Amir Ziashahabi

University of Southern California, Los Angeles, CA
ziashaha@usc.edu github.com/amir-zsh

Education

University of Southern California **Aug 2021 – Dec 2026 (Expected)**
Ph.D. in Electrical Engineering (GPA: 4/4) *Los Angeles, CA*

Sharif University of Technology **Sep 2017 – Jan 2020**
M.Sc. in Computer Engineering –Software (GPA: 19.38/20) *Tehran, Iran*

Iran University of Science and Technology **Sep 2013 – Sep 2017**
B.Sc. in Computer Engineering – Software (GPA: 17.66/20) *Tehran, Iran*

Research Interests

- Efficient Machine Learning
- Privacy-Preserving Machine Learning
- Large Language Models
- Distributed Systems

Research and Professional Experience

Machine Learning Researcher **Aug 2021 – Present**
vITAL Lab, University of Southern California (Advisor: Prof. Salman Avestimehr) *Los Angeles, CA*

- Conducting machine learning research with a focus on efficiency in system and algorithmic design.
- Areas of focus include large language models, federated learning, secure multiparty computation, distributed ML, and diffusion models.

Machine Learning Intern **May 2022 – Aug 2022**
FedML, Inc. *Los Angeles, CA*

- Designed and developed hierarchical federated learning systems and analyzed communication technologies performance.

Senior Developer & Distributed Systems Researcher **Feb 2018 – Aug 2021**
Rastak Media Sepehr Co. *Tehran, Iran*

- Led several projects, including developing a distributed ledger application with Hyperledger Fabric and a ride-hailing platform on Ethereum.
- Engineered solutions using Kubernetes, Solidity, React, NestJS and Web3 APIs.

Distributed Systems Researcher **Jan 2018 – Jan 2020**
Decentralab Lab, Sharif University of Technology (Advisor: Prof. Mohammad Ali Maddah-Ali) *Tehran, Iran*

- Investigated scalability solutions for blockchain with a focus on layer-2 solutions.
- Conducted a thesis on a blockchain-based social news aggregator.
- Contributed to curriculum design for a blockchain technology course.

Distributed Systems Researcher **Dec 2016 – Sep 2017**
Cloud Computing Center, Iran University of Science & Technology (Advisor: Prof. Mehrdad Ashtiani) *Tehran, Iran*

- Researched IaaS cloud computing solutions; deployed and monitored OpenStack services.

Technical Skills

Programming: Python, JavaScript, Solidity, CSS, HTML, Go, C#, PHP, Java, C/C++
Tools and Technologies: PyTorch RPC, gRPC, MPI, Docker, Kubernetes, Redis, Nginx, Unity Game Engine, Truffle
Libraries/Frameworks: PyTorch, Transformers, Diffusers, vLLM, GPT-Fast, Express, NestJS, React, Django, Flask, Laravel

Academic Honors

- Annenberg Fellowship, University of Southern California, 2021
- Ranked 2nd among M.Sc. students in Software Engineering, Sharif University, 2020
- Full-Tuition Scholarship, Ranked 14th among 25K+ in Nationwide M.Sc. University Entrance Exam, 2017
- Ranked 4th among B.Sc. students in Software Engineering, IUST, 2017
- Full-Tuition Scholarship, Top 1% of 251K+ in Nationwide B.Sc. University Entrance Exam, 2013

Selected Publications and Preprints

Amir Ziashahabi[*], Yavuz Faruk Bakman[*], Duygu Nur Yaldiz, Mostafa El-Khamy, Sai Praneeth Karimireddy, Salman Avestimehr. Reject Only Critical Tokens: Pivot-Aware Speculative Decoding. *NeurIPS 2025 Efficient Reasoning Workshop*, 2025 (accepted).

Lei Gao[*], **Amir Ziashahabi**[*], Yue Niu, Salman Avestimehr, Murali Annavaram. MobiZO: Enabling Efficient LLM Fine-Tuning at the Edge via Inference Engines. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025 (accepted).

Narges Ghasemi[*], **Amir Ziashahabi**[*], Salman Avestimehr, Cyrus Shahabi. GeoToken: Hierarchical Geolocalization of Images via Next Token Prediction. *IEEE International Conference on Data Mining (ICDM)*, 2025 (accepted).

