Dr. rer. nat. Sebastian Lapuschkin (*né Bach*)

* December 16, 1986 in Würzburg

Fraunhofer Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI
Einsteinufer 37, 10587 Berlin
http://www.hhi.fraunhofer.de
sebastian.lapuschkin@hhi.fraunhofer.de • +49 (30) 31002-371 • +49 (177) 483-2754

github.com/sebastian-lapuschkin • linkedin.com/in/sebastian-lapuschkin •
scholar.google.com/citations?user=wpLQuroAAAAJ

Short Bio

Sebastian Lapuschkin is the Head of the Explainable Artificial Intelligence research group at Fraunhofer Heinrich Hertz Institute (HHI) in Berlin.

He received his Ph.D. degree with distinction from the Berlin Institute of Technology in 2018 for his pioneering contributions to the field of Explainable Artificial Intelligence (XAI) and interpretable machine learning. From 2007 to 2013 he studied computer science (B. Sc. and M. Sc.) at the Berlin Institute of Technology, with a focus on software engineering and machine learning.

Sebastian is the recipient of multiple awards, including the Hugo-Geiger-Prize for outstanding doctoral achievement and the 2020 Pattern Recognition Best Paper Award.

His work is focused on pushing the boundaries of XAI, e.g, for achieving human-understandable explanations, and towards the effective and efficient utilization of interpretable feedback for the improvement of machine learning systems and data.

Since 2024 he is co-organizing The World Conference on eXplainable Artificial Intelligence and serves as a Topic Editor on "Opportunities and Challenges in Explainable Artificial Intelligence" for the MDPI Open Access Journals.

Further research interests include efficient machine learning and data analysis, as well as data and algorithm visualization.

Professional Experience

Multidisciplinary Digital Publishing Institute (MPDI)

Topic Editor *2024 - 2026*
for "Opportunities and Challenges in Explainable Artificial Intelligence".
Submission pre-screening, review management and decision handling.

XAI4Science

Organizer *2024 - 2025*
of the Workshop "XAI4Science: From Understanding Model Behavior to Discovering New Scientific Knowledge (2025)", co-located with ICLR 2025 at Singapore EXPO, Singapore.

World Conference on eXplainable Artificial Intelligence

Steering Committee Member *2024 - 2025*
for the 3rd XAI World Conference (2025) in Istanbul, Turkey.
Conference and Special Track co-organization.

Programme Committee Chair *2023 - 2024*
for the 2nd XAI World Conference (2024) in Valetta, Malta.
Conference and Special Track co-organization.

Fraunhofer Heinrich-Hertz-Institute — Berlin, Germany

Ethics Committee Member *2023 -*
Founding member of the first ethics committee at Fraunhofer HHI.

Head of Explainable Artificial Intelligence *2021 -*
Research Group Leadership and direction of XAI research & applications.

Tenured Researcher *2019 - 2020*
PostDoc research position in the Machine Learning Group at Fraunhofer HHI.

Research Associate *2014 - 2018*
Founding member of the Machine Learning Group at Fraunhofer HHI.

Berlin Institute of Technology — Berlin, Germany

Research Associate *2013 - 2014*
Supervision by Prof. Dr. Klaus-Robert Müller and Prof. Dr. Alexander Binder.

Student Research- & Teaching Assistant *2011 - 2013*
Research assistant to Prof. Dr. Alexander Binder at the machine learning group at TU Berlin.
Teaching assistant to Prof. Dr. Klaus-Robert Müller, Prof. Dr. Dr. Franz Király,
Dr. Irene Dowding (née Winkler) and Dr. Daniel Bartz.

Student Teaching Assistant *2009 - 2011*
Teaching assistant to Prof. Dr. Marc Alexa, Prof. Dr. Odej Kao and Prof. Dr. Oliver Brock.

Education

Berlin Institute of Technology — Berlin, Germany

PhD in Machine Learning (*summa cum laude*) *2018*
Date of oral defense: December 19th, 2018.
Dean's signature on Doctorate Certificate dated January 23rd, 2019.

