Fatih İlhan
Resume

School of Computer Science, College of Computing
Georgia Institute of Technology, Atlanta, GA, USA

e-mail: filhan@gatech.edu
web: fatih-ilhan.github.io
github: github.com/fatih-ilhan
google scholar profile: DHB3X18AAAAJ
orcid id: 0000-0002-0173-7544

Research Interests

Efficient and Scalable AI Inference/Finetuning Systems, AI Workload Management, Distributed/Federated Learning, Edge-Cloud Computing

Academic Experience

Georgia Institute of Technology **Atlanta, GA, USA**

Ph.D. in Computer Science, CGPA: 3.84/4.00, Supervisor: Prof. Ling Liu August 2021 – Present

- Research focus areas: efficient inference/finetuning, federated/distributed learning, large language models, computer vision systems, AI safety and alignment
- Published 13 papers (5 as first author) in top venues such as CVPR, NeurIPS, ICLR, EMNLP, WACV, WWW and ICDCS.
- Served as reviewer for CVPR, AAAI, ICML, ICDCS, IEEE PAMI and IEEE TOIT.
- Head TA for the Advanced Internet Systems course with 5 TAs and 100-150 students, selected as the outstanding Head TA for OMSCS program.

Bilkent University **Ankara, Türkiye**

M.Sc. in EEE, CGPA: 3.58/4.00, Supervisor: Prof. Serdar Kozat September 2019 – August 2021

- Thesis: Nonstationary Time Series Prediction with Markovian Switching RNNs
- Research focus areas: Nonstationary time series prediction, spatiotemporal event modeling
- Published 3 papers in top IEEE journals, served as reviewer for IEEE TNNLS and IEEE TSP.
- Served as TA for the courses: Statistical Learning and Data Analytics, Neural Networks.

B.Sc. in EEE, CGPA: 3.81/4.00 January 2018 – June 2019

- Senior Project: GPS-independent outdoor localization system
- Specialization in signal processing, machine learning, communications
- Attended exchange program at Nagoya University, Japan (Spring 2018) and studied intelligent automobile systems.

Ankara Science High School **Ankara, Türkiye**

High School Degree, Natural Sciences Track, CGPA: 95.26/100 September 2010 – June 2014

Work Experience

IBM Thomas J. Watson Research Center **Yorktown Heights, NY**

Research Intern, Mentor: Dr. Gong Su, Manager: Dr. Donna Dillenberger May-Aug 2022/2023/2024

- I worked on memory-efficient decoding with KV caching compression for long-context inference with LLMs (Summer 2024)
- I researched efficient pruning for LLM finetuning in resource-constrained environments (Summer 2023)
- I worked on computation-efficient federated learning under heterogeneous settings with resource-constrained devices (Summer 2022)

DataBoss Analytics **Ankara, Türkiye**

Machine Learning Engineer August 2018 – July 2021

- Built end-to-end AI pipelines for a large-scale online prediction and anomaly detection system - Predy.AI - and analyzed complex spatiotemporal traffic, crime, weather and consumption data.

Roketsan **Ankara, Türkiye**

Engineering Intern June 2017 – July 2017

- Worked on integrating GPS and INS data using Extended Kalman Filter.
- Built a Labview application that enables communication with a GPS receiver and displays/records the position, velocity, heading and time data.

Conference Papers

[C19] T. Huang, S. Hu, **F. Ilhan**, S. F. Tekin, and L. Liu, "Booster: Tackling Harmful Fine-tuning for Large Language Models via Attenuating Harmful Perturbation", *International Conference on Learning Representations (ICLR)*, 2025. *(oral)*

[C18] **F. Ilhan**, G. Su, S. F. Tekin, T. Huang, S. Hu, and L. Liu, "Resource-Efficient Transformer Pruning for Finetuning of Large Models", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[C17] **F. Ilhan**, KH. Chow, S. Hu, T. Huang, S. F. Tekin, W. Wei, Y. Wu, M. Lee, R. Kompella, H. Latapie, G. Liu, L. Liu, "Adaptive Deep Neural Network Inference Optimization with EENet", *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.

