Curriculum Vitae

Ali Safaya

asafaya19@ku.edu.tr
Koç University

Github: alisafaya
Twitter: ali_safaya
Semantic-Scholar: Ali Safaya
ORCID: 0000-0003-0945-8897
Personal Website: asafaya.me

Research Interests

Natural Language Understanding, Large Language Models, Memory Augmentation

Employment

July 2025— Present — **Position:** AI Research Scientist; **Where:** HyperbeeAI, California

January 2024— July 2025 — **Position:** AI Research Scientist; **Where:** Codeway, Istanbul

April 2021— January 2024 — **Position:** NLP Research Scientist; **Where:** HyperbeeAI, California

November 2017— September 2020 — **Position:** ML/NLP Engineer; **Where:** Casemice Digital, Istanbul

Education

May 2020— 2026 (expected) — **Degree:** Doctor of Philosophy in Computer Science; **Where:** Koç University, Istanbul; **GPA:** 3.72 of 4.00

PhD research fellow at KUIS AI Center.

August 2015— June 2019 — **Degree:** Bachelor of Science in Computer Engineering; **Where:** Sakarya University, Sakarya; **GPA:** 3.89 of 4.00

Valedictorian for the graduation ceremony, for graduating with the highest GPA throughout the degree program.

Research

September 2022— Present — **Project:** Longform Language Models and Neural Memory Augmentation; **Where:** KUIS AI Center, Koç University, Istanbul; **Advisor:** Prof. Dr. Deniz Yuret

Contributions: I am working on separating semantic and episodic memory in language models, as well as developing improved memory architectures for modeling long sequences. This is my main research focus for my PhD study.

January 2021— Present — **Project:** Turkish Data Depository; **Where:** KUIS AI Center, Koç University, Istanbul; **Advisor:** Prof. Dr. Deniz Yuret

Contributions: I am leading multiple teams of researchers to develop datasets and tools for Turkish NLP. One example of our work was to deliver a new spell-checker to *LibreOffice* team for Turkish language. Also, we built a website to share all available Turkish NLP datasets in an open and documented manner. Another recent project is *Mukayese*: an

extensive NLP benchmarking platform for Turkish, which provides test sets and baselines for most of NLP tasks. See `tdd.ai`.

August 2019— August 2020

Project: ERC funded project of Emerging Markets Welfare
Where: Koç University, Istanbul
Advisor: Dr. Ali Hürriyetoglu
Contributions: I worked as Project Engineer on building a Protest Events Database using Natural Language Processing and Machine Learning, where I gained experience on being a part of a long term research project, and collaborating with interdisciplinary academics. I had the chance to co-organize workshops at top NLP venues like ACL and LREC during my work time on this project.

Teaching

2020— Present

Position: Teaching Assistant
Where: Computer Science Dept., Koç University
Teaching assistant for Data Structures and Algorithms *COMP-202*, Algorithms and Complexity *COMP-305*, Deep Learning *COMP-541*, and Natural Language Processing *COMP-542*

Publications

- Ali Safaya and Deniz Yuret. "Neurocache: Efficient Vector Retrieval for Long-range Language Modeling". In: *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Mexico City, Mexico: Association for Computational Linguistics, June 2024, pp. 870–883. URL: https://aclanthology.org/2024.naacl-long.50
- Ali Safaya and Engin Erzin. "HuBERT-TR: Reviving Turkish Automatic Speech Recognition with Self-supervised Speech Representation Learning". In: *Computing Research Repository* arXiv:2210.07323v2 (2022). DOI: 10.48550/arXiv.2210.07323. URL: http://arxiv.org/abs/2210.07323v2
- Aarohi Srivastava et al. "Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models". In: *Computing Research Repository* arXiv:2206.04615 (2022). DOI: 10.48550/ARXIV.2206.04615. URL: https://arxiv.org/abs/2206.04615
- Ali Safaya et al. "Mukayese: Turkish NLP Strikes Back". In: *Findings of the Association for Computational Linguistics: ACL 2022*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 846–863. DOI: 10.18653/v1/2022.findings-acl.69. URL: https://aclanthology.org/2022.findings-acl.69
- Ali Hürriyetoğlu et al. "Event Coreference Resolution for Contentious Politics Events". In: *Computing Research Repository* arXiv:2203.10123 (2022). DOI: 10.48550/ARXIV.2203.10123. URL: http://arxiv.org/abs/2203.10123
- Ali Safaya, Moutasem Abdullatif, and Deniz Yuret. "KUISAIL at SemEval-2020 Task 12: BERT-CNN for Offensive Speech Identification in Social Media". In: *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Barcelona (online): International Committee for Computational Linguistics, 2020, pp. 2054–2059. URL: https://aclanthology.org/2020.semeval-1.271
- Ali Hürriyetoğlu et al. "Automated Extraction of Socio-political Events from News (AESPEN): Workshop and Shared Task Report". English. In: *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*. Marseille, France: European Language Resources Association (ELRA), 2020, pp. 1–6. ISBN: 979-10-95546-50-4. URL: https://aclanthology.org/2020.aespen-1.1
- Ali Hürriyetoğlu et al. "COVCOR20 at WNUT-2020 Task 2: An Attempt to Combine Deep Learning and Expert rules". In: *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*. Online: Association for Computational Linguistics, 2020, pp. 495–498. DOI: 10.18653/v1/2020.wnut-1.75. URL: https://aclanthology.org/2020.wnut-1.75
- Ali Safaya. "Event Sentence Detection Task using Attention model". In: *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, 07*. Lugano, Switzerland: CEUR Workshop Proceedings, 2019. URL: http://ceur-ws.org/Vol-2380/paper_122.pdf

Presentations

- Ali Safaya and Taner Sezer. *Turkish Data Depository Project: Towards a Unified Turkish NLP Research Platform*. Seminar presented at KUIS AI meeting, Istanbul (online). 2022
- Ali Safaya. *Reproducing Single Headed Attention - Recurrent Neural Networks*. Seminar presented at KUIS AI meeting, Istanbul (online). 2020

Languages

- Arabic — Native-Bilingual Proficiency
- English — Working Proficiency
- Turkish — Native-Bilingual Proficiency

Technical experience

PyTorch, Jax, CUDA

Honors and awards

- KUIS AI Center Fellowship (2020 — present)
- Turkey's National NLP Hackathon — 1st place award (2021)
- Inzva Computer Engineering PhD Scholarship (2020 — 2021)
- Graduated with Honor Degree of BS program (Valedictorian of 2019)
- Türkiye Scholarship (2015 — 2019)