Alexey Kravets

Contact Info: +393405523595, alexey.kravets012@gmail.com **Right to work**: Italian citizen, UK settlement status holder **Google scholar**: link, **Medium**: https://medium.com/@alexml0123, **GitHub**: https://github.com/akres001

Publications

Alexey Kravets, Da Li, Chuan Li, Da Chen, Vinay P. Namboodiri "Interpretability Transfer from Language to Vision via Sparse Autoencoders", International Conference on Machine Learning (**ICML**), 2026

- We align textual Sparse Autoencoders with visual representations transferring interpretability to large vision-language models (VLMs) at minimal cost, enabling both interpretability and steering of concepts within VLMs.

Alexey Kravets, Da Chen, Vinay P. Namboodiri "Rethinking Few Shot CLIP Benchmarks: A Critical Analysis in the Inductive Setting", International Conference on Computer Vision (**ICCV**), 2025

- We identified a flaw in evaluation of existing few-shot methods with CLIP and proposed a pipeline to evaluate them more fairly using unlearning.

Alexey Kravets, Vinay P. Namboodiri "Zero-shot CLIP class forgetting via text-image space adaptation", Transactions on Machine Learning Research (TMLR), 2025

- We propose a class removal technique for CLIP without using any images, but only text of the class to forget. The method relies on changing the text projection matrix in CLIP.

Alexey Kravets, Vinay P. Namboodiri "Zero-Shot Class Unlearning in CLIP with Synthetic Samples", The Winter Conference on Applications of Computer Vision (**WACV**), 2025

- We propose a method to unlearn classes from CLIP without real data. We do it by generating synthetic samples and using Lipschitz regularization to unlearn.

Olga Loginova, Oleksandr Bezrukov, Ravi Shekhar, **Alexey Kravets** "Addressing Blind Guessing: Calibration of Selection Bias in Multiple-Choice Question Answering by Video Language Models", The 63rd Annual Meeting of the Association for Computational Linguistics (**ACL**), 2025

- We show that Video-Language models suffer from positional bias in Multiple-Choice Question Answering. We then introduce a post-processing calibration technique based on fairness bias metrics.

Alexey Kravets, Vinay P. Namboodiri "CLIP Adaptation by Intra-modal Overlap Reduction", British Machine Vision Conference (**BMVC**) - **Oral**, 2024

- We improve the performance of few-shot training-free learning methods that use CLIP model by reducing the intra-model overlap in the image encoder of CLIP.

Work Experience

PhD Student, University of Bath

April 2023 - Present
My current research primarily focuses on vision and language models, few-shot learning, machine unlearning and more recently mechanistic interpretability. I published first author papers in venues such as ICCV, WACV, BMVC and TMLR journal. I have also co-authored a paper accepted in ACL.

Lead Data Scientist, Aviva

April 2018 - April 2023
In Aviva I've led significant machine learning projects in the Healthcare claims team where we were developing NLP tools for claims predictions.

Educational Background

BSc Economics and Statistics (2013-2016)

University of Florence (107/110)

MSc Quantitative Finance (2016-2018)

University of Florence (110/110 cum laude)

MSc Artificial Intelligence (2021-2023)

University of Bath (Distinction)

Skills

- Python (PyTorch, Transformers) and some experience in R, C++.
- Deployment: Dataiku, Flask, Gradio.

Languages

Italian (Native), Russian (Native), English (Proficient), Polish (A2)