Ömer Veysel Çağatan

omerveyselacademic@gmail.com • https://asparius.github.io

Research Interests

AI Safety, Alignment, RL for LLMs, Deep RL.

Education

B.Sc., Computer Engineering Aug 2020 – Jun 2024
Koç University, Turkey
Tracks: Artificial Intelligence, Data Analytics
Award: Vehbi Koç Scholar, Anatolian Scholarship Program

Publications

Peer-reviewed

- **O. V. Cagatan**, O. F. Tal, and M. E. Gursoy. *Adversarial Robustness of Discriminative Self-Supervised Learning in Vision.* In Proceedings of IEEE/CVF International Conference on Computer Vision, **ICCV 2025**.
- **O. V. Cagatan** and B. Akgun. *Uncovering RL Integration in SSL Loss: Objective-Specific Implications for Data-Efficient RL.* In Proceedings of Reinforcement Learning Conference, **RLC 2025**.
- MMTEB Team, **O. V. Cagatan**. *MMTEB: Massive Multilingual Text Embedding Benchmark.* In Proceedings of International Conference on Learning Representations, **ICLR 2025** .
- **O. V. Cagatan**. *UNSEE: Unsupervised Non-contrastive Sentence Embeddings.* In Proceedings of the European Chapter of the Association for Computational Linguistics, **EACL 2024**.
- **O. V. Cagatan** and B. Akgun. *BarlowRL: Barlow Twins for Data-Efficient Reinforcement Learning.* In Proceedings of Asian Conference on Machine Learning (ACML), 2023.
- **O. V. Cagatan**. *ToddlerBERTa: Exploiting BabyBERTa for Grammar Learning and Language Understanding.* In Proceedings of CoNLL–CMCL 2023 Shared Task: The BabyLM Challenge (CONLL), 2023.

Workshop Papers

- **O. V. Cagatan** and B. Akgun. *Uncovering RL Integration in SSL Loss: Objective-Specific Implications for Data-Efficient RL.* In NeurIPS 2024 Workshop: Self-Supervised Learning – Theory and Practice, 2024.
- **O. V. Cagatan**. *SigCLR: Sigmoid Contrastive Learning of Visual Representations.* In NeurIPS 2024 Workshop: Self-Supervised Learning – Theory and Practice, 2024.

Preprints

- **O. V. Cagatan**, G. Sahin, B. Akgun and X. Zhao . *Clipping-Free Policy Optimization for Large Language Models.* arXiv preprint, 2026.
- **O. V. Cagatan** and B. Akgun. *Failure Modes of Maximum Entropy RLHF.* arXiv preprint, 2025.

Professional Experience

Research Engineer Apr 2025 – Present
KUIS AI, Koç University

- Authored research proposal for **multi-turn orchestration agents with RL** for banking chatbots at İş Bankası (Turkey's largest bank); leading 100M TL (∼2M EUR).
- Developing **AI-driven discovery methods** for novel conductive metal-organic frameworks (MOFs) in collaboration with Asst. Prof. Büşra Dereli (Chemistry, Koç University)
- Conducting research on **Test-Time scaling**, **Deep RL** and **RLHF/Reasoning** through collaborative projects:
  - Test-time scaling without RL based on base model characteristics (with Asst. Prof. Gözde Gül Şahin)
  - Scaling resolution for deep RL environments such as Procgen (with Raphael Trumpp, TUM; preprint forthcoming)
  - Developing pessimistic and stable GRPO replacement (with Dr. Xuandong Zhao, UC Berkeley and Asst. Prof. Barış Akgün; preprint forthcoming)

NLP Intern Jul 2022 – Oct 2022
FineSci Technology (Supervised by Assoc. Prof. Alptekin Küpçü)
Built Turkish sentiment analysis models; created large datasets and conducted cross-lingual benchmarking

Research Experience

Research Fellow, SPAR Sep 2025 – March 2026
UC Berkeley (Supervised by Dr. Xuandong Zhao)

- Investigating reward hacking behavior in LLMs by adapting AI Safety Gridworlds for large language model settings. Analyzing existing RL objectives proposed for reasoning and developing novel algorithms to address their limitations. Project ongoing.

Undergraduate Research Assistant Nov 2022 – Mar 2025
Koç University

- Data-efficient reinforcement learning (Supervised by Asst. Prof. Barış Akgün) — **RLC 2025**, **ACML 2023**, **NeurIPS 2024 Workshop**
- Robustness of self-supervised models (Supervised by Asst. Prof. M. Emre Gürsoy) — **ICCV 2025**
- Novel vision self-supervised learning objectives (Single-author) — **NeurIPS 2024 Workshop**
- Non-contrastive sentence embeddings (Single-author) — **EACL 2024**

Teaching, Service & Skills

Teaching: Guest Lecturer (RL, NLP), Teaching Assistant (Probability), Tutor (Differential Equations), Koç University, 2022–2025
Reviewing: ICLR 2025–2026, ICML 2026, ACML 2023
Technical Skills: Python, PyTorch, Flax/JAX, Transformers, TRL, verl, C/C++, Java, LaTeX