Nino Scherrer

Mail: nino.scherrer@gmail.com

Website: ninodimontalcino.github.io
Github: ninodimontalcino

EDUCATION

M.S. Computer Science *Sep 2019 – Sep 2022*
ETH Zurich, Switzerland

B.S. Computer Science *Sep 2015 – Sep 2019*
ETH Zurich, Switzerland

RESEARCH EXPERIENCE

Research Scientist, Patronus AI (Full-Time) — Zurich, Switzerland, *Nov 2023 – ongoing*

- Working on scalable approaches for language model evaluations (e.g., automated construction of large-scale evaluation datasets, automated red-teaming, uncertainty quantification in evaluations)

Independent Researcher (Part-Time) — Zurich, Switzerland, *Jul 2023 – ongoing*

- Investigating cognitive biases / theory of mind capabilities in LLMs (in collaboration with Stanford University and University of Warsaw)
- Collaborated with Google Research and ETH Zurich on a mechanistic interpretability project (in final preparation for journal submission)

Visiting Researcher, FAR AI / Columbia University (Full-Time) — New York, United States, *Oct 2022 – Jun 2023*

- Evaluated the moral beliefs encoded in large language models (NeurIPS Spotlight Paper)
- Collaborated with Claudia Shi, Amir Feder, and Prof. David Blei

Visiting Researcher, Vector Institute (Full Time) — Toronto, Canada, *Jun 2022 – Aug 2022*

- Developed methods for visual scene understanding/causal reasoning
- Supervised by Prof. Animesh Garg

Visiting Researcher, Mila Quebec AI Institute (Full-Time) — Montreal Canada, *Nov 2021 – May 2022*

- Investigated the synergies of causal structure in machine learning models on out-of-distribution generalization (ICML Workshop Paper)
- Advised a project on causal experimental design (NeurIPS Paper)
- Conducted a systematic review of neural causal discovery (in Submission)
- Supervised by Prof. Yoshua Bengio and Nan Rosemary Ke (DeepMind)

Thesis Student, Max Planck Institute for Intelligent Systems — Tubingen, Germany
Apr 2021 - Oct 2021

- Designed an experimental design method for causal structure learning (NeurIPS Workshop Paper)
- Contributed to a Bayesian causal discovery method (ICML Workshop Paper)
- Supervised by: Prof. Stefan Bauer

Thesis Student + Follow-Up Research, ETH Zurich (Part-Time) — Zurich, Switzerland
Sep 2019 - Sep 2020

- Developed a simulation pipeline to generate synthetic stroke MR images
- Contributed to a journal paper on brain stroke segmentation (Radiology: AI)
- Supervised by: Dr. Christian Federau

WORK EXPERIENCE (NON-RESEARCH)

Data Analyst / Civil Servant, University Hospital of Zurich (Full-Time) — Zurich Switzerland
Sep 2020 - Apr 2021

- Implemented a COVID monitoring tool to prevent nosocomial infections
- Developed a tool for automated vaccine dose planning for high-risk patients

Software Engineer, Self-Employed (Part-Time during studies) — Zurich, Switzerland
Sep 2015 - Oct 2021

- Various individual projects (full stack, web, and search engine optimization)

Systems Engineer, SFS Group AG (Full-Time) — Heerbrugg, Switzerland
Aug 2009 - Jun 2015

- Process Automation and Systems Engineering on server/client level

PUBLICATIONS

> CONFERENCE PAPERS / JOURNAL ARTICLES

[1] **Evaluating the Moral Beliefs Encoded in LLMs,** Scherrer, N.*, Shi, C.*, Feder, A., & Blei, D., *NeurIPS 2023* ***Spotlight***, 2023

[2] **Trust Your ∇ : Gradient-based Intervention Targeting for Causal Discovery,** Olko, M.*, Zając, M.*, Nowak, A.*, Scherrer, N., Annadani, Y., Bauer, S., Kuciński, L. & Miłoś, L. , *NeurIPS 2023*, 2023

[3] **Radial Matrix Constraint Influences Tissue Contraction and Promotes Maturation of Bi-Layered Skin Equivalents**, Polak, J., Sachs, D., Scherrer, N., Süess, A., Liu, H., Levesque, M., Werner, S., Mazza, E., Restivo, G., Meboldt, M. & Giampietro, C., *Biomaterials Advances*, 2023

[4] **Improved Segmentation and Detection Sensitivity of Diffusion-Weighted Stroke Lesions with Synthetically Enhanced Deep Learning**, Federau, C., Christensen, S., Scherrer, N., Ospel, J., Schulze, V., Schmidt, N., Breit, H., Maclaren, J., Lansberg, M. & Kozerke, S., *Radiology: Artificial Intelligence,* 2020

> WORKSHOP PAPERS

[5] **On the Generalization and Adaption Performance of Causal Models,** Scherrer, N., Goyal, A., Bauer, S., Bengio, Y. & Ke, N.R., *ICML 2022 - SCIS and BeyondBayes Workshop*, 2022

[6] **Learning Neural Causal Models with Active Interventions,** Scherrer, N., Bilaniuk, O., Annadani, Y., Goyal, Y., Schwab, P., Schölkopf, B., Mozer, M.C., Bengio, Y., Bauer, S. & Ke, N.R., *NeurIPS 2021 - WHY-21 Workshop,* 2021