[C16] T. Huang, S. Hu, **F. Ilhan**, S. F. Tekin and L. Liu, "Lazy Safety Alignment for Large Language Models against Harmful Fine-tuning", *Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

[C15] S. F. Tekin, **F. Ilhan**, T. Huang, S. Hu and L. Liu, "LLM-TOPLA: Efficient LLM Ensemble by Maximising Diversity", *ACL Conference on Empirical Methods in Natural Language Processing (EMNLP Findings)*, 2024.

[C14] KH. Chow, Sihao Hu, Tiansheng Huang, **Fatih Ilhan**, Wenqi Wei, and Ling Liu, "Diversity-driven Privacy Protection Masks Against Unauthorized Face Recognition", *Privacy Enhancing Technologies Symposium (PETS)*, 2024

[C13] **F. Ilhan**, G. Su, Q. Wang and L. Liu, "Scalable Federated Learning with System Heterogeneity", *IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2023. *(demo)*

[C12] **F. Ilhan**, G. Su and L. Liu, "ScaleFL: Resource-Adaptive Federated Learning with Heterogeneous Clients", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[C11] **F. Ilhan**, S. F. Tekin, S. Hu, T. Huang, KH Chow, L. Liu, "Hierarchical Deep Neural Network Inference for Device-Edge-Cloud Systems", *ACM International World Wide Web Conference (WWW)*, 2023. *(poster)*

[C10] T. Huang, S. Hu, KH. Chow, **F. Ilhan**, S. F. Tekin and L. Liu, "Lockdown: Backdoor Defense for Federated Learning with Isolated Subspace Training", *Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

[C9] KH. Chow, L. Liu, W. Wei, **F. Ilhan** and Y. Wu, "STDLens: Securing Federated Learning Against Model Hijacking Attacks", *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[C8] W. Wei, L. Liu, KH. Chow, **F. Ilhan** and Y. Wu, "Model Cloaking against Gradient Leakage", *IEEE International Conference on Data Mining (ICDM)*, 2023.

[C7] S. Hu, T. Huang, **F. Ilhan**, S. F. Tekin, L. Liu, "Large Language Model-Powered Smart Contract Vulnerability Detection: New Perspectives", *IEEE International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (IEEE TPS-ISA)*, 2023.

[C6] **F. Ilhan**, S. F. Tekin and B. Aksoy, "Spatio-Temporal Crime Prediction via Temporally Hierarchical Convolutional Neural Networks", *28th IEEE Signal Processing and Communications Applications Conference*, 2020.

[C5] **F. Ilhan**, N. M. Vural and S. S. Kozat, "LSTM-Based Online Learning with Extended Kalman Filter Based Training Algorithm", *28th IEEE Signal Processing and Communications Applications Conference*, 2020.

[C4] **F. Ilhan** and E. Mumcuoglu, "Performance Analysis of Semi-Supervised Learning Methods under Different Missing Label Patterns", *28th IEEE Signal Processing and Communications Applications Conference*, 2020.

[C3] **F. Ilhan**, S. F. Yilmaz and S. S. Kozat, "A Two-Stage Multi-Class Classification Approach Based on Anomaly Detection", *28th IEEE Signal Processing and Communications Applications Conference*, 2020. *(poster)*

[C2] N. M. Vural, B. Altas, **F. Ilhan** and S. S. Kozat, "Shortest Path Learning in Non-Stationary Environments via Online Convex Optimization", *28th IEEE Signal Processing and Communications Applications Conference*, 2020.

[C1] N. M. Vural, B. Altas, **F. Ilhan** and S. S. Kozat, "Online Shortest Path Learning via Convex Optimization", *28th IEEE Signal Processing and Communications Applications Conference*, 2020.

Journal Papers

[J3] **F. Ilhan**, O. Karaahmetoglu, I. Balaban and S. S. Kozat, "Markovian RNN: An Adaptive Time Series Prediction Network with HMM-based Switching for Nonstationary Environments", *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[J2] N. M. Vural, **F. Ilhan**, S. F. Yilmaz, S. Ergüt and S. S. Kozat, "Achieving Online Regression Performance of LSTMs with Simple RNNs", *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[J1] **F. Ilhan** and S. S. Kozat, "Modeling of Spatio-Temporal Hawkes Processes with Randomized Kernels", *IEEE Transactions on Signal Processing*, 2020.

