Ismet Dagli

Computer Science Department
Colorado School of Mines

Mail: ismetdagli@mines.edu
Phone: +1 951 446 67 24
Address: Golden, CO, USA

Short Bio & Research Interest

I'm currently working at Microsoft as Sr. Supercomputer Software Engineer on Microsoft Azure AI high performance computing team. I received my PhD at Colorado School of Mines, advised by Dr. Mehmet Belviranli. My broader research interest spans high-performance computing and energy-efficient computing on heterogeneous platforms, deep learning (DL) inference acceleration, inference optimization and model deployment, scalable resource management and scheduling, and edge/cloud distributed systems for DL workloads. I have published several works in **DAC, ICS, PPoPP, DATE, and ICCAD**. My research studies were recognized by ACM *MICRO* and *CGO* SRC (Bronze Medals). I was selected as **2024 MLCommons ML and System Rising Stars**. I presented my studies at **NVIDIA GTC'25.**

Education

Colorado School of Mines **Jan 2020 - Aug 2025**

- *Doctor of Philosophy in Computer Science* *Golden, CO*
  Thesis Title: Enabling Collaborative, Efficient, and Scalable Resource Management for Heterogeneous Edge Computing Architectures
  GPA: 3.84

Boğaziçi University **Sept 2015 - Dec 2019**

- *Bachelor of Science in Computer Engineering* *Istanbul, Turkey*

Work Experience

Microsoft **Oct 2025 – Present**
Senior Supercomputing Software Engineer *Seattle, WA*

- Advisor: Dr. Prabhat Ram
- Working with Microsoft Azure AI high performance computing team.
- Working on looking for systems engineers to enable customers in deploying, monitoring, profiling, and debugging their applications on hyperscale cloud infrastructure. Azure enables the largest supercomputing deployments to tackle complex computational problems in public cloud.

Colorado School of Mines **Jan 2020 – July 2025**
Research Assistant *Golden, CO*

- Advisor: Mehmet E. Belviranli
- Engineered and implemented novel scheduling and resource management frameworks for AI/DL workloads (including CNNs, RNNs, and multi-DNN systems) on heterogeneous accelerators (GPUs, DSAs) on edge platforms (NVIDIA Jetson Orin/Xavier, Qualcomm Snapdragon). Achieved up to 65% energy efficiency gains and 32% performance improvements through optimized mapping techniques and contention-aware runtime systems.
- Utilized PyTorch/TensorRT/Tensorflow/CoreML with ONNX. I also used profiling techniques for several AI workloads (including DNN and traditional ML applications) on heterogeneous platforms for execution time, memory utilization, power consumption and extensively used Nsight systems and Nsight Compute.
- Designed graph-based, multi-layer hardware representation scheme that is capable of expressing arbitrary topologies and multi-tiered decentralized orchestrator mechanism that scalably finds a mapping of a task to local or remote processing units for distributed computing and edge-cloud environments, reaching up to

99.2% accuracy.

Colorado School of Mines **Jan 2024 – May 2024**
Adjunct Faculty for CSCI564: Advanced Computer Architecture *Golden, CO*

- Primary instructor in a grad-level class of nearly 100 students, CSCI564: Advanced Computer Architecture. I was responsible for teaching the entire class and managing all course workloads.

Pacific Northwest National Laboratory **May 2021 – Dec 2021**
Ph.D. Intern *Richland, WA*

- **Supervisor: Antonino Tumeo**
- Leveraged MLIR to apply compiler optimizations and developed evolutionary algorithms for High-Level Synthesis (HLS) within the SODA toolkit, aimed at designing and optimizing custom hardware accelerators for ML workloads.

Colorado School of Mines **August 2020 – May 2021**
Teaching Assistant *Golden, CO*

- Taught as a TA in a senior class (CSCI442: Operating System) of more than 250 students for two semesters.
- Responsible for grading the homeworks/projects/midterm/final as well as managing weekly office hours.

Tubitak Research Center **August 2019 – Dec 2019**
Part-Time Artificial Intelligence Engineer *Istanbul, Turkey*

- Supervisor: Dr. Ali Rıza Ekti
- Worked on recognizing the class of voices in a domestic environment on STM32 and implemented several RNN models to evaluate under several real-life conditions, reaching up to %95 F-1 score.

Baykar Tactical UAS **Jul 2018 – Sept 2018**
Software Engineer Intern *Istanbul, Turkey*

- Supervisor: Tolga Büyükyazı
- Developed a CNN semantic segmentation model on Jetson TX2 for sky objects for small-scale unmanned air vehicles (UAVs).

Publications

- **Ismet Dagli**, Justin Davis, Mehmet Belviranli, "HARNESS: Holistic Resource Management for Diversely Scaled Edge Cloud Systems", in 39th ACM International Conference on Supercomputing, **ICS 2025**
- **Ismet Dagli**, James Crea, Mehmet Belviranli, "MC^3: Memory Contention based Covert Channel Communication on Shared DRAM System-on-Chips", in Design, Automation and Test in Europe Conference, **DATE 2025**
- Alexander Cieslewicz, Ismet Dagli, Soner Seckiner, Jake Hertz, Bo Wu, Selcuk Kose, Mehmet E. Belviranli, "Title kept hidden due to anonymity", International Conference on Parallel Architectures and Compilation Techniques (PACT), **PACT 2025** ***(under submission)***
- Justin Davis, Ismet Dagli, Mehmet Belviranli, "Title kept hidden due to anonymity", Design Automation Conference, **MobiCom 2025** ***(under submission)***
- **Ismet Dagli**, Mehmet Belviranli, "Shared Memory-contention-aware Concurrent DNN Execution for Diversely Heterogeneous System-on-Chips", Proceedings of the 29th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming, **PPoPP 2024**
- Justin Mcgowen, Ismet Dagli, Neil Dantam, Mehmet Belviranli. "Scheduling for Cyber-Physical Systems with Heterogeneous Processing Units under Real-World Constraints" in 38th ACM International Conference on Supercomputing, **ICS 2024**

