Cavit **Cakir**

Munich, Germany

✉ cavitcakir3@gmail.com | cavitcakir | cavitcakir

Summary

Machine Learning Engineer with 4+ years of experience in AI, specializing in Computer Vision (CV) and Natural Language Processing (NLP). Holds a Master's degree in Informatics from the Technical University of Munich, with a focus on CV, NLP, and ML. Skilled in building scalable AI applications using state-of-the-art zero-shot models and efficient data pipelines. Improved vectorization speed by 5x in image processing tasks at the current position, demonstrating expertise in optimizing AI workflows. Proven track record in deploying AI solutions on cloud platforms (AWS, GCP) and developing robust backend systems with FastAPI. Passionate about leveraging technology to drive innovation, with strong skills in Python, C++, PyTorch, and Numpy.

Education

Technical University of Munich (TUM)
Munich, Germany

Master of Science in Informatics
Oct 2021 - Mar 2024

- **Thesis:** Self-Supervised Feature Learning for 3D LiDAR Semantic Segmentation with Neural Radiance Fields (NeRFs)
- **Specializations:** *Machine Learning, Computer Vision, Natural Language Processing, Artificial Intelligence*
- **Grade:** 2.0/5.0 *(German grading system, 1.0 highest - 5.0 lowest)*

Sabanci University
Istanbul, Turkey

Bachelor of Science in Computer Science and Engineering
Sep 2016 - Jun 2021

- **Graduation Project:** Meeting Scheduler Chatbot
- **CS GPA:** 3.78/4.0 *(US grading system, 4.0 highest - 0.0 lowest)*

Skills

Programming	Python, C++, JavaScript
Machine Learning	PyTorch, NumPy, Pandas, TensorFlow, Scikit-Learn, OpenCV, BentoML, Qdrant
Web Development	React, Node.js, FastAPI, SQL, MongoDB
DevOps & Cloud	Git, Docker, Bash, AWS, GCP
Natural Languages	Turkish (*native*), English (*advanced*), German (*basic - learning*)

Experience

Machine Learning Engineer
Munich, Germany

Quasara | Python, Pytorch, LLMs, AWS, GCP, FastAPI, Qdrant
May 2024 - Current

- Designed and implemented a scalable vectorization pipeline that leverages state-of-the-art zero-shot multimodal models to process large-scale image datasets from S3 and store them in a Qdrant database.
- Increased **vectorization speed by 5x**, significantly improving processing efficiency and reducing computational costs, resulting in significant savings for the company.
- Enhanced small object recognition using multimodal LLMs for image captioning and detection, along with computer vision techniques. Achieved a **2x speed increase with only 3% accuracy loss**, and a **3x speed increase with 5% accuracy loss**.
- Implemented a benchmarking script to evaluate vectorization performance in terms of accuracy and precision.
- Developed and maintained a FastAPI backend for the company's AI product, and collaborated with client engineering teams.
- Deployed scalable solutions on cloud platforms, including Google Compute Engine and AWS EC2.
- Adapted to rapidly changing customer requirements, ensuring timely project delivery.

Master Thesis Student
Ingolstadt, Germany

CARIAD - a subsidiary of Volkswagen | Python, Pytorch, NeRF | 📄 View Thesis
May 2023 - Dec 2023

- Developed a novel self-supervised learning approach for 3D LiDAR semantic segmentation, utilizing Neural Radiance Fields (NeRFs) to extract volumetric features from image data.
- Distilled 2D image features into 3D domains, allowing LiDAR segmentation models to be trained with geometrically aware features without relying on labeled data.
- Demonstrated the potential of using self-supervised techniques to overcome the limitations of densely labeled 3D datasets.

Computer Vision Engineer Working Student
Munich, Germany

Quasara | Python, Pytorch
Dec 2022 - May 2023

- Led the end-to-end development of a *Damage Classification Project* using state-of-the-art transformer models, overcoming imperfect labeled data through advanced cleaning and augmentation techniques, achieving over **93% accuracy** in image classification.
- Demonstrated project ownership and expertise by applying cutting-edge AI technologies to solve real-world challenges.
- Enhanced client communication through regular updates and detailed presentations of project outcomes, promoting strong engagement with customers.