[7] **Variational Causal Networks: Approximate Bayesian Inference over Causal Structures,** Annadani, Y., Rothfuss, J., Lacoste, A., Scherrer, N., Goyal, A., Bengio, Y. & Bauer, S., *KDD 2021, **Oral** at Workshop on Bayesian causal inference for real world interactive systems*, 2021

> PREPRINTS / PAPERS IN SUBMISSION

[8] **Uncovering Mesa-Optimization Algorithms in Transformers,** von Oswald, J.*, Niklasson, E.*, Schlegel, M.*, Kobayashi, S., Zuccet, N., Scherrer, N., Miller, N., Sandler, M., Vladymyrov, M., Agüera y Arcas, B., Pascanu, R., & Sacramento, J., *Arxiv Preprint – In Preparation for Journal Submission*, 2023

[9] **Deep Learning for Causality: A Unifying Perspective on Neural Causal Structure Learning,** Scherrer, N., Annadani, Y., Bauer, S., Goyal, A., Ke, N.R. & Bengio, Y., *In Final Preparation*, 2023

[10] **Federated Causal Discovery From Interventional and Observational Data**, Abyaneh, A., Scherrer, N., Schwab, P., Bauer, S., Schölkopf, B. & Mehrjou, A., *Arxiv Preprint – In Conference Submission*, 2023

[11] **FinanceBench: A New Benchmark For Financial Question Answering**, Islam, P.*, Kannappan, A.*, Kiela, D.*, Qian, R.*, Scherrer, N.*, & Vidgen, B.*, *Arxiv Preprint – In Preparation for Submission,* 2023

[12] **SimpleSafetyTests: A Test Suite for Identifying Critical Safety Risks in Large Language Models**, Vidgen, B., Scherrer, N., Kirk, H.R., Qian, R., Kannappan, A, Hale, S.A. & Röttger, P., *Arxiv Preprint – Under Review,* 2024

INVITED TALKS

- **Evaluating Beliefs Encoded in LLMs,** Ada Lovelace Institute, London, 2024 (Upcoming)
- **Evaluating The (Moral) Beliefs Encoded in LLMs,** ML/AI Meetup, Zurich, 2023
- **Deep Learning for Causality,** AI for Actional Impact Lab, Imperial College, London, 2023
- **On the Synergies of Causality and Deep Learning,** Neuroscience in ML, ETH Zurich, 2022
- **Learning Neural Causal Models with Active Interventions,** Explainable AI, Imperial College, 2021
- **Learning Neural Causal Models with Active Interventions,** Causality Group, TU Darmstadt, 2021

PRESS COVERAGE

- **FinanceBench:** CNBC and Fortune,
- **SimpleSafetyTests:** VentureBeat and ComputerWorld,
- **Evaluating the Moral Beliefs Encoded in LLMs:** Heise Podcast (German)

AWARDS / SCHOLARSHIPS / GRANTS

- **NeurIPS Spotlight Paper**, Awarded to the top 3% paper submissions
- **NeurIPS Scholar Award,** Covering registration and hotel costs for NeurIPS Conf. in New Orleans, 2023
- **Unrestricted Research Grant** (7500 CAD), Vector Institute, 2022
- **Research Scholarship**, Covering costs of internship with Prof. Yoshua Bengio, Mila AI Institute, 2021
- **Apprenticeship Award**, Merit-based award for vocational diploma, Hans Huber Foundation, 2013
- **Golden Book Entry,** Merit-based award for best graduation of class, SFS Group AG, 2013
- **Promotion Prize,** Merit-based award for vocational graduation, Hilti AG, 2013

ACADEMIC SERVICE / MENTORSHIP

> **Reviewing:** JMLR (2022), ICML (2022, 2023), NeurIPS (2022), ICML, & NeurIPS Workshops (2022, 2023)

> **Teaching Assistance**: Information Retrieval (ETH Zurich, 2019)

> **Volunteering:** First Year Student Day at ETH Zurich (2016, 2017)

> **Current Student Mentorship:**
- Mariia Minaeva (PhD Student @ Technical University of Munich / Helmholtz AI)
- Gracjan Goral (PhD Student @ University of Warsaw)

> **Past Student Mentorship:**
- Amin Abyaneh (Intern @ MPI Tubingen, now PhD Student @ McGill University)

SKILLS

> **Programming:**	Python, SQL
> **Scripting:**	Bash, PowerShell
> **Frameworks:**	PyTorch, NumPy
> **General:**	Ethics, Cognitive Science, Neuroscience
> **Social:**	I love collaborating and mentoring!

LANGUAGES

> **German:**	Native
> **English:**	Proficient
> **French:**	Basic
> **Slovak:**	Basic

REFERENCES

Prof. Stefan Bauer
Associate Professor
TU Munich / Helmholtz AI
st.bauer@tum.de

Nan Rosemary Ke
Senior Research Scientist
DeepMind
nke@google.com

Amir Feder
Postdoctoral Fellow / Faculty Researcher
Columbia University / Google Research
amir.feder@columbia.edu