Ugur Cekmez, PhD

Phone: +90 (554) 253 5399 Email: ucekmez@gmail.com

Summary

With a diverse background in AI and backend development, I have co-founded tech startups, secured multiple investments, and served as an AI Advisor for several companies, while also making academic contributions as an Assistant Professor.

Work Experience

CTO, Konstellate, Inc., USA
Aug 2024–Present

Leading the development and deployment of cutting-edge AI-driven solutions that revolutionize the analysis of Key Opinion Leader (KOL) content in the pharmaceutical sector.
ugur@konstellate.com

Co-Founder, Technic Intelligence, Inc., USA
March 2023–Present

technic.ai, an early-stage technology company, is developing an AI inference engine that can operate efficiently on various accelerated hardware. The product holds potential in edge computing and cloud applications, where it can optimize decision-making processes, enhance reliability, and address privacy concerns. Tools I've been utilizing: NVIDIA DeepStream, Triton, Apache TVM, OpenVino, transfer learning, labeling and CV & LLM inferencing...
ugur@technic.ai

Board, SCADR Space, LLC., USA
May 2023–Present

SCADR Space is an early-stage space-tech startup focusing on space-based computation and data routing solutions. By leveraging state-of-the-art technologies, SCADR points out to the satellite industry by providing efficient data processing and transmission services, reducing latency, bandwidth requirements, and power consumption. I've been advising them on their AI infrastructure needs (mostly AI engines, object detection and tracking tasks)
ugur@scadr.space

AI Advisor, HiveLogic AI, USA
March 2023–Present

HiveLogic specializes in unifying diverse sensors to provide actionable business insights through visualization and Agentic AI. As AI Advisor, I guide the integration of AI technologies into their platform.

AI Advisor, Hiri AI, TR
Jan 2023–Present

Hiri is an AI-backed HR platform that transforms the recruitment process with its smart and adaptive testing capabilities. It offers online tests including personal inventories, language & skill assessments, uniquely tailored to each candidate based on their responses. With its custom interview workflow generation, Hiri brings efficiency and accuracy to HR departments. In this position I've been advising (and writing) on LLM-powered hiring process management (smart candidate CV evaluation and interview postreports).
ugur@hiri.ai

Assistant Professor, Biruni University, Computer Engineering Department, TR
March 2023–Present

I conduct "AI Entrepreneurship" and "AI Applications" courses where I mostly focus on recent LLM architectures and micro-saas projects on top of language and vision models.
ucekmez@biruni.edu.tr

***Technical Lead, AI Solutions*, Hai Cora Ltd.**, USA
Feb 2024–Aug 2024

Led AIOps @hai Cora (early-stage), creating truly human-like AI that can complete various tasks, engage in meaningful conversations, provide emotional support, advice, and companionship. In this project, I focused on developing and integrating multiple concepts such as scalable RAG management, collaborative agents, function-calling LLMs, creating personas and managing context windows, as well as speech-to-text, text-to-speech, voice cloning and testing different types of frameworks (langchain, llamaindex, memgpt, crewai, autogen).

Part-time Lecturer, *Marmara University, Computer Engineering Department*, TR
Oct 2022–July 2023

After receiving my PhD, I conducted two courses: "Artificial Neural Networks" and "Introduction to Machine Learning"

Lead MLOps Engineer, Chooch Intelligence Technologies, Co., USA
June 2021–March 2023

I was responsible for designing, implementing and deploying the whole MLOps pipeline of a Visual AI product, leading 6+ people by being fully active in almost every task in this flow (including customer meetings and commercial & military-related deployments)

Techs I utilized: Python & FastAPI & Django, OCI (Docker), Tensorflow & PyTorch & ONNX, Nvidia Triton & DeepStream, Redis & Kafka & MQTT, Node.js & Next.js, WebSockets

Concepts I worked: Continuous & production-ready scalable ML model deployment, GPU-powered video decoding with GStreamer, robust API design with FastAPI, Airgapped-ready product deployment with Docker, Fullstack Edge-Inference pipeline

Research Engineer, TRT - Turkish Radio and Television Corporation, Data&Insight Lab, TR
April 2020–June 2021

I was responsible for developing and deploying AI-powered tasks, ETLs, DevOps pipelines, infrastructures at TRT AI Lab. I also led the data engineers team there and implemented and deployed the personalisation engine of TRT Dinle [1] as well as the data infrastructure of TRT's digital products.

