Curriculum Vitae/Resume

Ekdeep Singh Lubana
Email: eslubana@umich.edu

Education

Ph.D. Candidate, University of Michigan, Ann Arbor *August, 2019–ongoing*
Co-affiliated with Center for Brain Science, Harvard University
Advisors: Robert P. Dick and Hidenori Tanaka

B.Tech., Indian Institute of Technology, Roorkee *July, 2015–May, 2019*
Major: Electronics and Communication Engineering
Thesis: Resource Efficient Techniques for Embedded Machine Vision (Nominated for Best Bachelor's Thesis)

Areas of Interest

· AI Alignment, Science of Deep Learning, Interpretability

Experience

· **Research Intern**, Qualcomm AI Research, Amsterdam *June., 2023–Nov., 2023*
Mentors: Taco Cohen, Johann Brehmer, and Pim de Haan

· **Research Affiliate**, Krueger AI Safety Lab *Aug, 2022–Present*
Mentor: David Krueger

· **Research Intern**, Bell Labs Cambridge, UK *Sept., 2021–Dec., 2021*
Mentor: Akhil Mathur

· **Research Intern**, Physics and Informatics Lab, NTT Research Inc. *May, 2021–Aug., 2021*
Mentor: Hidenori Tanaka

Publications (* denotes equal contribution)

1. Usman Anwar, Abulhair Saparov*, Javier Rando*, Daniel Paleka*, Miles Turpin*, Peter Hase*, **Ekdeep Singh Lubana***, Erik Jenner*, Stephen Casper*, Oliver Sourbut*, Benjamin Edelman*, Zhaowei Zhang*, Mario Gunther*, Anton Korinek*, Jose Hernandez-Orallo*, Lewis Hammond, Eric Bigelow, Alexander Pan, Lauro Langosco, Tomasz Korbak, Heidi Zhang, Ruiqi Zhong, Seán Ó hÉigeartaigh, Gabriel Rachet, Giulio Corsi, Alan Chan, Markus Anderljung, Lillian Edwards, Yoshua Bengio, Danqi Chen, Samuel Albanie, Tegan Maharaj, Jakob Foerster, Florian Tramer, He He, Atoosa Kasirzadeh, Yejin Choi, and David Krueger. Foundational challenges in assuring alignment and safety of large language models. *arXiv preprint arXiv:2404.09932*, 2024

2. Samyak Jain*, Robert Kirk*, **Ekdeep Singh Lubana***, Robert P. Dick, Hidenori Tanaka, Edward Grefenstette, Tim Rocktaschel, and David Krueger. Mechanistically Analyzing the Effects of Fine-Tuning on Procedurally Defined Tasks. In *Proc. Int. Conf. on Learning Representations (ICLR)*, 2024.

3. Eric Bigelow, **Ekdeep Singh Lubana**, Robert P. Dick, Hidenori Tanaka, and Tomer Ullman. In-Context Learning Dynamics with Random Binary Sequences. In *Proc. Int. Conf. on Learning Representations (ICLR)*, 2024.

4. Rahul Ramesh, **Ekdeep Singh Lubana**, Mikail Khona, Robert P. Dick, and Hidenori Tanaka. Compositional Capabilities of Autoregressive Transformers: A Study on Synthetic, Interpretable Tasks. *arXiv preprint arXiv:2311.12997*, 2024. (In submission.)

5. Mikail Khona, Maya Okawa, Jan Hula, Rahul Ramesh, Kento Nishi, Robert P. Dick, **Ekdeep Singh Lubana***, and Hidenori Tanaka*. Towards an Understanding of Stepwise Inference in Transformers: A Synthetic Graph Navigation Model. *arXiv preprint arXiv:2402.07757*, 2024. (In submission.)

6. Maya Okawa*, **Ekdeep Singh Lubana***, Robert P. Dick, and Hidenori Tanaka*. Compositional Abilities Emerge Multiplicatively: Exploring Diffusion Models on a Synthetic Task. In *Proc. Adv. in Neural Information Processing Systems (NeurIPS)*, 2023.