Preprints

[P11] **F. Ilhan**, G. Su and L. Liu, "Memory-Efficient Decoding with KV Cache Compression for Long-Context LLMs", *in progress*, 2025.

[P10] **F. Ilhan**, S. F. Tekin, S. Hu, T. Huang and L. Liu, "Fed4LM: Efficient Federated Finetuning under Data and Resource Heterogeneity with a Mixture of Masked Adapters", *in progress*, 2025.

[P9] Z. Yahn, S. F. Tekin, **F. Ilhan**, S. Hu, T. Huang, M. Loper and L. Liu, "Attention-Based Adversarial Attacks on Large Vision Transformers for Object Detection", *in progress*, 2025.

[P8] T. Huang, S. Hu, **F. Ilhan**, S. F. Tekin, and L. Liu, " Harmful Fine-tuning Attacks and Defenses for Large Language Models: A Survey", *in progress*, 2025.

[P7] T. Huang, S. Hu, **F. Ilhan**, S. F. Tekin, and L. Liu, " Virus: harmful fine-tuning attack for Large Language Model bypassing Guardrail Moderation", *in progress*, 2025.

[P6] T. Huang, S. Hu, **F. Ilhan**, S. F. Tekin, W. Wei, and L. Liu, "Backdoor Defense for Decentralized Learning with Fisher Information Guidance", *in progress*, 2025.

[P5] S. F. Tekin, **F. Ilhan**, T. Huang, S. Hu, Z. Yahn and L. Liu, "H^3 Fusion : Helpful, Harmless, Honest Fusion of Pretrained-LLMs", *in progress*, 2025.

[P4] S. F. Tekin, **F. Ilhan**, T. Huang, S. Hu and L. Liu, "Multi-Agent Reinforcement Learning with Focal-Diversity Optimization", *in progress*, 2025.

[P3] S. F. Tekin, **F. Ilhan**, T. Huang, S. Hu and L. Liu, "FusionShot: Boosting Few Shot Learners with Focal-Diversity Optimized Ensemble Method", *in progress*, 2025.

[P2] S. Hu, T. Huang, **F. Ilhan**, S. F. Tekin and L. Liu, "A Survey on Large Language Model-Based Game Agents", *in progress*, 2025.

[P1] S. Hu, T. Huang, KH. Chow, **F. Ilhan**, S. F. Tekin and L. Liu, "Linking Ethereum Accounts with Pseudo-supervised Pre-trained Language Models", *in progress*, 2025.

Awards and Honors

- 191st among 2M high school graduates in University Entrance Examination.
- 80th among 0.2M university graduates in ALES (National GRE).
- Full Scholarship from the Scientific and Technological Research Council of Türkiye for M.Sc. studies.
- JASSO Scholarship for Exchange Program at Nagoya University.
- Full Scholarship from Bilkent University during B.Sc. and M.Sc. Studies.
- Bilkent University High Honor Student during B.Sc. Studies.

Skills

Programming: Python, SQL, R, C++, Java, Matlab, Assembly (8051), VHDL
Tools: Deep Learning Libraries (Tensorflow, PyTorch, Keras), MLOps Tools (Kubernetes, Polyaxon, MLFlow), Other Tools (Docker, Flask, Django, Kafka, Spark), Agile (Gitlab, Atlassian Tools)
Test Scores: TOEFL iBT: 108, GRE: 149/170/3.5

Languages: Turkish (Native), English (Advanced), Japanese (Lower intermediate ∼N4)

Social Activities

- Bass Guitarist in “Parallel Park” (2022-2023)
- Bilkent University Music Club Member (2014-2017)
- Bass Guitarist in “Freud Goes Technical” (2014-2017)
- Bilkent IEEE Student Branch Member (2014-2016)
- Bilkent University Open Software and Internet Technologies Club Member (2014-2015)
- Ankara Science High School Electronics Club Member (2012-2014)
- Ankara Science High School Physics Olympiads Team Member (2010-2012)

Hobbies

- Backpacking, overnight camping, being on the road
- Playing bass, discovering new music genres