

Justin Shenk, Ph.D.

AI Safety Researcher

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

Professional Experience

AI 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).

Research management & mentorship

- Supervised 20+ early-career researchers across AI Safety Camp, SPAR, and CORDA Democracy in 5 cohorts.
- Outcomes: 4 mentees placed in AI safety roles or fellowships.
- Invited speaker at Apart Research hackathons and research sprints: Realism in Control Benchmarks, Secret Loyalties, AI Incident Response, Digital Minds.
- Assess applicants and proposals, and facilitate BlueDot Impact's AGI Strategy, Technical AI Safety, and Frontier AI Governance courses.

Engineering & teaching

- Deliver corporate training on transformers, computer vision, and edge AI (NobleProg, Data Science Retreat), including Volkswagen autonomous-driving engineers.

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.

AI Researcher – University of Osnabrück, Germany

2016 – 2017

- Built an interactive Hilbert basis visualization (Normaliz) for the Mathematics Department.
- Implemented a real-time viewpoint-detection algorithm in an Android app for the Computer Science Department.

Senior Research Assistant – Research Imaging Institute, San Antonio, TX

2009 – 2010

- Developed software for neuropsychometric experiments and analysis of brain-activation imaging data.

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 & Presentations

Foresight Institute (Berlin Node), AI & Democracy	2026
Neocypherpunk Summit , Building for Epistemic Accountability — Berlin	2026
Computational Consciousness: CIMC X Berlin , Time: the Final Frontier (lightning talk)	2026
Venture Café Berlin , Panel Moderator: Cybersecurity and AI — Berlin	2025
Berlin Blockchain Week (Decentralized Media Summit), Panel: Content Integrity in the Age of AI — Berlin	2025
#datalift Summit , Keynote: Developing and monitoring AI solutions with limited labeled data — Berlin	2022
CV in Production , Developing and monitoring AI solutions with limited labeled data — Berlin	2022
Data Science UA , Automating Food Recognition for Canteens with Machine Vision — Kiev	2020
Nordic Data Science & ML Summit , Pitfalls of Emotion Detection in Production — Stockholm	2019
PyCon LT , Peltarion's Deep Neural Network Platform Use Cases — Vilnius	2019
ML Prague , Spectral Decomposition for Neural Network Topology Optimization (poster) — Prague	2019
PyData Warsaw , Optimizing Deep Neural Network Topology with Delve — Warsaw	2018
Intel AI Developer Conference , Benchmarking Evolutionary Algorithms (poster) — San Francisco	2018

Publications

Machine Learning & AI Safety

LinuxArena: A Control Setting for AI Agents in Live Production Software Environments – Tracy, ..., **Shenk**, et al. *NeurIPS* 2026, Evaluations & Datasets Track (Poster); also *ICML Workshop AIWILD* 2026 (Oral, Best Paper Award; Presenter)

Compute-Sovereignty Capture: A Multi-Agent Threat Model for Civilian-Control Failure – Sahoo & **Shenk**. *NeurIPS Workshop AI4Good* 2026 (Poster)

Better Call Reward: Reward Hacking as Strategic Abstention in Legal Reasoning Models – Sahoo & **Shenk**. *ICML Workshop AI4Law* 2026

Intertemporal Preference Steering in Qwen3 via Contrastive Activation Addition – Mráz & **Shenk**. *ICML Workshop MechInterp* 2026

Creative Collision: Directorial Persona Steering and Competition in Large Language Models – Sahoo & **Shenk**. *ICML Workshop GenAICreativity* 2026

Temporal Preference Concepts and their Functions in a Large Language Model – Rios-Sialer, ..., **Shenk**, et al. *BlackboxNLP* 2026 (Extended Abstract)

(Input) Size Matters for CNN Classifiers – Richter, Byttner, ..., **Shenk**. *ICANN* 2021

Feature Space Saturation during Training – Richter, **Shenk**, et al. *BMVC* 2021

Spectral Analysis of Latent Representations – **Shenk**, Richter, et al. *arXiv:1907.08589* 2019

Open-Source Software

Delve: Neural Network Feature Variance Analysis – **Shenk**, Richter & Byttner. *J. Open Source Software* 7(69) 2022

Traja: A Python Toolbox for Animal Trajectory Analysis – **Shenk** et al. *J. Open Source Software* 6(63) 2021

FER: Facial Expression Recognition – **Shenk**, CG & Arriaga. *Zenodo* 2021

Environmental & Climate Policy

Transformation and Enhancement of Climate Change Policy Indicators – Herman & **Shenk**. *Fondo Editorial FCSH* 2022

Pattern Discovery for Climate and Environmental Policy Indicators – Herman & **Shenk**. *Env. Sci. & Policy* 120 2021

Neuroscience & Biomedical

Trajectory Data Mining in Mouse Models of Stroke – **Shenk**. *Ph.D. Dissertation, Radboud University* 2022

12 further neuroscience publications (2006–2025), list available on Scholar.

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