7. **Ekdeep Singh Lubana**, Eric J Bigelow, Robert P. Dick, David Krueger, and Hidenori Tanaka. Mechanistic Mode Connectivity. In *Proc. Int. Conf. on Machine Learning (ICML)*, 2023.

8. **Ekdeep Singh Lubana**, Johann Brehmer, Pim de Haan, and Taco Cohen. FoMo Rewards: Can we cast foundation models as reward functions? In *NeurIPS Foundation Models for Decision Making Workshop*, 2023

9. Liu Ziyin, **Ekdeep Singh Lubana**, Masahito Ueda, and Hidenori Tanaka. What Shapes the Loss Landscape of Self-Supervised Learning? In *Proc. Int. Conf. on Learning Representations (ICLR)*, 2023.

10. Puja Trivedi and **Ekdeep Singh Lubana**, Mark Heimann, Danai Koutra, and Jay Jayaraman Thiagarajan. Analyzing Data-Centric Properties for Contrastive Learning on Graphs . In *Proc. Adv. in Neural Information Processing Systems (NeurIPS)*, 2022.

11. **Ekdeep Singh Lubana**, Ian Tang, Fahim Kawsar, Robert P. Dick, and Akhil Mathur. Orchestra: Unsupervised Federated Learning via Globally Consistent Clustering. In *Proc. Int. Conf. on Machine Learning (ICML)*, 2022. (Accepted for **Spotlight** presentation.)

12. **Ekdeep Singh Lubana**, Robert P. Dick, and Hidenori Tanaka. Beyond BatchNorm: Towards a Unified Understanding of Normalization in Deep Learning. In *Proc. Adv. in Neural Information Processing Systems (NeurIPS)*, 2021.

13. **Ekdeep Singh Lubana** and Robert P. Dick. A Gradient Flow Framework for Analyzing Network Pruning. In *Proc. Int. Conf. on Learning Representations (ICLR)*, 2021. (Accepted for **Spotlight** presentation.)

14. **Ekdeep Singh Lubana**, Puja Trivedi, Danai Koutra, and Robert P. Dick. How do Quadratic Regularizers Prevent Catastrophic Forgetting: The Role of Interpolation. In *Proc. Conf. on Lifelong Learning Agents (CoLLAs)*, 2022.

15. **Ekdeep Singh Lubana**, Robert P. Dick, Vinayak Aggarwal, and Pyari Mohan Pradhan. Minimalistic Image Signal Processing for Deep Learning Accelerators. In *Proc. Int. Conf. on Image Processing (ICIP)*, 2019.

16. **Ekdeep Singh Lubana**, Vinayak Aggarwal, and Robert P. Dick. Machine Foveation: An Application-Aware Compressive Sensing Framework. In *Proc. Data compression Conference (DCC)*, 2019.

17. **Ekdeep Singh Lubana** and Robert P. Dick. Digital Foveation: An Energy-Aware Machine Vision Framework. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, pages 2371–2380, 2018.

Technical Service

· Reviewer for NeurIPS, ICML, ICLR, AISTATS, IEEE TPAMI, IEEE TNNLS *2021–present*
· **Top Reviewer**, NeurIPS. *2023*
· **Top Reviewer**, ICLR. *2022*
· **Top Reviewer**, NeurIPS. *2022*

Technical Awards

· Awarded the **BIRAC-GYTI award** by the **President of India.** *2018*
· Winner of the **Ericsson Innovation Challenge** held at the Nobel Museum, Stockholm, Sweden. *2017*
· Winner of the **Jury's choice award** at the **Accenture Innovation Challenge.** *2017*
· **Gold medal** and **winner of Engineers' Conclave** at **Inter-IIT Tech meet.** *2018*

Academic achievements & Scholarships

· Awarded the **KVPY (Kishore Vaigyanik Protsahan Yojna)** Fellowship by Govt. of India. *2015*
· Awarded the **NTSE (National Talent Search)** Scholarship by N.C.E.R.T., New Delhi. *2014*
· Ranked amongst **Top 300** students in **National Standard Examination in Astronomy.** *2015*
· Ranked amongst **Top 300** Students in the **Indian National Mathematics Olympiad.** *2015*