Experienced R&D Engineer, SIEMENS, Technology & Innovation Unit, TR
November 2018–February 2020

Here I was responsible for researching, proposing and implementing innovative tech ideas. I used all my tech stack that had mentioned in previous positions. I also worked as a solution architect jointly with other members of the team to develop in-house projects involving AI, hybrid/mobile, cloud and container technologies. Some key projects I was involved :

- *Virtual Device Management:* I designed and implemented a digital twin concept for smart IoT devices capable of installing custom apps and monitoring possible versioning and hardware loading scenarios. It was build on a Docker-powered light environment where multiple devices can be created, customized and monitored

- *DigiHome:* I was responsible for re-implementing the architecture of a low-cost digital home concept with Amazon AWS services where Amplify, GraphQL, AppSync, DynamoDB, Cognito and related services were highly used. The front-end was made with React.

Senior Researcher, TUBITAK BILGEM Cloud Computing & Big Data Research Lab, TR
April 2017–November 2018

My position involved using cloud technologies, making inferences with data intelligence and empowering apps with artificial intelligence approaches. I worked on parallel deep learning approaches. Some of my projects were determining anomaly behaviors of ECG data, predicting possible future CPU loads of virtual machines in cloud environment and developing a machine learning as a service (MLaaS) product (including re-implementation of many ML and DL applications from both their scientific papers and github repos those are ranging from text summarization/tagging/diff, object recognition/detection, plate recognition to anomaly detection and stock prediction etc. and corresponding UIs with python and react). While researching the AI models, I also developed their engine to run as scalable containers by using microservice architectures. I also prepared backends and APIs (Falcon/fastAPI of Python). I was also one of the members of cloud team that's responsible of developing our DevOps processes. For this aim, we developed a Jenkins-centered DevOps pipeline. The requirements included github and unit-tests integration, code coverage tests and deploying the developer codes out to microservices via mentioned tools.

[1] www.trtdinle.com

Co-Founder, General Manager, Fili Labs R&D and Software Solutions Ltd
November 2014–December 2019

I founded a start-up company to carry out a government-funded (TUBITAK TEYDEB 1512) research project which was about developing an interactive personality inventory to make better decisions of hiring.

Research&Teaching Assistant, Yıldız Technical University, Computer Engineering Department, TR
February 2013–April 2017

I was responsible of managing and assisting some of the computer hardware and science lessons for the 1st, 2nd and 3rd year students.

Lectures I helped with: Logical Circuits, Circuit Theory, Electronic Circuits, Microprocessors, Formal Languages, Network Technologies, Data Structures and Algorithms, Operating Systems, Programming Languages, Introduction to Computer Graphics

Freelance, Researcher & Full Stack Developer
2007–Present

Along with my regular jobs, I also work as a freelancer researcher/app developer if I see an awesome opportunity to strengthen my abilities. Currently, I'm able to design and implement an end-to-end full-stack system that's composed of 1) machine learning models (working as a computation engine), 2) Async Python APIs (communicating with models, databases and frontends), 3) backends / frontends (with node.js using next.js/react) and 4) microservice architecture (to containarize applications in a highly available environment with Docker and Kubernetes/Cloud Foundry).

Part-Time Developer, @ Various Companies:
June 2010–September 2011

Software Developer (Mostly Python/Django Development), Egnity: I used wordpress and django to create web applications answering customer needs and prepared backends for some Facebook apps using Facebook API back then.

Webmaster, Istanbul Bilgi University - Computer Science Department: I designed, implemented and maintained a web application that's able to stream videos of some of the computer science lectures (website once called video.cs.bilgi.edu.tr).

Webmaster, Akinon Design Studio: Here I figured out the power of Python and Django. Akinon was one of the early (Turkey-based) commercial users of Django. While contributing to the customer projects, I learned many things which I still benefit from.

Education

PhD. Computer Engineering, Institute of Pure and Applied Sciences, Marmara University, 2022,
MSc. Computer Engineering, Aeronautics and Space Technologies Institute, Turkish Air Force Academy, 2014,
BSc. Computer Science, Faculty of Arts and Sciences, Istanbul Bilgi University, 2012,
BSc. (erasmus) Computer Science, Science Faculty, Aarhus University, Denmark, 2011,

Study/Work fields

Machine Learning, Deep Learning, Data Intelligence, Massively Parallel Systems (GPGPU), Evolutionary Computation, Microservices, Software Engineering, Software Project Management, Innovation.

