Dr. Timo Freiesleben

Curriculum Vitae

Google Scholar
Personal Website

Research Interests

Areas of Specialization	Explainable AI, Philosophy of Science, Philosophy of AI
Areas of Competence	Ethics of AI, Causality, Decision Theory, Logic

Positions

10/2022–Now **Research Fellow**, *Machine Learning for Science Cluster, Eberhard Karls Universität Tübingen*, "Certification and Foundations of Safe Machine Learning Systems in Healthcare" funded by the Carl Zeiss Foundation

Education

10/2019–05/2023 **Ph.D. in Neurophilosophy**, *Graduate School of Systemic Neurosciences München (LMU)*,
Thesis on **What Does Explainable AI Explain?**
Thesis advisory committee: Prof. Dr. Stephan Hartmann (main supervisor), Dr. Álvaro Tejero Cantero, Prof. Dr. Agnieszka Wykowska, Prof. Dr. Paul Taylor
External reviewer: Prof. Dr. Jan-Willem Romeijn

10/2018–09/2019 **M.Sc. in Computer Science**, *Ludwig-Maximilians-Universität München (LMU)*,
Taken courses on Deep Learning & AI
Without Graduation

10/2016–09/2018 **M.A. in Logic and Philosophy of Science**, *Munich Center for Mathematical Philosophy (MCMP), Ludwig-Maximilians-Universität München (LMU)*,
Very Good,
Thesis on **Incorporating Intuitions into Decision Making Rationally**
Supervised by Dr. Rush Stewart and Prof. Dr. Dr. Hannes Leitgeb

10/2012–09/2016 **B.Sc. in Mathematics**, *Eberhard Karls Universität Tübingen*,
Very Good,
Thesis on **Ramification and Arithmetic Schemes**
Supervised by Prof. Dr. Jürgen Hausen

08/2015–07/2016 **Erasmus Exchange Year**, *University of Oslo*,
With a focus on Mathematical Logic and Computability Theory

09/2010–07/2012 **Abitur**, *in the Wirtschaftsoberschule at the KS-Künzelsau*, *Very Good*

Publications

Journal Publications

2024 **Scientific Inference With Interpretable Machine Learning: Analyzing Models to Learn About Real-World Phenomena**, *Minds and Machines*
Freiesleben, T., König, G., Molnar, C., & Tejero-Cantero, A.

2023 **Beyond Generalization: A Theory of Robustness in Machine Learning**, *Synthese*
Freiesleben*, T. & Grote*, T.

2022 **The Intriguing Relation Between Counterfactual Explanations and Adversarial Examples**, *Minds and Machines*
Freiesleben, T.

Books

2024 **Supervised Machine Learning for Science: How to stop worrying and love your black box** , *https://ml-science-book.com/*
Molnar*, C. & Freiesleben*, T.

2023 **What Does Explainable AI Explain?**, *Dissertation LMU Munich*
Freiesleben, T.

Book Chapters

Forthcoming **Artificial Neural Nets and the Representation of Human Concepts**, *Philosophy of Science for Machine Learning: Core Issues and New Perspectives*, edited by Juan Durán and Giorgia Pozzi, Synthese Library
Freiesleben, T.

2022 **General pitfalls of model-agnostic interpretation methods for machine learning models**, *In Lecture Notes on Artificial Intelligence 13200 xxAI — Beyond explainable AI, Cham. Springer International Publishing.*
Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B.

Conference Publications

2024 **CountARFactuals – Generating plausible model-agnostic counterfactual explanations with adversarial random forests**, *Proceedings of World XAI Conference*
Freiesleben*, T., Dandl*, S., Blesch*, K., König*, K., Kapar, J., Bischl, B., & Marvin Wright

2023 **Dear XAI Community, we Need to Talk! Fundamental Misconceptions in Current XAI Research**, *Proceedings of World XAI Conference*
Freiesleben, T., & König, G.