Amir Ziashahabi[*], Narges Ghasemi[*], Sajjad Shahabi, John Krumm, Salman Avestimehr, Cyrus Shahabi. OSMGen: Highly Controllable Satellite Image Synthesis using OpenStreetMap Data. *NeurIPS 2025 UrbanAI Workshop*, 2025 (accepted).

Amir Ziashahabi, Baturalp Buyukates, Artan Sheshmani, Yi-Zhuang You, Salman Avestimehr. Frequency Domain Diffusion Model with Scale-Dependent Noise Schedule. *IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2024.

Hamza Saleem, **Amir Ziashahabi**, Muhammad Naveed, Salman Avestimehr. Hawk: Accurate and Fast Privacy-Preserving Machine Learning Using Secure Lookup Table Computation. *Privacy Enhancing Technologies Symposium (PETS)*, 2024.

Artan Sheshmani, Yi-Zhuang You, Baturalp Buyukates, **Amir Ziashahabi**, Salman Avestimehr. Renormalization Group flow, Optimal Transport and Diffusion-based Generative Model. *Physical Review E*, 2024.

Shen Li, Pritam Damania, Luca Wehrstedt, [et al., including **Amir Ziashahabi**]. PyTorch RPC: Distributed Deep Learning Built on Tensor-Optimized Remote Procedure Calls. *Proceedings of Machine Learning and Systems (MLSys)*, 2023.

Amir Ziashahabi, Chaoyang He, Salman Avestimehr. Understanding the Communication Backends of Federated Learning, under review.

[*]Equal contribution

Teaching Experience

Applied and Cloud Computing for Electrical Engineers — Teaching Assistant; University of Southern California
Foundations of Blockchain Technology — Teaching Assistant; Sharif University of Technology
Ethereum Programming Workshops — Instructor; Various Locations
Iran Blockchain Summer School — Co-organizer and Lecturer; Sharif University of Technology
Hands-on Blockchain Workshop — Co-organizer and Mentor; Sharif University of Technology
Database Design — Teaching Assistant; Sharif University of Technology
Compiler Design — Teaching Assistant; Iran University of Science and Technology

Selected Projects and Coursework

Pivot-Aware Speculative Decoding *(Python, PyTorch, Transformers, vLLm, GPT-Fast)*
Speculative decoding that rejects only pivot-critical tokens to keep accuracy while boosting throughput.
GeoToken: Hierarchical Image Geolocalization *(Python, PyTorch, Transformers)*
Geolocalization framed as next-token prediction for accurate location inference.
OSMGen: Controllable Satellite Image Synthesis *(Python, PyTorch, Transformers, Diffusers)*
Satellite image generation conditioned on OpenStreetMap for structure-faithful, controllable outputs.
MobiZO: Efficient LLM Fine-Tuning at the Edge *(Python, PyTorch, Transformers, PEFT, ExecuTorch)*
Edge-focused fine-tuning leveraging inference engines to remove backpropagation and cut VRAM on mobile/edge hardware.
FDDM: Frequency-Domain Diffusion *(Python, PyTorch)*
Diffusion in the frequency domain with a scale-dependent noise schedule; achieving faster training and sampling.
Hierarchical Federated Learning *(FedML, PyTorch, PyTorch RPC, MPI, gRPC, MQTT)*
Added hierarchical FL to FedML OSS stack by implementing distributed training primitives.
Software Fault Prediction *(R)*
Designed and evaluated a concise fault-prediction method on software-evolution data.
Toy Language + JVM Backend *(ANTLR, Jasmin)*
Built a small language with a compiler front-end and JVM bytecode backend.
Hunterguh: 2D Multiplayer *(Unity Game Engine, C#)*
Implemented 2D multiplayer action game.
2D Game Engine (from scratch) *(Python, PyGame, PyQt)*
Lightweight engine with user interface, input handling, and a scene/update loop.
Arcade Game *(C++)*
Implemented fixed shooter arcade game from scratch.
ChaapArt: Card-Design Web App *(PHP, Laravel)*
Co-founded and built the backend for a templated card-design platform.