

Serhan YILMAZ

✉ serhan.yilmaz@sabanciuniv.edu | [in yilmaz-serhan](https://www.linkedin.com/in/yilmaz-serhan) | [GH serhanylmz](https://github.com/serhanylmz) | [Globe serhanyilmaz.org](https://serhanyilmaz.org)

RESEARCH INTERESTS: **Multi-Agent Systems, Agentic AI Frameworks, Conversational AI**

EDUCATION

Computer Science — Bachelor of Science

SEP 2021 - JUN 2025

Sabanci University

cGPA: **3.63/4**

- **Relevant Coursework:** Machine Learning, Image Processing, Linear Algebra, Statistics, Algorithms, Data Structures, Advanced Programming, Operating Systems, Database Systems
- Ranked **top 0.02%** among 2.9 million students in the National University Entrance Exam.

PUBLICATIONS

[1] Serhan Yilmaz and Kemal Oflazer, "**CASCADES: Compound AI Systems for Controlled And Diverse Question Generation**," in preparation, 2024.

[2] Angelika Romanou, [...] Serhan Yilmaz, [...] Sara Hooker, Antoine Bosselut, et al., "**INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge**," arXiv:2411.19799 [cs.CL], under review at ICLR 2025, 2024.

RESEARCH EXPERIENCE

Carnegie Mellon University — Research Internship Pittsburgh, PA (Remote) | DEC 2023 - PRESENT

- Research internship with Professor Kemal Oflazer at Carnegie Mellon University on advanced question generation and multi-agent frameworks.
- Developed a novel compound AI system utilizing multiple specialized language models in structured workflows for enhanced question generation.
- Implemented sophisticated LLM-as-a-Judge methodologies with Chain-of-Thought reasoning for semantic and structural assessment.
- Designing iterative feedback loops in multi-agent collaboration frameworks, achieving **60% improvement** over baseline generation methods.
- Authored novel paper detailing the CASCADES framework [1], introducing compound AI systems for controlled question generation through structured LLM orchestration and iterative refinement (in preparation, 2024).

Expedition Aya - Cohere for AI — Research Contributor Remote | AUG 2024 - OCT 2024

- Contributor to INCLUDE [2], a multilingual benchmarking dataset for LLMs (ICLR 2025 submission).
- Built Python pipelines using LLM and VLMs for automated extraction of multichoice questions from educational resources across diverse languages.
- Involved in dataset curation efforts and mentoring international contributors.
- Enhanced cross-lingual evaluation methodologies for question-answering tasks, improving benchmark reliability.

TU Darmstadt — Research Intern Hessen, Germany (Remote) | JUL 2024 - SEP 2024

- Developed a self-sustaining agentic chatbot system for psychological assessment of Ukrainian refugees under Prof. Iryna Gurevych.
- Implemented complex decision-making pipelines for autonomous question planning, contextual understanding, and verdict generation.
- Optimized the system for deployment at Charité Berlin, integrating cultural sensitivity parameters and robust assessment protocols.

KTH Royal Institute of Technology — Research Intern Stockholm, Sweden | JUN 2024 - SEP 2024

- Advanced voice activity prediction models under Prof. Gabriel Skantze, achieving significant improvements through attention mechanism optimization.
- Implemented novel data augmentation techniques for audio processing, enhancing model robustness in real-world scenarios.
- Successfully deployed enhanced models on Furhat Robotics' humanoid robots, improving conversational turn-taking accuracy by 94%.
- Developed multimodal training frameworks integrating speech, text, and contextual signals for enhanced interaction capabilities.

EPFL — *Summer@EPFL Intern*

Lausanne, Switzerland | JUN 2023 - SEP 2023

- Initiated **Project Charisius**, for developing robust, privacy-preserving Federated ML environments.
- Optimized gradient aggregators, leveraging **PyTorch** and **CUDA** to achieve performance improvements of **2200%** compared to previous versions.
- Published the Charisius library for accelerated Statistics and Federated ML applications, reducing runtime by up to **96%**.
- Developed documentation website and Flask-backed benchmarking leaderboard for the project.

Boston University — *Research Assistant*

Boston, MA (Remote) | JUN 2022 - SEP 2022

- Conducted research on Threat Modeling and Component Design for large-scale server systems under Prof. Rabia Tugce Yazicigil.
- Developed a C++ tool providing **336** models to design secure servers for resource-intensive applications.

LEADERSHIP EXPERIENCE

kAi Sabanci — *Founder & President*

Istanbul, Turkey | MAR 2023 - PRESENT

- Founded and led Sabanci University's premier AI club, growing membership to **600+** with **80% active participation**.
- Organized **17 events** in one term, with workshops on Gen AI, NLP, ML, and DL, serving as a **voluntary TA**.
- Secured kAi Sabanci's position as a **founding member** of the **NVIDIA Student Network**.
- Facilitated high-profile visits, including hosting **NVIDIA AI chief Simon See** and organizing a visit to the Turkish Presidency's Digital Transformation Office.
- Supervised **6 major projects**, including ChatSU, and NeRF campus modeling.
- Selected for the prestigious **NVIDIA Student Spotlight** series for outstanding leadership and achievement.

WORK EXPERIENCE

Yapi Kredi Bank — *NLP R&D Engineering Intern*

Istanbul, Turkey | OCT 2023 - APR 2024

- Enhanced document interpretation algorithms using OCR-free **Donut Transformers**.
- Fine-tuned language models, reducing annotation effort by approximately **40%** through transfer learning.

Sabanci University — *Undergraduate Teaching Assistant*

Istanbul, Turkey | SEP 2022 - JAN 2023

- Wrote homework assignments for students and held weekly recitation and office hours.

Pharus Tech — *Intern*

Stockholm, Sweden (Remote) | JUL 2020 - SEP 2020

- Analyzed and visualized data from the ESCO dataset, with information on **3008** occupations and **13890** skills.
- Developed data processing pipelines, improving data analysis efficiency by an estimated **30%**.

AWARDS

Sabanci University — *Sakip Sabanci Award for Outstanding Success*

SEP 2022 -

- Ranked in the **top 4%** among students in my cohort, demonstrating academic excellence.

The Royal Swedish Academy of Engineering Sciences — *Innovation in Crisis Award*

MAY 2020 -

- Secured **1st place** as part of team United in Crisis for innovative solutions during challenging times.

The New York Academy of Sciences — *Junior Academy Membership*

SEP 2019 -

- Selected for the prestigious "The Junior Academy" program with a highly competitive **8%** acceptance rate.

SKILLS

- **Languages & Tools:** Python, Docker, Linux, AWS, Riva, Git, Javascript, HTML/CSS, C, C++
- **Libraries & Frameworks:** PyTorch, TensorRT, RAPIDS, JAX, cuDNN, OpenCV, CUDA, Keras, Pandas, NumPy, Transformers, LangChain, Ray
- **Specialized Skills:** Multi-Agent Systems, LLM Fine-tuning, Prompt Engineering, Audio Processing, Attention Mechanisms