2023 **Relating the Partial Dependence Plot and Permutation Feature Importance to the Data Generating Process**, *Proceedings of World XAI Conference*
Freiesleben*, T., Molnar*, C., König*, G., Herbinger, J., Reisinger, T., Casalicchio, G., Wright, M. N., & Bischl, B.

2023 **Improvement-Focused Causal Recourse (ICR)**, *Proceedings of AAAI Conference on Artificial Intelligence*
König, G., Freiesleben, T., & Grosse-Wentrup, M.

2021 **A causal perspective on meaningful and robust algorithmic recourse**, *ICML 2021 workshop on Algorithmic Recourse*
König, G., Freiesleben, T., & Grosse-Wentrup, M.

Teaching

Main Instructor

Explainable Artificial Intelligence, *MCMP & Statistics Department*, Jointly with Gunnar König, Winter Term 21/22

Causality and Machine Learning, *Statistics Department*, Jointly with Gunnar König and Susanne Dandl, Sommer Term 21

Philosophy of Artificial Intelligence, *MCMP*, Jointly with Prof. Stephan Hartmann, Winter Term 20/21

Ethics of Artificial Intelligence, *Statistics Department*, Jointly with Florian Pfisterer, Christoph Molnar, Gunnar König, and Susanne Dandl, Winter Term 20/21

Teaching Assistant

Formal Methods II: Models and Simulations, *MCMP*, Led by Dr. Rush Stewart, Summer Term 20

Central Topics in Philosophy of Science, *LMU*, Led by Dr. Jürgen Landes, Winter Term 19/20

Linear Algebra 1, *Mathematics Department*, Led by Dr. Peter Philip, Winter Term 18/19

Linear Algebra 2, *Mathematics Department*, Led by Prof. Dr. Fabien Morel, Summer Term 18

Linear Algebra 1, *Mathematics Department*, Led by Prof. Dr. Fabien Morel, Winter Term 17/18

Topology and multivariable differential calculus, *Mathematics Department*, Led by Prof. Dr. Franz Merkl, Summer Term 17

Analysis 1, *Mathematics Department*, Led by Prof. Dr. Franz Merkl, Winter Term 16/17

Academic Service and Organization

Reviewing **Synthese, Nature Machine Intelligence, FAccT, Minds and Machines, Philosophy of Science, Erkenntnis, Studies in History and Philosophy of Science, Philosophy & Technology, Austrian Science Fund (FWF), European Journal for Philosophy of Science, ICML workshop, NeurIPS workshop, World XAI Conference**

Workshop Organizer **Philosophy of Science Meets Machine Learning**, *Tübingen University*, 11-13 September 2024, Tübingen
jointly with Konstantin Genin, Thomas Grote, Markus Ahlers, Raysa Benatti, & Sebastian Zezulka

Tübingen, Hannover, and Friends Network Workshop, *Tübingen University*, 04-05 March 2024, Tübingen
jointly with Thomas Grote & Sebastian Zezulka

Philosophy of Science Meets Machine Learning, *Tübingen University*, 12-14 September 2023, Tübingen
jointly with Thomas Grote, Konstantin Genin & Sebastian Zezulka

LMU-Cambridge Workshop, *Topic: "AI in Science: Foundations and Applications"*, 9-10 June 2022, Munich
jointly with Stephan Hartmann & Tom Sterkenburg

Reading Group Organizer

ML Cluster Tübingen, *Topic: "Philosophy of Machine Learning"*, since winter term 2022/2023, Tübingen
jointly with Sebastian Zezulka and Benedikt Höltgen

MCMP, *Topic: "Philosophy of Machine Learning"*, since summer term 2022, online
jointly with Tom Sterkenburg

Skills

Languages: German (native speaker), English (fluent), Spanish (very good command), Norwegian (good command).

Computer Skills: Python (++), LaTeX(+++), MATLAB/Octave (++), Java (++), R (+), NetLogo (++), JavaScript (++), HTML (++), PHP (+), WebPPL (+), SQL (+).