Natural Language Processing (NLP) Intern   *Istanbul, Turkey*

FineSci Technology | Python, Pytorch   *Jul 2020 - Oct 2020*

- Contributed to the *News Classification and Clustering Project*, focusing on classifying Turkish news.
- Applied state-of-the-art transformer language models to classify news articles, achieving over **96% accuracy** and significantly improving the efficiency and accuracy of news categorization.

Undergraduate Teaching Assistant   *Istanbul, Turkey*

Sabanci University | C++, Teaching   *Feb 2019 - Feb 2020*

- Independently managed lab sessions for 20-30 students, providing hands-on support and guidance, and offered personalized mentoring during office hours to enhance students' understanding and engagement with course materials.
- Utilized C++ as the primary programming language to facilitate practical learning in the *Introduction to Computing* course, demonstrating strong technical expertise and application skills.

Selected Projects in Computer Vision

Implementation of Panoptic Neural Field   *Technical University of Munich*

Advanced Practical Course | Python, Pytorch, NeRF | 📄 View Project   *Oct 2022 - Mar 2023*

- Implemented the *Panoptic Neural Field (PNF)* paper from scratch using *Kaolin Wisp* and the *KITTI 360 dataset*, demonstrating initiative and expertise in translating theoretical concepts into practical, executable code without existing implementations.
- Optimized the *PNF* architecture for better performance, employing advanced techniques to significantly improve model efficiency and accuracy, highlighting my ability to innovate and enhance existing methodologies.

Improving Point Cloud Transformer using Curve Aggregation   *Technical University of Munich*

ML for 3D Geometry Course Project | Python, Pytorch, Transformers | ⊙ View Project   *Apr 2022 - Sep 2022*

- Enhanced the shape analysis capabilities of a *Point Cloud Transformer* by implementing a *Curve Aggregation* method, improving model performance for detailed 3D point cloud processing by 0.35%. This innovation led to more accurate shape interpretations on the *ShapeNet Parts* dataset.
- Converted the *Point Cloud Transformer* implementation from *Jittor* to *PyTorch*, demonstrating technical proficiency and versatility in adapting complex models to widely used frameworks.

Selected Projects in NLP

NLP and Knowledge Graphs for Research Cluster Prediction and Analysis   *Technical University of Munich*

TUM-DI-LAB Interdisciplinary Project | Python, Pytorch, Transformers | 📄 View Project   *Oct 2022 - Mar 2023*

- Contributed to the *Unsupervised Classification of Research Papers Project* by proposing a novel *Hierarchical Classification Method* and successfully applying existing techniques.
- Used the cutting-edge *SPECTER* embedding model to enhance the quality of embeddings, enabling a more precise understanding of research paper content and significantly contributing to the project's success.

Emotional Clustering of Social Media Users   *Technical University of Munich*

Advanced Practical Course | Python, Pytorch, BERT | 📄 View Project   *Apr 2022 - Sep 2022*

- Utilized a *BERT model* to extract embeddings and cluster users based on their textual data, demonstrating proficiency in advanced NLP techniques for user segmentation.
- Preprocessed Reddit user posts to meet the requirements of the *BERT model*, showcasing the ability to prepare and adapt large-scale datasets for complex NLP tasks.
- Applied various *Dimensionality Reduction Methods*, including *HDBSCAN* and *KMeans*, to efficiently analyze and cluster high-dimensional data.

Meeting Scheduler Chatbot   *Sabanci University*

Bachelor's Graduation Project | Python, JavaScript, React, Node.js, Docker | ⊙ View Project   *Sep 2020 - Jun 2021*

- Developed an advanced chatbot using the *RASA Bot Framework*, integrating pre-trained natural language understanding methods to accurately process and interpret user inputs. This demonstrated expertise in utilizing state-of-the-art AI technologies to create intelligent conversational agents.
- Implemented a graphical user interface for the chatbot with *React* and *Node.js*, encapsulated in *Docker* containers to ensure ease of deployment and scalability. This end-to-end development highlighted my full-stack development skills and commitment to delivering robust, scalable applications.