

Johannes von Oswald

jvoswald@google.com | research.google/people/johannesvonoswald

I am a research scientist at Google Research working with Blaise Agüera y Arcas and João Sacramento at the Paradigms of Intelligence Team. Until 2023, I was a PhD student supervised by Angelika Steger and João Sacramento at the Institute of Theoretical Computer Science, ETH Zurich. My research is focused on learning algorithms for neural networks to tackle problems with continuous and sparse data. Under these assumptions, the trade-off between model bias and sample efficiency lies at the heart of my interests in biological and artificial neural networks, statistical learning theory and reinforcement learning. Currently, I am working on mesa-optimization: the emergence of algorithms within neural networks, in [paper link 1](#) and [paper link 2](#) we showed that gradient descent is a mechanism learned by pure self-supervised learning which allows transformers to learn and generalize at test time to novel data provided in-context. I am currently working on incorporating these findings into large language models.

Education

ETH Zurich

Zurich | 2018 - 2023

PhD Student at the Institute of Theoretical Computer Science
Advisor: Prof. Angelika Steger & Dr. João Sacramento
Deep learning, in particular continual, meta and mesa-optimization.

Technical University Munich

Munich | 2014 - 2017

Master in Mathematics
Advisor: Prof. Daniel Cremers
Thesis: Bayesian Optimal Flow Prediction with Neural Networks
I studied mathematics with a minor in computer science and focused on computer vision, inverse problems and dynamical systems.

ETH Zurich

Zurich | Sommer 2016

Visiting student under the Swiss-European Mobility Programme.
I worked at the Institute for Visual Computing of Prof. Marc Pollefeys on neural networks based object recognition incorporating depth information.

Hongkong University of Science and Technology

Hongkong | Autumn 2015

TU Munich exchange program. HKUST Dean's List for outstanding attainment.

Technical University Berlin

Berlin | 2011 - 2014

Bachelor in Mathematics
Advisor: Prof. Yuri Suris
Thesis: Special Solutions of the N-body Problem.
I studied mathematics with a minor in physics and a focus on differential geometry and mathematical physics.

Fellowships & Awards

Swiss Data Science Center Fellowship (\approx CHF 200.000 - 4 years of PhD salary)	2019 - 2023
Google Cloud Research Grant (\$ 5000)	2021
ZNZ Travel Grant (CHF 1000)	2021
Pi School of Artificial Intelligence (EUR 2000)	2018

Tasks & Poster Presentations

ASAP Seminar - MesaNet: Sequence Modeling by Locally Optimal Test-Time Training virtual	June 2025
Oral presentation at ICLR - Learning Randomized Algorithms with Transformers Singapore	April 2025
Talk at the Helmholtz-ELLIS Workshop on Foundation Models in Science on modern RNNs Berlin	March 2025
Tutorial at CoLLAs 2024 on the mechanics of in-context learning Pisa	July 2024
Talk at the Alignment Workshop - Mechanistic Interpretability of in-context learning New Orleans	Dec 2023
Talk at Mila, Montreal/ServiceNow - Uncovering mesa-opt. algorithms in Transformers virtual	Oct 2023
Talk at ChatGPT Zurich Group - In-context learning in large language models Zurich	Sep 2023
Oral presentation at NeurIPS - Transformers learn in-context by gradient descent Hawaii	July 2023
Talk at SDSC - On Transformers and how they can learn in-context Zurich	April 2023
Talk at Google Brain, San Francisco - On Transformers and how they can learn in-context virtual	Feb 2023
Talk at Jagiellonian University, Krakau - Transformers learn in-context by gradient descent virtual	Jan 2023
Talk at Sapienza University, Rome - Transformers learn in-context by gradient descent virtual	Jan 2023
Talk at Google DeepMind - Transformers learn in-context by gradient descent virtual	Dec 2022
Talk at Google DeepMind - On HyperNetworks and the prefrontal cortex London	Aug 2022
Poster presentation at NeurIPS - Learning where to learn Offsite at EPFL	Dec 2021
Poster presentation at ICML - Meta-learning via hypernetworks virtual	July 2021
Talk at the Swiss Data Science Center - Hypernetworks for sparse and continuous data virtual	Aug 2020
Poster presentation at ICLR - Continual learning with hypernetworks virtual	June 2020
Spotlight poster presentation at NeurIPS - On adversarially-trained hypernetworks Montreal	Dec 2018

