Arda Yüksel

Darmstadt, Germany | rd.yuksel07@gmail.com | linkedin.com/in/arda-yüksel | github.com/ArdaYueksel

Education

Technical University of Munich (TUM), M.Sc. in Informatics — Sept 2021 – June 2024
- CGPA: 1.7
- **Master Thesis:** Turkish Open Book Question Answering Benchmark Generation

Bilkent University, B.Sc. in Electrical and Electronics Engineering and Physics(Second Major) — Sept 2016 – June 2021
- CGPA: 3.64
- **Academic Achievement:** Comprehensive Scholarship recipient, ranked 691 among 2 million in University Entrance Exams

Experience

PhD Candidate, TU Darmstadt – Darmstadt, Germany — Feb 2025 – present
- Researched on Multimodal Fact Checking.
- Organized Special Reading Groups on Human AI collaboration and Multimodal AI.
- Collaborated with TrustHLT and UKP Lab members on various research projects.
- Led weekly meetings and research projects with Industry partners.

AI Developer, Neuland.AI – Cologne, Germany — Aug 2024 – Jan 2025
- Developed and deployed multilingual RAG-based chatbots using LlamaIndex and OpenAI SDK.
- Enhanced application containerization on Azure AI and Docker for scalable production.
- Boosted generation performance by up to 20% through prompt engineering and advanced RAG methodologies.
- Designed Graph and Vector database schemas for optimized data retrieval with Neo4j and Qdrant.
- Implemented microservices and unit tests via FastAPI within CI/CD pipeline.

AI Researcher - Working Student, Fraunhofer IKS – Munich, Germany — May 2022 – June 2024
- Conducted research on object detection systems (DETR, YOLO, Faster-RCNN).
- Developed a pipeline for genetic algorithm-based adversarial attacks to assess model vulnerabilities.
- Reduced detection accuracy in object detectors through effective adversarial attack implementations.
- Published findings at AI.BAY'23 and Design, Automation and Test in Europe Conference'23 conferences.

NLP Researcher - Undergraduate Intern, Aykut Koç Lab - UMRAM – Ankara, Turkey — Feb 2019 – Aug 2021
- Engineered a pipeline to detect semantic changes using Gaussian word embeddings.
- Leveraged Google Books' Ngram and SemEval datasets for word embedding alignment in multiple languages.
- Achieved state-of-the-art performance in multilingual semantic change detection for SemEval.
- Published a first-author paper in IEEE's TASLP.

Publications

TurkishMMLU: Measuring Massive Multitask Language Understanding in Turkish
Arda Yüksel, Abdullatif Köksal, Lütfi Kerem Şenel, Anna Korhonen, Hinrich Schütze
aclanthology.org/2024.findings-emnlp.413 (Findings of EMNLP 2024)

Butterfly Effect Attack: Tiny and Seemingly Unrelated Perturbations for Object Detection
Nguyen Anh Vu Doan, Arda Yüksel, Chih-Hong Cheng
doi.org/10.23919/DATE56975.2023.10137164 (Design, Automation & Test in Europe Conference & Exhibition)

Semantic Change Detection With Gaussian Word Embeddings
Arda Yüksel, Berke Uğurlu, Aykut Koç
doi.org/10.1109/TASLP.2021.3120645 (IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP))

Technologies

Languages: English, Turkish, German (B1)

Programming Languages: Proficient in Python, C++, Matlab, TeX, Shell
Intermediate: Java, JavaScript

Tools and Frameworks: Azure, Qdrant, Docker, Neo4j, LlamaIndex, PyTorch, HuggingFace, Git, NLTK, OpenCV, Keras, TensorBoard, Deap, Scikit-Learn, Pandas, Numpy, MongoDB

Theory: Natural Language Processing (NLP), Large Language Models (LLM), Retrieval Augmented Generation (RAG), Graph Databases, Vector Databases, Artificial Intelligence, Machine & Deep Learning, Adversarial Attacks, Out-of-Distribution Detection, Genetic Algorithms, Computer Visio

Projects

Speech and Text Language Identification
- Designed and implemented an advanced language detection pipeline for both text and speech in the FLEURS dataset.
- Enhanced performance by 50% through integration of Text-to-Speech and Automatic Speech Recognition technologies.
- Achieved state-of-the-art (SOTA) performance, effectively supporting various low-resource languages.

Sentimental Polysemy Detection
- Conducted in-depth analysis of polysemous words within sentiment detection using Large Language Model (LLM) embeddings.
- Extracted and utilized contextual embeddings from models such as BERT, BART, and RoBERTa via HuggingFace.
- Applied dimensionality reduction techniques (PCA, UMAP, T-SNE) and visualized complex embeddings on TensorBoard, uncovering nuanced semantic variations.

Abstractive Dialogue Summarization
- Developed an abstractive summarization pipeline, transforming raw dialogue sequences into concise summaries using LLM models.
- Collaborated with a team to clean and optimize real-world dialogue datasets, ensuring data quality.
- Fine-tuned BART and GPT models, achieving robust performance in summarizing complex conversational data.

BookEx Exchange Books
- Collaborated in a cross-functional team to architect and develop BookEx, a dynamic book exchange platform.
- Engineered the backend for user and book data management, enhancing platform functionality and user experience.
- Built a messaging feature for seamless user interaction, contributing to both frontend and backend development.

Trajectory Prediction
- Predicted vehicle trajectories using time series analysis, leveraging deep learning architectures (GRU, LSTM).
- Constructed GRU and LSTM models from scratch, demonstrating strong foundations in deep learning.
- Benchmarked against Linear & Ridge Regression, achieving superior predictive accuracy with neural network models.

Extractive News Summarization
- Analyzed state-of-the-art (SOTA) performance of Large Language Models (LLMs) on extractive news summarization.
- Cleaned and preprocessed extensive datasets (CNN, BBC) using NLTK, ensuring high-quality data for model training.
- Fine-tuned sentence embeddings with HuggingFace models, extracting accurate and relevant summaries.