Sweta Mahajan

✉ sweta.mahajan@mpi-inf.mpg.de
LinkedIn & Github & Twitter & Website

Education

PhD in Computer Science, *Max Planck Institute for Informatics, Germany*, supervised by Prof. Bernt Schiele, and Prof. Alexander Koller. Nov 2023- present

- First cohort of the Research Training Group, Neuroexplicit Models of Language, Vision, and Action.
- Research focus: Building reliable and trustworthy large vision-language models.

BS-MS Dual Degree in Mathematics and Statistics, *Indian Institute of Science Education and Research (IISER) Kolkata, India*, supervised by Prof. Anirvan Chakraborty. 2015 - 2020

- Master's thesis: Hidden Markov Modelling in Prokaryotic Gene Finding and Sequence Alignment.

Research Experience

Research Assistant, Yardi School of AI, *Indian Institute of Technology (IIT) Delhi, India*, supervised by Prof. Chetan Arora and Prof. Parag Singla. Aug 2021 - Oct 2023

Publications

[1] TEVI: Text-Conditioned Editing of Visual Representations via Sparse Autoencoders for Improved Vision-Language Alignment, ***Sweta Mahajan****, *Sukrut Rao**, *Jiahao Xie, Alexander Koller, Bernt Schiele*, EMNLP (Findings) 2026. [Paper] 2026

[2] Discover-then-Name: Task-Agnostic Concept Bottlenecks via Automated Concept Discovery, *Sukrut Rao**, ***Sweta Mahajan****, *Moritz Böhle, Bernt Schiele*, ECCV 2024. [Paper][Code] Also presented as a poster at WiCV, CVPR 2025. 2024

[3] Neural Analogical Reasoning, *Atharv Sonwane, Abhinav Lalwani,* ***Sweta Mahajan****, Gautam Shroff, Lovekesh Vig*, International Workshop on Neural-Symbolic Learning and Reasoning (NeSy) 2022. [Paper][Code] 2022

* equal contribution

Talks

- Talk at Multimodal AI group, TU Darmstadt. 2026
- Talk at CVML group, Tübingen AI Center. 2026
- Talk at DWS-Colloquium, University of Mannheim. 2024

Academic Services

- **Area Chair**: CVPR 2027.
- **Organizer**: eXCV@ECCV Workshop 2026, eXCV@ICCV Workshop 2025.
- **Reviewer**: ICLR 2027, ECCV 2026, CVPR 2026, XAI4CV@CVPR workshop 2024–2025, eXCV@ECCV Workshop 2024, WiCV@CVPR 2025.
- **Sub-Reviewer**: ICML 2024, ICML 2025, ICML 2026.

Student Supervision

- Ankit Khatiwada Nov 2026 – Present
- Tuğçe Türkmenler Nov 2026 – Present
- Tejas Dhopavkar Mar 2026 – Present

Teaching Assistant

- Explainable Machine Learning Seminar in **Saarland University**, Winter 2026.
- Undergraduate course on Machine Learning (COL341) in **Indian Institute of Technology** Delhi, Spring 2023.
- Undergraduate course on Machine Learning (COL774) in **Indian Institute of Technology** Delhi, Autumn 2022.
- Undergraduate course on Intro. to Artificial Intelligence, **Indian Institute of Technology**, Delhi, Autumn 2021.
- Postgraduate courses in **Indian Institute of Management**, Calcutta, July 2020 to July 2021.
- Undergraduate course on Probability in **IISER** Kolkata, Spring 2020.
- Undergraduate course on Python programming for computational Physics in **IISER** Kolkata, Autumn 2019.

Research Schools and Conferences

- International Computer Vision Summer School (ICVSS), Sicily, Italy. [Website] 2024
- Research Week with Google, Bangalore, India. [Website] 2022 & 2023

Awards and Scholarships

- Awarded a fully funded PhD position in the first cohort of the DFG-funded Research Training Group "Neuroexplicit Models of Language, Vision, and Action" (RTG 2853), Saarland Informatics Campus. 2023
- Ranked All India 111 (top 0.36%) out of 31144 students in National Eligibility Test (NET) in Mathematics, and received a 5-year scholarship from the Govt. of India. 2020
- Ranked All India 110 (top 8%) out of 1346 students in Graduate Aptitude Test in Engineering (GATE) in Statistics. 2019
- Ranked All India 36 (top 0.7%) out of 5161 students in KVPY Examination - received scholarship from the Govt. of India to fund my bachelors and masters. 2015

Skills

Programming: Python, Bash, Git, LaTeX.

ML libraries: PyTorch, Torchvision, Hugging Face, OpenCV, Detectron2.

Research areas: Vision-language models, interpretability, sparse autoencoders, concept bottleneck models, computer vision.

Mathematical background: Probability, statistics, optimization, combinatorics, machine learning, deep learning.

Compute: Multi-GPU training, SLURM-managed HPC clusters.

Leadership & Outreach

- PhD student representative for the RTG 2853 "Neuroexplicit Models of Language, Vision, and Action", Saarland Informatics Campus. Oct 2026 – Present
- Volunteer for Ek-Pehal, a non-profit initiative providing education to local children. Organized outreach activities in local villages, taught classes and coordinated events at IISER Kolkata. 2015–2020
- Founding member and Public Relations Officer of the Entrepreneurship Cell, IISER Kolkata. 2017

Last updated: November 20, 2026