Sukrut Rao

sukrutrao.github.io | sukrut.rao@mpi-inf.mpg.de | Google Scholar | GitHub | LinkedIn

Education

Max Planck Institute for Informatics and Saarland University — Saarbrücken, Germany
Ph.D. in Computer Science, focusing on building trustworthy and reliable vision models — *2021 – 2026*
Advisor: Prof. Dr. Bernt Schiele

Max Planck Graduate Center for Computer and Information Science — Saarbrücken, Germany
Ph.D. Preparatory Phase (Graduate Coursework) — *2019 – 2021*

Indian Institute of Technology Hyderabad — Hyderabad, India
Bachelor of Technology (Honors) in Computer Science and Engineering — *2015 – 2019*
GPA: 9.95/10.00, President of India Gold Medal

Research and Industry Experience

Amazon — Melbourne, Australia
Applied Scientist Intern — *Jun 2025 – Dec 2025*
Developed an end-to-end pipeline for identifying and fixing defects in product images from the Amazon product catalog. Delivered a working pipeline using state-of-the-art vision-language models that achieved high precision across highly diverse product types and was used for a live A/B test on amazon.com.au.

Research Training Group on Neuroexplicit Models of Language, Vision, and Action — Saarbrücken, Germany
Associated Member — *Jul 2023 – Present*
Building robust and interpretable vision and language models combining neural and explicit components, e.g. [5], [6], [7].

Machine Intelligence Laboratory, The University of Tokyo — Tokyo, Japan
Research Intern — *May 2018 – Jul 2018*
Co-developed, in a team of students, an improved method for the detection of object-object and object-attribute relationships from images. Our team secured the 6^{th} position out of 232 teams in the Google AI Open Images Challenge 2018.

Bosch Engineering and Business Solutions — Bengaluru, India
Research Intern — *May 2017 – Jul 2017*
Developed automated deep learning-based methods to diagnose among five grades of diabetic retinopathy from fundus images.

Publications

Peer-Reviewed

[1] CFM: Language-aligned Concept Foundation Model for Vision — ECCV 2026 (Spotlight)
Kai Wittenmayer, **Sukrut Rao**, Amin Parchami-Araghi, Bernt Schiele, Jonas Fischer

[2] Align Once to Explain: Feature Alignment for Scalable B-cosification of Foundational Vision Transformers — CVPR 2026
Raphael Maser*, Siddhartha Gairola*, **Sukrut Rao**, Bernt Schiele

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

[4] FaCT: Faithful Concept Traces for Explaining Neural Network Decisions — NeurIPS 2025
Amin Parchami-Araghi, **Sukrut Rao**, Jonas Fischer, Bernt Schiele

[5] B-cos LM: Efficiently Transforming Pre-trained Language Models for Improved Explainability — TMLR 2025
Yifan Wang, **Sukrut Rao**, Ji-Ung Lee, Mayank Jobanputra, Vera Demberg

[6] B-cosification: Transforming Deep Neural Networks to be Inherently Interpretable — NeurIPS 2024
Shreyash Arya*, **Sukrut Rao***, Moritz Böhle*, Bernt Schiele

[7] Discover-then-Name: Task-Agnostic Concept Bottlenecks via Automated Concept Discovery — ECCV 2024
Sukrut Rao*, Sweta Mahajan*, Moritz Böhle, Bernt Schiele

[8] Good Teachers Explain: Explanation-Enhanced Knowledge Distillation — ECCV 2024
Amin Parchami-Araghi*, Moritz Böhle*, **Sukrut Rao***, Bernt Schiele

*Authors contributed equally

[9] Better Understanding Differences in Attribution Methods via Systematic Evaluations — TPAMI 2024
Sukrut Rao, Moritz Böhle, Bernt Schiele

[10] Studying How to Efficiently and Effectively Guide Models with Explanations — ICCV 2023
Sukrut Rao*, Moritz Böhle*, Amin Parchami-Araghi, Bernt Schiele

[11] Towards Better Understanding Attribution Methods — CVPR 2022
Sukrut Rao, Moritz Böhle, Bernt Schiele

[12] Adversarial Training Against Location-Optimized Adversarial Patches — ECCV Workshops 2020
Sukrut Rao, David Stutz, Bernt Schiele

[13] Open-WBO-Inc: Approximation Strategies for Incomplete Weighted MaxSAT — JSAT 2019
Saurabh Joshi†, Prateek Kumar†, **Sukrut Rao**†, Ruben Martins†

[14] Approximation Strategies for Incomplete MaxSAT — CP 2018
Saurabh Joshi†, Prateek Kumar†, Ruben Martins†, **Sukrut Rao**†

[15] Fast-Dawid-Skene: A Fast Vote Aggregation Scheme for Sentiment Classification — WISDOM @ KDD 2018
Vaibhav B Sinha, **Sukrut Rao**, Vineeth N Balasubramanian

Preprints

[16] DisParQ: Self-Supervised Part Concepts for Interpretable Vision Foundation Models — arXiv 2026
Adam Pardyl*, Siddhartha Gairola*, **Sukrut Rao**, Adam Wróbel, Bartosz Zieliński, Bernt Schiele, Dawid Rymarczyk

[17] AnyBottle: A Recipe to Only Keep the Concepts You Really Need — arXiv 2026
Wolfgang Stammer, **Sukrut Rao**, Hevra Petekkaya, David Steinmann, Bernt Schiele

Talks

Google — Munich, Germany
Stanford University — Stanford, USA
UC Berkeley — Berkeley, USA
UC San Diego — La Jolla, USA
Northeastern University — Boston, USA
Boston University — Boston, USA
MIT CSAIL — Cambridge, USA
University of Washington — Seattle, USA
Towards Improving Multimodal Models using Interpretability Tools — *Jun 2026*

