

# Farrin Marouf Sofian

farrinsofian@gmail.com | [Github](#) | [Website](#) | +1(657)642-2996

## EDUCATION

---

### University of California, Irvine

*M.S. Computer Science*

Irvine, CA

9/2023 – 6/2025

- Machine Learning, Deep Generative AI, Image Understanding, Probabilistic Models

### Koç University

*B.Sc in Computer Science*

Istanbul, Turkey

9/2019 – 6/2023

- Minor: Mathematics
- Honors: Vehbi Koç Scholar for outstanding academic performance for 4+ semesters.
- Autonomous Driving, Natural Language Processing, Computer Vision, Real Analysis

## WORK & RESEARCH EXPERIENCE

---

### Machine Learning Engineer Intern

*Adobe*

06/2024 – Present

*San Jose, California*

- Worked on Improving the AI assistant in the Adobe Acrobat.

### Graduate Researcher at Mandt Lab

*University of California, Irvine*

1/2024 – Present

*Irvine, California*

- Implementing generative models for super-resolution in climate simulation data.

### Research Assistant at Computer Vision Lab

*Koç University & İş Bank Artificial Intelligence Center*

9/2021 – 12/2023

*Istanbul, Turkey*

- Conducted research in Turkey's most competitive AI lab. Worked on sound-guided image manipulation and generation as well as audio identification and re-synthesis projects.
- Leveraged Stable Diffusion while designing and implementing various adapter architectures for efficient audio conditioning.
- Tested multiple image inversion techniques to identify the method that best preserves image identity while facilitating editing.
- Conducted experiments and fine-tuned the framework with audio encoding models including AudioCLIP and CLAP, to assess their performance as audio encoders in the proposed framework.
- Conducted under the supervision of [Dr. Duygu Ceylan](#) and [Prof. Aykut Erdem](#). Work currently under submission.

### Research Assistant at NLP Lab

*KUIS AI lab*

10/2022 – 7/2023

*Istanbul, Turkey*

- Created the first high-quality dataset of 130k parallel sentences in Turkish with 25 expert curated writing rules for grammatical error schema.
- Implemented and fine-tuned NMT baseline and sequence tagger model for grammatical error correction and detection.
- Worked with High Performance Computing (HPC) clusters and contributed on writing a [detailed documentation](#) on how to use HPC clusters for graduate students and companies.
- Published in IJCNLP-AAACL 2023.

### Machine Learning Engineer Intern

*RadiusAI*

6/2022 – 9/2022

*Tempe, Arizona*

- Researched and proposed deep learning-based solutions for camera calibration, tailored to the company-specific scene geometry.
- Evaluated the effectiveness of deep learning as well as other traditional methods to solve the problem.
- Optimized the computation of a semantic segmentation scores matrix by nearly 50%, utilizing dynamic programming in the pre-processing step.
- Created the company's first fully automated camera calibration, reducing the company's reliance on expensive annotators.

### Software Engineer Intern, Backend

*Digitopia*

6/2021 — 9/2021

*Istanbul, Turkey*

- Designed the company's flagship product that provides marketing analytics for large enterprises.
- Developed a dynamic PDF generator from user dashboard results using Handlebars, JavaScript and Node.js.

## PROJECTS

---

### Wallpaper Recommender

10/2022 – 12/2022

- Developed a Deep Learning framework to find and recommend similar wallpapers based on image similarity.
- Experimented with unsupervised learning approaches to represent images, (e.g SimCLR and SwAV).
- Generated similar new wallpapers using diffusion models and created a demo website for the project.

### Birds of Istanbul App

5/2021 — 7/2022

- Led a group of 4 students and collaborated with a psychology department to create Turkey's first audio-based bird identification mobile application.
- Curated a dataset of 355k bird songs local to Turkey and pre-processed the audio using various models such as Sound Separation proposed by Google.
- Examined several self-supervised transformer-based audio models and fine-tuned them on the dataset.

### Audio Re-synthesis and Conversion

1/2022 – 4/2022

- Performed extensive literature study to identify the best suitable model for converting voice of a species into another one.
- Fine-tuned 3 pre-trained encoders, utilizing VQ-VAE and HiFi-GAN as the content encoder and vocoder respectively, on a custom dataset.
- Trained and extracted species' sound embeddings from transformer-based audio encoder model.

## PUBLICATIONS

---

A. Kara, **F. M. Sofian**, A. Y. X. Bond, and G. G. Sahin, "GECTurk: Grammatical Error Correction and Detection Dataset for Turkish" The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AAACL) 2023.

## VOLUNTARY EXPERIENCE

---

### ACM Corporate Relations Team

*Koç University*

10/2020 – 1/2023

*Istanbul, Turkey*

- Assisted in organizing and coordinating hackathons and workshops with prominent companies.
- Organized 2 different company-led mock interviews to enhance students' job-seeking prospects.

### Undergraduate Teaching Assistant

*Koç University*

2021 & 2023

*Istanbul, Turkey*

- **Comp 301: Programming Languages Concepts** and **Comp 132: Advanced Programming**: Created lab assignments, projects, quizzes and graded them. Held office hours and organized lab sessions.

## SKILLS & PROFICIENCIES

---

**Tools:** Python, Pytorch, Tensorflow, OpenCV, Scikit Learn, C, Git/GitHub, Java, JavaScript, Node.js, React

**Languages:** Persian (Native) – English (Fluent) – Turkish (Fluent) – German (Intermediate)