☎ +49 152 079 60625 (EU); +1 (332) 205-1286 (US)
Website: https://georglange.com/
E-Mail: mail@georglange.com

Georg Lange

Work Experience

06/23 – 01/24 **SERIMATS** (Stanford Existential Risk Initiative, ML alignment theory scholars) scholar, Berkely (CA) and London

- Researched **Mechanistic Interpretability** for LLMs with Alex Makelov, mentored by **Neel Nanda** (3 months fulltime, 4 months part time)
- Worked on Sparse Autoencoders and Distributed Alignment Search (Geiger et al, 2023) for feature detection and subspace activation patching

06/21 – 03/23 Consultant for Data Science and Cloud Computing, Datametric, Amsterdam

- Developing Data Science solutions for major companies (Vattenfall, Ikea, T-mobile, Vesting-finance) with AWS, Sagemaker, and Azure ML (part time)

10/20 – 07/21 Student Consultant at SUGAR network, HPI, KIT, Germany

- Developed a new data-driven product from scratch for a major German insurance company; used Design Thinking and data to guide decisions; developed an App

08/19 – 11/20 Engineer (Intern and working student) for Artificial Intelligence, QiO Technologies Ltd., Potsdam, part-time

- Responsible for an Artificial Intelligence project for a British water company
- Built an end-to-end Computer Vision pipeline for damage detection in sewers, applying Transfer Learning techniques, data cleaning, hyperparameter tuning, and image visualization, implementing state-of-the-art AI research

09/18 – 10/19 Research Assistant at Digital Health Center, Hasso-Plattner-Institute

09/20 – 11/20

- Worked on the frontend for a molecular tumor board
- Developed a psychological test battery for epilepsy research

Education

09/20 – 08/21 **M.Sc. Artificial Intelligence at University of Amsterdam**

09/22 – 04/24

- Visited in-depth courses about Machine Learning, Deep Learning, Information Retrieval and did a research project about Privacy in complex-valued DNNs
- Research in complex-valued neural networks for privacy protection and equivariant spatiotemporal CNNs for scene representation learning
- **Thesis** on brain-like interpretable spatiotemporal Computer Vision models with adaptation mechanisms, supervised by **Iris Groen** and Amber Brands

08/21 – 01/23 **M.Sc. Cognitive Neuroscience** at Graduate Center and Graduate Student at **Jeff Beeler Lab**, Queens College, **CUNY**, Fulbright Scholar

- Investigating neural correlates of motivation and effort-based decision making in hyperdopaminergic (DAT-KD) and conditional D2-receptor KO (fDRD2 x Adora2a::Cre) mice using fiber photometry in Nucleus Accumbens
- **Thesis** on interaction between dopamine and acetylcholine during cocaine- or amphetamine-induced drug sensitization
- Developed a software package for fiber photometry data analysis: https://fibermagic.org/
- Developed a platform to control operand boxes and neural data acquisition systems wirelessly: https://github.com/Goreg12345/magicbox

- Hired and trained two undergraduate students on experimental neurobiology

10/17 – 09/20 **B.Sc. IT-Systems-Engineering**, Hasso-Plattner-Institute, Potsdam

- Graduated with 1.5: “very good”
- Took part in a one-year long Connected Health Care project about Unobstrusive Health Monitoring for Wearable Devices
- **Thesis**: “Detecting Several Types of Distractions During Work Using Wireless EEG-Devices Applying Machine Learning Techniques”

07/17 A-Levels (Abitur) passed with final grade 1.1, “very good”

Programming Skills and Artificial Intelligence Knowledge

- Develop, debug and train neural networks using **Pytorch**, Lightning, WandB, Tensorflow
- Cloud Computing (**AWS**) to scale AI / ML models using Sagemaker, Athena, EMR
- Mechanistic Interpretability for transformers using transformer-lens, Pytorch, einops
- Python with **Pandas, NumPy, SciPy**, Sklearn, Matplotlib, Plotly, Seaborn, Statsmodels
- User-oriented product management using Design Thinking, rapid prototyping, interviews
- Kotlin for Android and Ktor, Java, SQL
- C, C++ for Operating Systems, Arduino, algorithms and data structures
- Building interactive devices using laser cutting, 3D printing, electronics, and CV

Neuroscience Skills

- Manage a transgenic mouse colony (300 animals) including PCR-based genotyping
- Stereotactic brain surgery including viral injections and fiber implantation
- Measure neural activity using dual-color fiber photometry and manufacturing implants
- Perfusion, cryostatic or vibratome slicing, immunohistochemistry
- Running behavioral experiments including conditioning, operand boxes and IP-injections
- Developing, maintaining, and optimizing behavioral setups using 3D-printing, Computer Vision, software development, and electronics
- Measuring, cleaning, and analyzing EEG data

Awards and Achievements

06/23 & 09/23 Research grants from AI safety support for SERIMATS and its extension

02/21 – 09/22 Scholar of the German-American **Fulbright Program**

4/22 & 11/22 Research Award and Assistantship of the Cognitive Neuroscience program at Graduate Center, CUNY

02/20 – 03/22 Organized a buddy program between students and pupils from non-academic families: www.senkrechtstarter.org

01/18 – 01/23 Scholar of **Konrad-Adenauer-Foundation**, Political Foundation

- Financial stipend that covers cost of living
- Participated in several seminars and many one-day events

10/19 – 04/20 **Led and created a four-week Online Course** (MOOC) about Deep Learning, Neural Nets, and image recognition for open.HPI

- Currently 12900 enrollments on open.hpi.de/courses/neuralnets2020

- Member of the **commission of studies** for two years; revised B.Sc. IT-Systems Engineering and developed M.Sc. Cybersecurity, HPI
- Challenge winner of **Hack Zurich**, Europe’s largest Hackathon with >1000 participants

Publications

5/24 Alex Makelov* & Georg Lange*, Atticus Geiger, Neel Nanda. (2024). **Is This the Subspace You Are Looking for? An Interpretability Illusion for Subspace Activation Patching**. ICLR. https://doi.org/10.48550/arXiv.2311.17030

5/24 Alex Makelov* & Georg Lange*, Neel Nanda. (2024). **Towards Principled Evaluations of Sparse Autoencoders for Interpretability and Control.** SeT LLM (ICLR).

05/21 Arsen Sheverdin, Alko Knijff, Noud Corten, & Georg Lange. (2021). [Re] **Reproducibility report of "Interpretable Complex-Valued Neural Networks for Privacy Protection"**. Rescience C, 7(2), #20.