Andrii Shkabrii

andrii.shkabrii@univie.ac.at | github.com/shkabrii | Google Scholar | LinkedIn

Research Interests

My research interests lie in representation learning, optimization, and reinforcement learning. I focus on understanding how neural networks organize information in the embedding space and developing methods with theoretical guarantees to improve robustness and fairness of learned representations. I am interested in applying these ideas to AI safety, including unlearning, deception detection, and interpretability.

Education

M.Sc. Computer Science, Graduated with Distinction *October 2022 - March 2025* | Vienna, Austria

University of Vienna

- Thesis: *Investigating Plasticity Loss During the Self-Labeling Stage in Deep Clustering*

B.Sc. Computer Science *October 2019 - July 2022* | Vienna, Austria

University of Vienna

- Thesis: *Dynamic Feature Acquisition Strategies for Incomplete Data*

Work Experience

PhD Student *April 2025 - Present* | Vienna, Austria

University of Vienna, Research Group Data Mining and Machine Learning | Supervisor: Prof. Claudia Plant

- Developing HSIC-based methods for decomposing neural network representations, with provable guarantees on robustness and fairness of learned embeddings.
- Studying theoretical bounds linking embedding dimensionality to adversarial robustness, and extending to enforce fairness criteria such as demographic parity.
- Working on robustness of NLP classifiers (BERT-family, LLMs, guardrail models), investigating adversarial attacks targeting rationale vs. non-rationale parts of inputs.

Quantitative Developer *March 2023 - October 2024* | Remote

Private Crypto Trading Firm

- Designed low-latency, real-time distributed trading systems; built automated execution and margin management for crypto derivatives across 20+ accounts and 70+ assets.
- Built an LLM-based agentic system that translated natural language trading strategies into executable code, integrated with a custom backtesting engine.

Student Research Assistant *Sep 2022 - Dec 2022; Aug 2024 - Oct 2024* | Vienna, Austria

University of Vienna

- Investigated plasticity loss in deep learning; ran large-scale experiments on EuroHPC supercomputers. Contributed to deep clustering research (ICLR 2025, DSAA 2023).

Publications

*denotes equal contribution

Understanding and Improving Hyperbolic Deep Reinforcement Learning
T. Klein*, T. Lang*, **A. Shkabrii**, A. Sturm, K. Sidak, L. Miklautz, C. Plant, Y. Velaj, S. Tschiatschek
International Conference on Learning Representations (ICLR), 2026

H-SPLID: HSIC-based Saliency Preserving Latent Information Decomposition
L. Miklautz*, C. Shi*, **A. Shkabrii***, T. T. Davarakis, P. Lam, C. Plant, J. Dy, S. Ioannidis
Conference on Neural Information Processing Systems (NeurIPS), 2025

Breaking the Reclustering Barrier in Centroid-based Deep Clustering
L. Miklautz*, T. Klein*, K. Sidak*, C. Leiber, T. Lang, **A. Shkabrii**, S. Tschiatschek, C. Plant
International Conference on Learning Representations (ICLR), 2025

ReSL: Enhancing Deep Clustering Through Reset-based Self-Labeling
A. Shkabrii, T. Klein, L. Miklautz, S. Tschiatschek, C. Plant
Workshop on Spurious Correlation and Shortcut Learning (SCSL) and Workshop on Scaling Self-Improving Foundation Models (SSI-FM) @ International Conference on Learning Representations (ICLR), 2025

Non-redundant Image Clustering of Early Medieval Glass Beads
L. Miklautz, **A. Shkabrii**, C. Leiber, B. Tobias, B. Seidl, E. Weissensteiner, A. Rausch, C. Böhm, C. Plant

IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2023

Honors and Awards

UNIVIE Research Award for Students 2025/2026, awarded €1,400 | Austria
University of Vienna

Ernst Mach Scholarship (EM UKR/Batch I), awarded €12,000 | Austria
Republic of Austria, Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF), and the OeAD GmbH - Agency for Education and Internationalisation

Service

Conference Reviewer: ICLR, NeurIPS, ICML

Program Committee Member: ICDM 2025 LLM4Cluster Workshop

Skills

AI/ML: Representation learning, adversarial robustness, fairness, HSIC, optimization, clustering, reinforcement learning

Programming: Python, PyTorch, JAX, AWS, HPC, Java, C++

Languages: English (C1), German (C1), Ukrainian (Native)