Şafak Bilici

Date Of Birth: 08/1999
Address: Istanbul, Turkey
Website: safakkbilici.github.io
Github: github.com/safakkbilici
LinkedIn: linkedin.com/in/msafakbilici

Email: m.safak.bilici@gmail.com
Mobile: +90-553-385-3245

Education

Yildiz Technical University   Istanbul, Turkey
Bachelor of Science - Computer Engineering; GPA: 3.48/4.00   *2017 - 2022*
Graduation Thesis: *Deep Multimodal Learning with Vision-and-Language Transformers*

Experience

Software Engineer at Insider
Full-time   *Sep 2023 - Present*
- The search team provides SaaS search engine, to personalize search results for e-commerce partners.
- Developed a new concurrent caching mechanism between Search Backend and BFF API. Reduced the overall latency by 30 ms.
- Designed the infrastructure of AutoSuggestion API in AWS ECS and implemented its indexing pipeline as daily triggered AWS Lambda function.

Research Engineer at Huawei
Full-time   *Jan 2022 - Sep 2023*
- Developed AppGallery's re-ranker and dense retrieval model, trained with enriched low-resource user click history with weak-supervision. Implemented the back-end integration part for hybrid retrieval with Elasticsearch and distributed vector caching. This neural search pipeline improves the CTR by $\sim 15\%$
- Implemented low-latency multilingual spelling correction module. The model outperforms Huawaei's all previous spell checkers with ~ 20ms latency and more than 90% "Did You Mean" click through rate.
- Conducting A/B Testing experiments to evaluate and optimize NLP products for enhanced user experience.
- Leveraged distributed data processing and efficient model deployment techniques for enabling scalability and accelerating model inference using Distributed RPC, low-level quantization, and ONNX.
- Presented technical documentations to overseas.

Machine Learning Researcher at YTU NOVA Lab.
Full-time   *Dec 2020 - Feb 2022*
- Research lab under the supervision of Prof. Dr. Mehmet Fatih Amasyali.
- Research on intersection of Variational Inference and Low-Resource Language Models.
- Published two papers about a novel method on data augmentation for NLP.

Open Source Projects

- **x-tagger**: A Natural Language Processing toolkit for part of speech tagging and named entity recognition with various computational linguistics and deep learning methods. Has more than 4k installation on pip.
- **bayesmedaug**: A Python library that optimizes your data augmentation hyperparameters for medical image segmentation tasks by using Gaussian Process and Bayesian Optimization.

Skills Summary

- **Languages**: Python, Java, C/C++, Go.
- **Tools**: Docker, Kubernetes, Elasticsearch, Redis, PostgreSQL, Airflow, Spark, Spring Framework, Terraform.
- **Platforms**: Linux, AWS (certified).

Volunteer Experience

AI Program Specialist at inzva
Volunteer   *Aug 2021 - April 2023*
- Part of the team that organizes the syllabus of the AI programs.
- Technical mentoring for the project groups.
- GPU server administration.

Guide at inzva
Volunteer
- Guided numerous programs at inzva, focusing on topics such as anomaly detection, large language models, machine translation, search engines, word embeddings, and more.

Publications

- Can Özbey, Talha Çolakoğlu, **M. Şafak Bilici**, Ekin Can Erkuş. "A Unified Formulation for the Frequency Distribution of Word Frequencies using the Inverse Zipf's Law", in Special Interest Group on Information Retrieval (SIGIR), 2023.

- **M. Şafak Bilici**, Mehmet Fatih Amasyali. "Transformers as Neural Augmenters: Class Conditional Sentence Generation via Variational Bayes", 2022. arXiv: 2205.09391.

- E. Sadi Uysal, **M. Şafak Bilici**, B. Selin Zaza, M. Yiğit Özgenç, Onur Boyar, "Exploring The Limits Of Data Augmentation For Retinal Vessel Segmentation", 2021. arXiv: 2105.09365.

- **M. Şafak Bilici**, Mehmet Fatih Amasyali. "Variational Sentence Augmentation For Masked Language Modeling", in Innovations in Intelligent Systems and Applications Conference, 2021.