Computer Skills

Programming and Scripting : Python, JavaScript, LaTeX, Bash, CUDA C, Solidity
Frameworks : Tensorflow 2.0 for Deep Learning. React/React-Native/Next.js for full-stack/hybrid development. FastAPI and Express for backend development, Docker and related container technologies for deploying apps, Redis for cache and queue management, Kafka for message transmission, and pretty much the other stuff related to Machine Learning Systems Engineering (MLOps)

Honors & Awards & Investments

$100K pre-seed funding for Technic Intelligence, Inc. —— Certificate of Merit award for a research in 2014 [2] —— TUBITAK 1512 Start-up Company Investment Support, ≈ $45K, November 2014 —— BSc Honor Degree, Istanbul Bilgi University, 2012 —— European Union Grant for Exchange Studies, Denmark, Spring 2011 —— Istanbul Bilgi University Scholarship for BSc, 2009-2010 —— Student Selection and Placement Center (OSYM) Scholarship for BSc, 2008–2012.

Research

Completed Papers

Data-Oriented Network Attack Classification System with Deep Learning for the Internet of Things (PhD thesis)
Internet of Things (IoTs) Security: Intrusion Detection using Deep Learning [3]
DevOps Architecture in the Cloud, [4]
Network Anomaly Detection with Deep Learning, [5]
Multi-UAV Path Planning with Multi Colony Ant Optimization, [6]
Cells Classification with Deep Learning, [7]
Multi-UAV Path Planning with Parallel Genetic Algorithms on CUDA Architecture, [8]
Multi Colony Ant Optimization for UAV Path Planning with Obstacle Avoidance, [9]
Controlling Moving Targets by using Unmanned Aerial Vehicles, [10]
Parallel Solution of Large Scale Traveling Salesman Problems by using Clustering and Evolutionary Algorithms, [11]
UAV Path Planning with Parallel Genetic Algorithms on CUDA architecture, [12]
A UAV Path Planning with Parallel ACO Algorithm on CUDA Platform, [13]
Parallel Ant Colony Optimisation for UAV Route Planning on CUDA Architecture, [14]
Adapting the GA Approach to Solve Traveling Salesman Problems on CUDA Architecture, [15]
Adaptive Cut Level Selection at Agglomerative Clustering Algorithms in Co-Word Analysis, [16]

Book Chapters

Artificial Intelligence Studies in Health Services and Medical Images (in Turkish). [17]

Journal & Paper Reviews

IEEE Access Journal, '23
Journal of Web Engineering, '21
ICUAS '14 - '16 - '17, '20 International Conference on Unmanned Aircraft Systems, USA
SIU '16 - '17 - '18, IEEE Conference on Signal Processing and Communications Applications, Turkey
Neural Computing & Applications Journal, March '16.

For more: LinkedIn, Github

ps: Most of the events before 2014 has been truncated to maintain the 4 page length.

[2] UAV Path Planning with Parallel Genetic Algorithms on —— CUDA architecture, the World Congress on Engineering 2014, in London, U.K
[3] Published in Nov 2021, DOI: https://doi.org/10.13052/jwe1540-9589.2062
[4] 27th IEEE Conference on Signal Processing and Communications Applications, Sivas, Turkey, April 2019
[5] 26th IEEE Conference on Signal Processing and Communications Applications, İzmir, Turkey, May 2018
[6] 17th International Conference on Intelligent Systems Design and Applications, Delhi, India, Dec 2017
[7] 25th IEEE Conference on Signal Processing and Communications Applications, Antalya, Turkey, May 2017
[8] Presented in ACM GECCO'16, Defense, Security, and Risk Management Workshop, Denver, Colorado, The U.S.A, July 2016
[9] Presented in ICUAS '16, The 2016 International Conference on Unmanned Aircraft Systems, Arlington, Virginia, The U.S.A, June, 2016
[10] Presented in the 24th IEEE Conference on Signal Processing and Communications Applications, Zonguldak, Turkey. May, 2016
[11] Presented in the 24th IEEE Conference on Signal Processing and Communications Applications, Zonguldak, Turkey. May, 2016
[12] Presented in the World Congress on Engineering 2014, in London, U.K. July, 2014
[13] Presented in ICUAS '14, The 2014 International Conference on Unmanned Aircraft Systems, Orlando, Florida, The U.S.A, May, 2014
[14] Presented in the 22th IEEE Conference on Signal Processing and Communications Applications, Trabzon, Turkey. Apr, 2014
[15] Presented in 14th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary. Nov, 2013
[16] Bachelor thesis. June 05, 2012, Istanbul Bilgi University, Department of Computer Science
[17] Big Data and Open Data Analysis: Methods and Applications. Editor: Prof. Dr. Seref Sagiroglu