Recep Firat Cekinel

frtcek95@gmail.com —Webpage: https://user.ceng.metu.edu.tr/~rfcekinel/— Linkedin—GitHub

Passionate Machine Learning Researcher with hands-on experience in deploying NLP solutions at scale. Eager to leverage expertise in large language models, explainable AI, and multimodal reasoning to build impactful AI systems for real-world applications. Had an experience on downstream NLP tasks, including sarcasm detection, named entity recognition, sentiment analysis, and fact-checking. Seeking opportunities to apply hands-on experience in AI and NLP to tackle real-world challenges.

EDUCATION

Eberhard Karls University of Tübingen, Germany 2023 — 2024
Visiting predoctoral researcher at the Computational Linguistics Department

Middle East Technical University (METU), Turkey 2020 — 2025
Ph.D. in Computer Science, cGPA: 3.93 / 4.00
Thesis: Multilingual, Multimodal and Explainable Approaches in Automated Fact-Checking [PDF]

M.S. in Computer Science, GPA: 4.00 / 4.00 2017 — 2020
Thesis: Information Extraction from News related Texts using Graph Mining Techniques [PDF]

B.S. in Computer Science, GPA: 3.41 / 4.00 (11^{th} out of 110) 2013 — 2017

EXPERIENCE

Extreme Analytics for Mining Data Spaces (EXA4MIND) EU Horizon 2020
Machine Learning Researcher April 2023 - Ongoing

- Worked on the EXA4MIND project funded by EU Horizon Europe Research and Innovation program
- Developed an inference platform for advanced analytics on supercomputers and automated data management
- Utilized large language models to convert queries into SQL/NoSQL, enabling efficient data retrieval and workflow automation in high-performance computing (HPC) environments
- Keywords: Prompt Engineering, Text2SQL, Retrieval Augmented Generation
- Technologies: Fast API, Apache Airflow, Transformers, Apptainer, ZeroMQ

Middle East Technical University Ankara, Turkey
Machine Learning Researcher & Teaching Assistant June 2018 - Ongoing

- Conducted research on explainable AI for NLP tasks, including sarcasm detection, named entity recognition, sentiment analysis, and fact checking, to interpret the reasoning processes of language models
- Utilized frameworks such as LIME, SHAP, and attention-based visualizations to analyze and improve model interpretability
- Implemented machine learning models for misinformation detection, leveraging advanced NLP techniques to identify misinformation
- Designed evaluation frameworks to assess model accuracy, robustness, and scalability in detecting misinformation across diverse domains
- Instructor of CENG310: Algorithms and Data Structures with Python, a course that covers data structures and algorithms

Eberhard Karls University of Tübingen Tübingen, Germany
Machine Learning Researcher June 2023 - Jul 2024

- Designed and evaluated novel ML architectures for automated fact-checking task.
- Collected a fact-checking data set for Turkish and performed experiments with transfer learning through fine-tuning large language models and utilizing machine translation.
- Assessed the feasibility of transfer learning approaches for low-resource languages, the results provide some preliminary evidence for the type of information, knowledge or style, used in automated fact-checking models.
- Examined the utilization of multimodal information in Vision-Language Models (VLM), text and image, compared to embeddings extracted from discrete text-only and image-only models for the fact-checking problem
- Trained a probing classifier pipeline that involves extracting the last hidden layer's representation and using it as input for a small feed-forward neural network to explore whether recent VLMs can leverage multimodal information for the fact-checking problem.

ASELSAN Ankara, Turkey
Software Engineer Oct 2021 - May 2022

- Contributed to the development of a command and control software project, focusing on efficient processing of incoming sensor data to ensure real-time system performance
- Implemented a concurrent replay module interacting with a MySQL database using Hibernate to enable seamless data retrieval and playback for system analysis
- Designed and integrated additional GUI features to enhance user interaction and system usability to improve operator efficiency