Thesis: "Opening the machine learning black box with Layer-wise Relevance Propagation"
Supervision headed by Prof. Dr. Klaus-Robert Müller.

Master of Science in Computer Science *2013*
Focus on machine learning, computer vision and large scale data analysis.

Bachelor of Science in Computer Science *2010*
Focus on algorithms and software development

Deutschhaus-Gymnasium — Würzburg, Germany

Abitur (pre-university secondary education) *2007*

Teaching

Teaching of and teaching support for 19 university courses since 2009.

Talks & Lectures

Over 31 invited talks and individual lectures held since 2017.
Excludes teaching activities and internal/confidential events.

Third-Party Funded Research Projects

7 third-party funded research projects acquired and managed since 2018

Honors & Awards

Machine Learning and the Physical Sciences Reproducibility Badge *2024*
Stanford Top 2% Scientist Worldwide* *2021 - 2023*
Best Short Paper Award *2023*
Pattern Recognition Best Paper Award and Pattern Recognition Medal *2020*
Hugo-Geiger-Prize (1st place) *2019*
Freunde des HHI Nachwuchspreis *2019*
ERCIM Cor van Baayen Award (finalist) *2019*
Best Paper Award *2016*

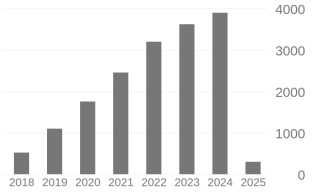
Patents

A Concept Representation of a Machine Learning Model *2024*
Relevance Score Assignment dealing with an Attention Module and Applications thereof *2024*
Analyzing an Inference of a Machine Learning Predictor *2023*
Method and System for Simulating an Optical Image of a Photonic and/or Electronic Device *2022*
Pruning and/or Quantizing Machine Learning Predictors *2020*
Relevance Score Assignment for Artificial Neural Networks *2016*

Publications

Summary of Scientific Impact

	All	Since 2020
# Publications	77	57
# Citations	17436	15292
h-index	33	31
i10-index	50	48



per Google Scholar, retreived on January 27th, 2025.

Selected Publications

1. Achtibat R, Dreyer M, Eisenbraun I, Bosse S, Wiegand T, Samek W and **Lapuschkin S** (2023). "From attribution maps to human-understandable explanations through Concept Relevance Propagation". In: *Nature Machine Intelligence* 5(9):1006–1019.
   https://github.com/rachtibat/zennit-crp | https://github.com/maxdreyer/crp-human-study
   A paper introducing the second generation of Explainable Artificial Intelligence with concept-based explanations.

2. Anders C J, Weber L, Neumann D, Samek W, Müller K-R and **Lapuschkin S** (2022). "Finding and Removing Clever Hans: Using Explanation Methods to Debug and Improve Deep Models". In: *Information Fusion* 77:261–295.
   The authors' first work of a series dedicated to the exploitation of knowledge derived from XAI to the improvement of performance and robustness of AI systems.

3. **Lapuschkin S**, Wäldchen S, Binder A, Montavon G, Samek W and Müller K-R (2019). "Unmasking Clever Hans Predictors and Assessing what Machines Really Learn". In: *Nature Communications* 10:1069.
   One of the first papers to rigorously perform model- and data analysis through the lens of XAI, adding a voice of caution to the ongoing excitement about machine intelligence.

4. Montavon G, **Lapuschkin S**, Binder A, Samek W and Müller K-R (2017). "Explaining NonLinear Classification Decisions with Deep Taylor Decomposition". In: *Pattern Recognition* 65:211–222.
   A paper discussing the mathematical foundation of LRP and its properties. Pattern Recognition Best Paper Award and Pattern Recognition Medal winner of 2020.

5. **Bach S**, Binder A, Montavon G, Klauschen F, Müller K-R and Samek W (2015). "On Pixel-wise Explanations for Non-Linear Classifier Decisions by Layer-wise Relevance Propagation". In: *PLoS ONE* 10(7):e0130140.
   A very influential and early work on local XAI, introducing the widely used Layer-wise Relevance Propagation method. This work has received over 5300 citations as counted by Google Scholar.