

Omer Veysel Cagatan

omerveyselacademic@gmail.com
<https://asparius.github.io>
<https://github.com/asparius>

Istanbul, Turkey

Education	Koc University , Sariyer, Istanbul Bachelor of Science Major: Computer Engineering Tracks: Artificial Intelligence, Data Analytics GPA: 3.71/4.00	Oct 2020-June 2024
Experience	MMTEB Frequent Contributor in The Massive Multilingual Text Embedding Benchmark (MMTEB) is a community-led extension of MTEB to cover embedding tasks for a massive number of languages. <ul style="list-style-type: none">Added multiple datasets for languages like Turkish, Kurdish, and Armenian and reviewed pull requests and issues.	Apr 2024-Present
	Adversarial Robustness of the Self-Supervised Models supervised by M. Emre Gürsoy Investigating the robustness of the Self-supervised visual representation learning models <ul style="list-style-type: none">We evaluate the adversarial attack performance of 8 top self-supervised models on more than 30 attacks that cover object recognition, object detection, and semantic segmentation. The preprint will be released soon.	Jan 2024-Present
	Novel Vision Self-Supervised Learning Objective Adapting sigmoid contrastive loss to do self-supervised pretraining. <ul style="list-style-type: none">Adjusted sigmoid loss to be used as a pretraining objective. Trained and achieved competitive results with SOTA models on CIFAR 10, CIFAR 100, and Tiny Imagenet. Accepted to NeurIPS 2024 Workshop: Self-Supervised Learning - Theory and Practice.	Nov 2023-Present
	Data-Efficient Deep RL supervised by Asst. Prof. Baris Akgun Working on Data-Efficient Reinforcement Learning. My primary aim is to utilize representation learning methods to make more efficient agents. <ul style="list-style-type: none">Led the BarlowRL project which has been accepted into the ACML 2023. After the success of the BarlowRL, we further explored the impact of different non-contrastive objectives in the SPR which is the baseline of state-of-the-art value-based agents. The preprint will be released soon.	June 2023-Present
	KoçAssistant Mar 2024-June 2024 <ul style="list-style-type: none">Developed a RAG-based chatbot that answers almost every question about Koç University. The database is created by scraping all university-affiliated websites and retrieving relevant documents with a hybrid search system with probabilistic and fine-tuned semantic models. The overall system is presented with the Cohere Toolkit for users to interact with.	
	Sample-Efficient Language Model in BabyLM Challenge Worked on building data-efficient language models in a small developmentally plausible corpus. <ul style="list-style-type: none">Scaled a previous small named BabyBERTa to achieve RoBERTa level grammar understanding.	Apr 2023-Aug 2023

Non-Contrastive Sentence Embeddings with Prof. Deniz Yuret and Prof. Alper Erdogan Nov 2022-Mar 2023

Worked on developing state-of-the-art sentence embedding models which are created after fine-tuning with non-contrastive objectives.

- Led this project which aimed to enhance BERT embeddings by fine-tuning it with non-contrastive objectives such as CorInfoMax and VICReg. This project has been accepted to EACL 2024.

NLP Intern at FineSci Technology with Assoc. prof. Alptekin Kupcu July 2022-Oct 2022
Worked on developing state-of-the-art Sentiment Analysis models in Turkish.

- Collected and curated a large dataset to obtain the best performant model. Lastly, wrote a comprehensive literature review about Sentiment Analysis in Turkish and the results of the findings from different languages employed in Turkish.

SPARK Autonomous Car Feb 2022-June 2022
Worked as a Computer Vision Engineer in the Software Team.

- SPARK is an autonomous vehicle Team that builds a car from scratch for the national competition. My primary work was the training of the Lane Detection System.

Teaching Experience *Teaching Assistant* for the ENGR200, Probability and Random Variables Oct 2022-Jan 2023.
Tutor for the MATH204, Differential Equations Feb 2022- June 2022

Skills
Programming: Python, C, C++, Java, Docker
Deep Learning: Pytorch, Flax/Jax, Transformers, Langchain, LlamaIndex.

Publications

Peer Reviewed Conferences

O. V. Cagatan and B. Akgun. *Uncovering RL Integration in SSL Loss: Objective-Specific Implications for Data-Efficient RL* Submitted to the Artificial Intelligence and Statistics, 2025 (AISTATS '25).

O. V. Cagatan, O. Tal and M.E. Gursoy. *Adversarial Robustness of Self-Supervised Learning in Vision* Submitted to the Thirteenth International Conference on Learning Representations (ICLR '25).

MMTEB Team, **O. V. Cagatan**. *MMTEB: Massive Multilingual Text Embedding Benchmark* Submitted to the Thirteenth International Conference on Learning Representations, 2025 (ICLR '25).

O. V. Cagatan. UNSEE: Unsupervised Non-contrastive Sentence Embeddings . To appear in the 18th Conference of the European Chapter of the Association for Computational Linguistics, (EACL '24).

O. V. Cagatan and B. Akgun. *BarlowRL: Barlow Twins for Data-Efficient Reinforcement Learning*. To appear in the Asian Conference on Machine Learning, 2023 (ACML '23).

Workshops and Shared Tasks

O. V. Cagatan and B. Akgun. *Uncovering RL Integration in SSL Loss: Objective-Specific Impli-*

cations for Data-Efficient RL To appear in the NeurIPS 2024 Workshop: Self-Supervised Learning - Theory and Practice, 2024 (NeurIPS SSL Workshop '24).

O. V. Cagatan *SigCLR: Sigmoid Contrastive Learning of Visual Representations* To appear in the NeurIPS 2024 Workshop: Self-Supervised Learning - Theory and Practice, 2024 (NeurIPS SSL Workshop '24).

O. V. Cagatan. *ToddlerBERTa: Exploiting BabyBERTa for Grammar Learning and Language Understanding.* To appear in the CoNLL–CMCL 2023 Shared Task: The BabyLM Challenge (CONLL'23).