Radboud University (external), The Netherlands
- M.S. Cognitive Science (Artificial Intelligence)** 2016 – 2018
University of Osnabrück, Germany
- M.S., B.S. Biology** 2002 – 2008
University of Texas at San Antonio, TX

Invited Talks

- Foresight Institute** (Berlin Node), AI & Democracy 2026
- Neocyberpunk Summit**, Building for Epistemic Accountability 2026
- Venture Café Berlin**, Panel Moderator: Cybersecurity and AI; **Berlin Blockchain Week**, Panel: Content Integrity 2025
- CV in Production; #datalift Summit**, Keynote: AI solutions with limited labeled data 2022
- Additional 4 invited talks at international AI conferences including **PyData Warsaw** and **ML Prague** 2018 – 2020

Selected Publications

- LinuxArena: A Control Setting for AI Agents in Live Production Software Environments (Tracy, ..., **Shenk**, et al., *NeurIPS* 2026, Evaluations & Datasets Track; *ICML Workshop AIWILD* 2026 — Oral, Best Paper Award).
- Intertemporal Preference Steering in Qwen3 via Contrastive Activation Addition (Mráz & **Shenk**, *ICML Workshop MechInterp* 2026).
- Trajectory Data Mining in Mouse Models of Stroke (**Shenk**, Ph.D. *Dissertation*, Radboud University 2022).
- Delve: Neural Network Feature Variance Analysis (**Shenk**, Richter & Byttner, *J. Open Source Software* 7(69) 2022).
- Feature Space Saturation during Training (Richter, **Shenk**, et al., *BMVC* 2021).

Technical Skills

- Frameworks & Tools:** Python, PyTorch, HuggingFace Transformers, Docker, GCP, agentic coding workflows (Claude Code)
- Research Areas:** AI Safety Evaluations, Mechanistic Interpretability, LLMs, Computer Vision