Selected Publications & Preprints

- ¹J. von Oswald, S. Kobayashi, Y. Akram, and A. Steger, “Learning randomized algorithms with transformers”, in ICLR - Oral presentation (2025).
- ²J. von Oswald, N. Scherrer, S. Kobayashi, L. Versari, S. Yang, M. S. et al., B. A. y Arcas, and J. Sacramento, “Mesanet: sequence modeling by locally optimal test-time training”, arXiv (2025).
- ³S. Kobayashi, Y. Akram, and J. V. Oswald, “Weight decay induces low-rank attention layers”, in NeurIPS (2024).
- ⁴J. von Oswald, E. Niklasson, E. Randazzo, J. Sacramento, A. Mordvintsev, A. Zhmoginov, and M. Vladymyrov, “Transformers learn in-context by gradient descent”, in ICML - Oral presentation (2023).
- ⁵J. von Oswald, M. Schlegel, A. Meulemans, S. Kobayashi, E. Niklasson, et al., R. Pascanu, and J. Sacramento, “Uncovering mesa-optimization algorithms in transformers”, arXiv (2023).
- ⁶A. Meulemans, N. Zucchet, S. Kobayashi, J. von Oswald, and J. Sacramento, “The least-control principle for learning at equilibrium”, in NeurIPS - Oral presentation (2022).
- ⁷N. Zucchet, S. Schug, J. von Oswald*, D. Zhao, and J. Sacramento, “A contrastive rule for meta-learning”, in NeurIPS (2022).
- ⁸J. von Oswald, S. Kobayashi, and et al., “Neural networks with late-phase weights”, in ICLR (2021).
- ⁹J. von Oswald, D. Zhao, S. Kobayashi, S. Schug, M. Caccia, N. Zucchet, and J. Sacramento, “Learning where to learn: Gradient sparsity in meta and continual learning”, in NeurIPS (2021).
- ¹⁰J. von Oswald, C. Henning, B. F. Grewe, and J. Sacramento, “Continual learning with hypernetworks”, in ICLR - Spotlight presentation (2020).

(* first author)

Teaching Experience (TA)

Algorithms and complexity	2022 & 2023
Algorithms and probability	2022 & 2023
Learning in deep artificial and biological neuronal networks	2019 & 2020 & 2021

Supervision

Julius Schulte (PhD Thesis) - Weight averaging through HyperNetworks
Oliver Sieberling & Yanick Schimpf (Deep learning project) - Swiss deep delta net
Vinay Hiremath (Master thesis) - Differentiable gated linear networks
Levan Varamashvili (Master thesis) - On how Transformers invert matrices
Laurent Verdan (Bachelor thesis) - Grokking beyond simple addition
Maximilian Schlegel (Bachelor thesis) - Transformers solve linear autoregressive tasks by gradient descent
Damian Cordes (Bachelor thesis) - On sparse continual meta-learning
Yuchen Chang (Bachelor thesis) - Understanding warm starting with the neural tangent kernel

Reviewing / Scientific Services

Area Chair - Neural Information Processing Systems (NeurIPS)	2025
Reviewer - Neural Information Processing Systems (NeurIPS), Top Reviewer in 2023	2021-2025
Reviewer - Conference on Lifelong Learning Agents (CoLLAs)	2023
Reviewer - International Conference on Machine Learning (ICML)	2021-2025
Reviewer - Transactions on Machine Learning Research (TMLR)	2022-2024
Reviewer - International Conference on Learning Representations (ICLR)	2021-2025
Reviewer - Workshop on Continual Learning in Computer Vision (CVPR)	2020 & 2021

Industry & Working Experience (Selection)

Google Research - Research Scientist Research on sequence modeling, transformers, reinforcement learning and mechanistic interpretability of deep learning models.	Zurich Sommer 2023 - today
--	------------------------------

Google Research - PhD Intern at Google Research Internship at Google in Seattle supervised by Max Vladymyrov.	Seattle Sommer 2023
---	-----------------------

Google Research - Research student at Google Research Research student at Google Zurich supervised by Alexander Mordvintsev.	Zurich Autumn 2022
--	----------------------

Pi School - School of Artificial Intelligence Awardee of a Pi school scholarship for an eight weeks hands-on machine learning course for industry applications.	Rome Sommer 2018
---	--------------------

Celonis - Business Process Mining Intern and working student developing tools for automated process creation for SAP databases.	Munich Sommer 2015
---	----------------------

Groupon - E-Commerce Marketplace Working student developing and implementing algorithms to automate long term return estimation of investments in online advertisement according to conversion rates.	Berlin 2012 - 2013
---	----------------------

Nice Bar - Bar Co-Owner and Barkeeper.	Berlin & Brussels 2011 - 2013
--	---------------------------------

Ronzino von Oswald SA - Textile Production Co-founder and board advisor of textile production and sourcing company based in Berlin and Porto, Portugal.	Berlin & Porto 2010 - Now
---	-----------------------------