Awards & Grants

10/2019–09/2022 **Graduate School of Systemic Neuroscience**, *Ph.D. fellowship*, 52,560€

25/07/2019 **Mobility Innovation Competition @ Campus**, *3rd prize in Startup competition*, Team: DeepGuardian, 8,000€
Deep-learning-software equipped camera board for violence detection that respects data privacy.

07/2018 **Oskar-Karl-Forster-Scholarship**, *book stipend*, 500€

06/2012 **School-Prize**, *best Abitur*

Conferences, Workshops, Talks, etc.

17/07/2024-19/07/2024 **World XAI Conference**, *Malta*, Oral Presentations on "CountARFactuals – Generating plausible model-agnostic counterfactual explanations with adversarial random forests"

28/05/2024 **Höchstleistungsrechenzentrum Stuttgart (HLRS)**, *Stuttgart*, Panelist at workshop on generalization and overfitting

23/05/2024 **Applied AI: seminar**, *online*, Talk on "Scientific Inference With Interpretable Machine Learning: Analyzing Models to Learn About Real-World Phenomena"

10/04/2024-12/04/2024 **ML, Explain Yourself!**, *Utrecht University*, Presentation with Sebastian Zezulka on "Science as a Kaggle Challenge: How Benchmarking Impacts Scientific Methodology"

02/02/2024 **Höchstleistungsrechenzentrum Stuttgart (HLRS)**, *Stuttgart*, Paper discussion of "Beyond Generalization: A Theory of Robustness in Machine Learning"

08/11/2023-10/11/2023 **AI policy**, *Utrecht University, chaired by Emily Sullivan*, Whitepaper to inspire AI policy

2/11/2023 **UQSay seminar series**, *Paris-Saclay University (online)*, Invited presentations on "Supervised Machine Learning in Science"

24/10/2023-26/10/2023 **Artificial Intelligence, Trustworthyness, Explainability (AITE) Conference**, *Tübingen*, Presentations on "Dear XAI Community, We Need to Talk! Fundamental Misconceptions in Current XAI Research"

19/10/2023 **Ethical Engineering**, *Munich*, Bidt: Bavarian Research Institute for Digital Transformation, Panelist

21/08/2023-25/08/2023 **European Congress of Analytic Philosophy**, *Vienna*, Oral Presentation on "Beyond Generalization: A Theory of Robustness in Machine Learning"

26/07/2023-28/07/2023 **World XAI Conference**, *Lisbon*, Two Oral Presentations on "Dear XAI Community, We Need to Talk! Fundamental Misconceptions in Current XAI Research" and "Relating the Partial Dependence Plot and Permutation Feature Importance to the Data Generating Process"

14/06/2023 **Helmholtz AI Conference**, *Hamburg*, Panelist on "When do we blindly trust in AI?"

11/05/2023-12/05/2023 **Tübingen-Hannover Network Workshop: Philosophy of Machine Learning**, *University of Hannover*, Presentation on "Contesting Counterfactual Explanations", Co-organizer on the Tübingen side

23/03/2023-24/03/2023 **Epistemology and Theory of Machine Learning**, *MCMP*, Invited Speaker on "Beyond Generalization: A Theory of Robustness in Machine Learning", Munich

07/02/2023-12/02/2023 **AAAI Conference**, *Washington D.C.*, Oral Presentation & Poster on "Improvement-Focused Causal Recourse (ICR)"

20/01/2023 **10 minutes Talk Series**, *ML Cluster Tübingen*, Talk on "What Does Explainable AI Explain?"

27/10/2022-28/10/2022 **Workshop: Responsible Machine Learning in Healthcare**, *University of Copenhagen*, Poster on "What Does Explainable AI Explain?"

