

Muhammed Yusuf Kocyigit

PHD, COMPUTER SCIENCE, BOSTON UNIVERSITY

111 Cummington Mall #B19, 02215, Boston, MA

✉ kocyigit@bu.edu | [in yusufkocyigit](#) | +16173377822

Education

Boston University

Boston, MA/USA

PHD IN COMPUTER SCIENCE / ADVISOR: PROF. DERRY WIJAYA, GPA: 4.0

2020 - Present

- Research in cross lingual representation learning, machine translation evaluation and bias detection
- Coursework include: Towards Universal NLP, Graduate Networks, Programming Languages, Random Computing

Bogazici University

Istanbul, TR

MS & BS IN ELECTRICAL AND ELECTRONICS ENGINEERING , MS GPA: 3.75 & BS GPA: 3.0

2012 - 2017 - 2020

- Research in Machine translation evaluation, Senior project: Handwritten text generation using conditional GANs
- Related Coursework: Graduate Deep Learning(AA), Advanced Speech Processing(AA), Machine Learning(BA), Pattern Recognition(AA), Bioinformatics(AA), Graduate Artificial Neural Networks(AA), Management for Engineers(AA)

Professional Experience

Boston University Artificial Intelligence Research Group

Boston, MA, USA

PHD, RESEARCH FELLOW

2020 - Present

- My main research focus is natural language evaluation. I am interested in many aspects of this problem from compositionality to robustness. Below are areas that I have worked on which I think link to this theme.
 - Evaluating and improving compositional generalization in neural semantic parsing models.
 - Using diverse augmentations and contrastive learning to improve sentence embeddings in downstream transfer performance.
 - Evaluating behavioral abilities of models in cross-task problems.
 - Learning better cross lingual representations through multitask and contrastive learning for quality estimation and parallel corpus mining.
- Working on bias detection in natural language data(books and news) through time within **2021 Google Research Scholar Program**.
- Leading research project for Ottoman handwritten text recognition which received **\$50K** from the Scientific Research Council of Turkey.

FAIR, Meta

London, UK

RESEARCH SCIENTIST INTERN

October 2023 - Present

- Working on defining and detecting data contamination between evaluation sets and pretraining data of Large Language Models
- Working on representing performance over-estimations caused by contamination and how to better understand model metrics withing current evaluation paradigm.

Amazon

Cambridge, UK

APPLIED SCIENCE INTERN

June 2022 - September 2022

- Worked in the Semantic Understanding team. My project was focused on better evaluating and improving compositional generalization abilities of Semantic Parsing models
- Created an extension to TMCD splits that has more robust convergence to compositionally challenging splits.
- Worked on improving compositional generalization through meta learning and chain-of-thought prompting.
- Worked quantifying characteristics of train-test split that make a split compositionally challenging.

TRIA AI

Istanbul, TR

Co-FOUNDER & CEO

2018 - 2020

- Founded company, led business development, closed contracts worth over **\$100K**, and managed client relations. Worked with the C-Suite of LC Waikiki, multi-billion retailer present in 40 countries, to guide their AI strategy to increase online sales.
- Deployed an ML model for LC Waikiki for behavioral customer segmentation, customer lifetime value prediction and churn analysis that achieved **+100%** increase in conversion rates for targeted ads.
- Worked with DEDAS, one of Turkey's biggest electric distributors, to detect illegal uses of electricity and increased the detection rate by **+150%** with estimated business impact of **\$500K**.

Turkcell Global Information

Istanbul, TR

TECHNICAL PROJECT LEAD, ML ENGINEER

2017 - 2018

- Joined AI team as a founding member. Engineered a text classification pipeline for routing customer issues with business impact of **\$250K**.
- Led conversational modeling project which secured **\$200K** award grant from Scientific Research Council of Turkey.

Publications

- [1] M. Y. Kocyigit, J. Lee, and D. Wijaya, "Better quality estimation for low resource corpus mining," in *Findings of ACL*, 2022.
- [2] Z. Tang, M. Y. Kocyigit, and D. Wijaya, "Augcse: Contrastive sentence embedding with diverse augmentations," in *AACL-IJCNLP*, 2022.
- [3] M. Y. Kocyigit, H. Kané, A. Abdalla, P. Ajanoh, and M. Coulibali, "Nubia: Neural based interchangeability assessor for text generation," in *INLG, EvalNLG Eval*, 2020.
- [4] M. Y. Kocyigit and E. E. Kucuk, "Western, religious or spiritual: An evaluation of moral justification in large language models," in *NeurIPS 2023, MP2 Workshop*, 2023.

- [5] M. Y. Kocyigit, H. Kané, P. Ajanoh, A. Abdalla, and M. Coulibali, "Towards neural similarity evaluator," in *NeurIPS 2019, Document Intelligence*, 2019.
- [6] A. Feyza Akyürek, S. Paik, M. Yusuf Kocyigit, S. Akbiyik, Ş. Leman Runyun, and D. Wijaya, "On Measuring Social Biases in Prompt-Based Multi-Task Learning," *Findings of NAACL*, 2022.
- [7] A. Feyza Akyürek, M. Yusuf Kocyigit, S. Paik, and D. Wijaya, "Challenges in Measuring Bias via Open-Ended Language Generation," *NAACL, GeBNLP Workshop*, 2022.
- [8] M. Y. Kocyigit, A. Andy, and D. Wijaya, *A novel method for analysing racial bias: Collection of person level references*, 2023. arXiv: **2310.15847 [cs.CY]**.

Leadership

COVID-19 Datathon - Turkey

Istanbul, TR

DIRECTOR - TRIA AI

2019

- Organized the first COVID-19 Datathon in Turkey to, digitally, bring people together and fund projects that can serve the public good amidst the crisis. With +60 teams and around 150 participants the top 4 projects were selected and awarded \$15K.

Critical Thinking Society, Bogazici University

Istanbul, TR

FOUNDER & DIRECTOR

2014 - 2015

- Found the first Critical Thinking Society in Bogazici University where we organized workshops for over 500 participants to teach and engage on how to use critical thinking for argumentation, media consumption and preventing misinformation.

Awards

2012 **Presidential High Honour Scholarship**, Turkish University Entrance Exam - Sciences Top 100 / 2M students

Istanbul, TR

Skills

Programming Language

Python, C, MATLAB, ATS, Assembly, Pytorch, Pytorch-Lightning, Hydra, WandB
English(Advanced), Arabic(Intermediate), German(Intermediate), Turkish(Native)