ORACLE Ankara, Turkey
Database Management Intern June 2016 - July 2016

- Gained hands-on experience with Oracle 11g Database management and configuration
- Demonstrated data masking and data encryption techniques using Oracle Advanced Security, showcasing methods to protect sensitive data in enterprise environments
- Applied learned concepts to practical scenarios, ensuring data security and compliance with organizational standards
- Technologies: Oracle 11g Database, Oracle Advanced Security

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, C++, HTML
- **Databases:**PostgreSQL, Neo4j, MongoDB, Cassandra
- **NLP Frameworks:** Hugging Face Transformers, SpaCy, NLTK, Ollama
- **Machine Learning Frameworks:** PyTorch, Pytorch Lightning, Scikit-learn
- **Web Frameworks:** FastAPI, Streamlit, Flask, Django
- **Big Data Frameworks:** ZeroMQ, Airflow, Spark, Storm

AWARDS

- 2023 DAAD Research Grant for Doctoral Candidates
- 2023 TUBITAK 2214-A Research Grant for Doctoral Candidates
- 2017 Best Poster Award, METU Engineering Day

TALKS & WEBINARS

- EuroCC@Turkey 2023 - Natural Language to SQL: LLMs and Recent Progresses [Video] 2023
- EuroCC@Turkey 2021 - Introduction to NoSQL Databases [Video] 2021
- EuroCC@Turkey Graph School 2021 - Text Graphs and Graph Databases [Video] 2021

SELECTED PUBLICATIONS

- **Journal Articles**
  1. A.G.Ozer, **R.F. Cekinel**, I.H.Toroslu, P.Karagoz, DocSpider: Cross-Domain Natural Language Querying Dataset for MongoDB. *Natural Language Processing*, 1-32 (2025). [Paper] [Github]
  2. **R.F. Cekinel**, P.Karagoz, Event prediction from news text using subgraph embedding and graph sequence mining. *World Wide Web* 25, 2403–2428 (2022). [Paper]
- **Conference Papers**
  1. **R.F. Cekinel**, P.Karagoz, C.Coltekin, Multimodal Fact-Checking with Vision Language Models: A Probing Classifier based Solution with Embedding Strategies, *COLING 2025*, 2025. [Paper] [Github]
  2. **R.F. Cekinel**, P.Karagoz, C.Coltekin, Cross-Lingual Learning vs. Low-Resource Fine-Tuning: A Case Study with Fact-Checking in Turkish, *LREC-COLING 2024*, 2024. [Paper] [Github]
  3. **R.F. Cekinel**, P.Karagoz, Text-Based Causal Inference on Irony and Sarcasm Detection, in *Big Data Analytics and Knowledge Discovery. DaWaK 2022*, 2022. [Paper] [Github]
- **Workshop Papers**
  1. A.U.Ozturk, **R.F. Cekinel**, P.Karagoz, Make Satire Boring Again: Reducing Stylistic Bias of Satirical Corpus by Utilizing Generative LLMs, *18th Workshop on Building and Using Comparable Corpora (BUCC), COLING 2025*. 2025. [Paper] [GitHub]
  2. **R.F. Cekinel**, P.Karagoz, Explaining Veracity Predictions with Evidence Summarization: A Multi-Task Model Approach, *IEEE Big Data 2024, 10th Special Session on Intelligent Data Mining*, 2024. [Paper] [Github]

PROJECTS

- Advanced Querying Inference System (AQIS) 2024 - Ongoing
  Developed an advanced querying interface using REST API in Python to support natural language queries on SQL/NoSQL databases. Additionally, adapted retrieval augmented generation using large language models to enhance query response accuracy. Currently, working on integrating the inference module into an HPC platform, enabling researchers to access the system seamlessly.
  Technologies: Transformers, FastAPI, Ollama, MongoDB, Apache Airflow, ZeroMQ
- Event Prediction from News Text using Graph Mining Approaches 2018 - 2020
  I was a researcher in this project which was funded by the Scientific and Technological Research Council of Turkey (TUBITAK). My main tasks were representing texts such as news in the form of a graph on Neo4j and contributing to the development of a graph sequence mining library. [Paper] [Github] [Web page]
  Keywords: Graph Mining, Graph Embeddings, Word Embeddings, Event Extraction