19/10/2022-22/10/2022 **Workshop: Philosophy of Science Meets Machine Learning**, *University of Tübingen*, Presentation on "Scientific Inference With Interpretable Machine Learning"

30/06/2022-01/07/2022 **Hannover-MCMP-Wuppertal Network Workshop: Philosophy of Science**, *University of Wuppertal*, Presentation on "Scientific Inference With Interpretable Machine Learning"

21/06/2022-24/06/2022 **FAccT Conference**, *Online Participation*

13/06/2022 **Science Summit of the Joint Research Centre of the European Commission**, *Topic: Science through the AI lens*, Panelist

09/06/2022-10/06/2022 **LMU-Cambridge Strategic Partnership Workshop, Topic: "AI in Science: Foundations and Applications"**, *Presentation on "Scientific Inference With Interpretable Machine Learning"*

09/11/2021-12/11/2021 **Workshop: Philosophy of Science Meets Machine Learning**, *University of Tübingen*, Presentation on "To Explain and to Predict – Explanatory Machine Learning Models in Science"

24/07/2021 **ICML workshop, Algorithmic Recourse**, *Online Event*, Poster on A Causal Perspective on Meaningful and Robust Algorithmic Recourse

19/05/2021 **MCMP-colloquium talks**, *Talk: "Embrace the Complexity: The Paradigm Shift in Science From Statistics to Machine Learning"*, München, Germany (Online Event), Jointly with Christoph Molnar

12/04/2021-14/04/2021 **NIAS-workshop, Explainable Medical AI: Ethics, Epistemology, and Formal Methods**, *Leiden, the Netherlands (Online Event)*

17/07/2020 **ICML workshop, XXAI: Extending Explainable AI Beyond Deep Models and Classifiers**, *Vienna, Austria (Online Event)*, Poster on Pitfalls to Avoid when Interpreting Machine Learning Models

29/06/2020-03/07/2020 **Summerschool: Regularization Methods for Machine Learning**, *Genova, Italy (Online Event)*, Led by Prof. Lorenzo Rosasco

17/02/2020 - 18/02/2020 **Workshop on Machine Learning: Prediction Without Explanation?**, *Karlsruhe (KIT)*, Talk on Counterfactual Explanations & Adversarial Examples

14/01/2020 **CTPS course**, *MCMP*, Topic: The Wisdom of Crowds, Guest Lecture

27/07/2018 - 28/07/2018 **Workshop on Decision Theory & the Future of Artificial Intelligence**, *München (Jointly organized by the MCMP, the CFI, and the CSER)*

Non-Academic Work

03/2019–09/2019 **Software Developer (working student)**, *Zentrum Digitalisierung.Bayern*, Garching,

Project: Working on the national research project MEMAP which contributes to the German energy transition strategy. MEMAP (Multi-Energy Management and Aggregation-Platform) optimally matches the local electricity- and heat demand/production for districts

Tasks: My work focused mainly on the software development of the platform in the programming language Java. In particular, I had the following tasks:

- programming the OPC-UA interfaces for handling live-data
- developing a Jetty-websocket and a website for online access to the platform (HTML, Javascript,etc.)
- configuration of server data for providing optimization results

References

Prof. Dr. Stephan Hartmann
Chair and Head of the Munich Center for Mathematical Philosophy
Department of Philosophy, Philosophy of Science and the Study of Religion
Ludwig-Maximilians-Universität München
Contact no: + 49 (0) 89 / 2180 - 3320
Email: S.Hartmann@lmu.de

Dr. Thomas Grote
Research Fellow – Ethics and Philosophy Lab
Cluster of Excellence – Machine Learning for Science
Eberhard Karls Universität Tübingen
Email: thomas.grote@uni-tuebingen.de

Prof. Dr. Jan-Willem Romeijn
Professor of Philosophy of science
Faculty of Philosophy
University of Groningen
Contact no: +31 50 36 36148
Email: j.w.romeijn@rug.nl

Dr. Álvaro Tejero-Cantero
Group Leader of the ML - Science Colaboratory
Cluster of Excellence – Machine Learning for Science
Eberhard Karls Universität Tübingen
Contact no: +49 176 2431 1515
Email: alvaro.tejero@uni-tuebingen.de

Dr. Rush Stewart
Lecturer
Department of Philosophy
King's College London
Email: rush.stewart@kcl.ac.uk