Multimodal AI Lab, TU Darmstadt — Darmstadt, Germany
Understanding and Improving Multimodal Models through the Lens of Interpretability — *May 2026*

Computer Vision and Multimodal Learning Un-Workshop, Tübingen AI Center — Tübingen, Germany
Bringing Interpretability to Vision-Language Foundation Models — *May 2026*

Guide Labs — San Francisco, USA
Towards Inherent Interpretability for Foundation Models — *Oct 2025*

IIIT Summer School on AI — Hyderabad, India
Explainable Computer Vision: Feature Attributions, Concepts, and Beyond — *Jul 2025*

Symposium on Explainable Artificial Intelligence Beyond Simple Attributions — Mainz, Germany
Keeping the Model in the Loop: Attribution Methods and Beyond — *Nov 2024*

Teaching and Student Supervision

Master Theses

- Görkem Yiğit — *Nov 2026 – Present*
- Raphael Maser, *Towards Inherent Interpretability of Foundation Vision Models* — *Aug 2025 – Mar 2026*
  Extended into [2], CVPR 2026
- Kai Wittenmayer, *Family-CBM: Automatic Discovery of Hierarchies in Concept Bottleneck Models* — *Mar 2025 – Dec 2025*
  Extended into [1], ECCV 2026 (Spotlight)

*Authors contributed equally
†Authors listed in alphabetic order, or grouped by affiliation

- Shreyash Arya, *Increasing Interpretability of Deep Neural Networks via B-cosification* — *Dec 2023 – Jun 2024*
  Extended into [6], NeurIPS 2024
- Amin Parchami-Araghi, *A Good Teacher Explains: Explanation-enhanced Knowledge Distillation* — *Dec 2023 – Jun 2024*
  Extended into [8], ECCV 2024

Other Supervision

- Tejas Dhopavkar (Student Assistant), *Benchmarking Reasoning of Multimodal Models* — *Mar 2026 – Present*
- Moussa Herrmann (Bachelor Thesis), *Refining Concept Bottlenecks with Symbolic Relations* — *Jul 2025 – Oct 2025*
- Nhi Pham (Research Immersion Lab), *Class Concept Discovery with B-cos Networks* — *Nov 2023 – Feb 2024*

Teaching Assistantships

- **Saarland University:** High-Level Computer Vision (Summer 2023, Summer 2024), Explainable Machine Learning Seminar (Winter 2025–26, Winter 2026–27)
- **IIT Hyderabad:** Applied Machine Learning (Spring 2019), Optimization Methods in Machine Learning (Fall 2018), Compilers-II (Fall 2018), Principles of Programming Languages-II (Spring 2018), Principles of Programming Languages-I (Fall 2017)

Academic Service

Organizer

- Explainable Computer Vision: Challenges and Opportunities in the Era of Foundation Models Workshop at ECCV 2026 — Malmö, Sweden
- Saarbrücken Symposium on Interpretable, Trustworthy, and Accountable Artificial Intelligence 2026 — Saarbrücken, Germany
- Explainable AI for Computer Vision (XAI4CV) Workshop at CVPR 2026 — Denver, USA
- Explainable Computer Vision: Quo Vadis? Workshop at ICCV 2025 — Honolulu, USA
- Explainable AI for Computer Vision (XAI4CV) Workshop at CVPR 2025 — Nashville, USA
- Explainable Computer Vision: Where are We and Where are We Going? Workshop at ECCV 2024 — Milan, Italy
- Explainable AI for Computer Vision (XAI4CV) Workshop at CVPR 2024 — Seattle, USA

Reviewer: CVPR 2023–26 (2x Outstanding Reviewer); ICCV 2023, 2025; NeurIPS 2023, 2025–26 (1x Top Reviewer); ICLR 2024, 2027; ICML 2024–26; ECCV 2024, 2026; IEEE TPAMI; XAIA @ NeurIPS 2023

Awards and Achievements

- Outstanding Reviewer (≈ top 5%) from CVPR 2026. — 2026
- Top Reviewer (≈ top 10%) from NeurIPS 2023. — 2023
- Outstanding Reviewer (≈ top 3.3%) from CVPR 2023. — 2023
- President of India Gold Medal from IIT Hyderabad for securing the highest GPA among graduating students. — 2019
- Institute Silver Medal from IIT Hyderabad for securing the highest GPA among graduating CSE students. — 2019
- Certificate of Appreciation in Research 2018–19 from IIT Hyderabad. — 2019
- Sixth position out of 232 teams in the Google AI Open Images Challenge 2018. — 2018
- First position secured by our Open-WBO-Inc MaxSAT solver ([13], [14]) in the 60s category in the weighted incomplete MaxSAT track at the MaxSAT Evaluation 2018. — 2018
- Academic Excellence Award 2017–18 from IIT Hyderabad. — 2018
- Honda Young Engineer and Scientist Award 2017 from Honda Foundation Japan (USD 10,000). — 2018
- Academic Excellence Award 2015–16 from IIT Hyderabad. — 2016
- All India Rank 668 in Joint Entrance Examination (JEE) (Advanced) 2015 out of ≈130,000 candidates. — 2015
- All India Rank 34 in Joint Entrance Examination (JEE) (Main) 2015 out of ≈1.3 million candidates. — 2015
- Top 0.1% in Mathematics and Physics in the All India Senior School Certificate Examination (Class XII). — 2015
- National Talent Search Scheme (NTS) Scholarship. — 2013