- Justin Mcgowen, Ismet Dagli, Neil Dantam, Mehmet Belviranli. "Constraint-aware resource management for cyber-physical systems" in Design, Automation and Test in Europe Conference as short paper, **DATE 2024.**
- Amid Morshedlou, Ismet Dagli, Jamal Rostami, Omid Moradian, Mehmet Belviranli, "Enhancing Reliability and Safety in Rock Excavation Using A Machine Learning Approach Through Wear Condition Identification" 58th US Rock Mechanics/Geomechanics Symposium, **ARMA 2024**
- H. Umut Suluhan, Serhan Gener, Alexander Fusco, Joshua Mack, Ismet Dagli, Mehmet E. Belviranli, Cagatay Edemen, Ali Akoglu. Title: A Runtime Manager Integrated Emulation Environment for Heterogeneous SoC Design with RISC-V Cores, in Heterogeneity in Computing Workshop (HCW), **IPDPS workshop 2024**
- Amid Morshedlou, Ismet Dagli, Austin Olltmans, Andrew Petruska, Mehmet Belviranli, Jamal Rostami, "Enhancing Safety Using Energy-Efficient Machine Learning Algorithms Through Prediction of Rock Type and Cutter Wear" Society for Mining, Metallurgy & Exploration: **SME MINEXCHANGE 2024**
- **Ismet Dagli**, Andrew Depke, Andrew Mueller, Sahil Hassan, Ali Akoglu, Mehmet Belviranli, "Contention-aware Performance Modeling for Heterogeneous Edge and Cloud Systems", 3rd workshop on Flexible Resource and Application Management on the Edge (FRAME), **HPDC Workshop 2023**
- **Ismet Dagli**, Alexander Cieslewicz, Jedidiah McClurg, Mehmet E. Belviranli, "AxoNN: Energy-Aware Execution of Neural Network Inference on Multi-Accelerator Heterogeneous SoCs", 59th ACM/IEEE Design Automation Conference, **DAC 2022**
- Justin McGowen, Ismet Dagli, Mehmet Belviranli, Neil Dantam; "Representations for Scheduling of Heterogeneous Computation to Support Motion Planning"; Implicit Representations for Robotic Manipulation, **RSS Workshop 2022**
- Antonino Tumeo, Marco Minutoli, Vito Giovanni Castellana, Limaye Ankur, Tan Cheng, **Ismet Dagli,** Nicolas Bohm Agostini, Serena Curzel, Amatya Vinay, Manzano Joseph; "Accelerating Data Processing at the Edge with Extreme Specialization"; 2022 Advanced Scientific Computing Research Workshop on the Management and Storage of Scientific Data, **ASCR Workshop 2022**
- **Ismet Dagli**, Mehmet E. Belviranli, "Multi-accelerator neural network inference in diversely heterogeneous embedded systems"; 2021 IEEE/ACM Redefining Scalability for Diversely Heterogeneous Architectures Workshop (RSDHA), **SC Workshop 2021**
- Serena Curzel, Nicolas Bohm Agostini, Shihao Song, Ismet Dagli, Ankur Limaye, Cheng Tan, Marco Minutoli, Vito Giovanni Castellana, Vinay Amatya, Joseph Manzano, Anup Das, Fabrizio Ferrandi, Antonino Tumeo; "Automated generation of integrated digital and spiking neuromorphic machine learning accelerators" 40th IEEE/ACM International Conference On Computer Aided Design, **ICCAD 2021**

Posters

- Ismet Dagli, Mehmet E. Belviranli, "Energy and Shared Memory Contention Aware Execution of DNN Inference on Multi-Accelerator Heterogeneous SoCs", **NVIDIA GTC 2025**
- James Crea, Ismet Dagli, Mehmet Belviranli, "Stealing Your Data: Memory-Contention Covert Channel Vulnerability in Smartphones", C-MAPP 2025
- Ismet Dagli, James Crea, Mehmet Belviranli, "MC^3: Memory Contention based Covert Channel Communication on Shared DRAM System-on-Chips", **MICRO Workshop 2024** (CWIDCA)
- Ismet Dagli**,** Mehmet E. Belviranli, "H-EYE: Holistic Performance Modeling for Diversely Scaled Systems", Student Research Competition (SRC) at **CGO 2024, Finalist (selected top-3)**
- Ismet Dagli, Mehmet E. Belviranli, "Layer-wise Concurrent DNN Execution Characterization and Scheduling for Heterogeneous System-on-Chips", C-MAPP 2023
- Ismet Dagli, Alexander Cieslewicz, Soner Seckiner, Jake Hertz, Bo Wu, Selcuk Kose and Mehmet Belviranli, "Extracting Neural Network Models via Contention-based Side Channel Attacks On Shared

Memory System-on-Chips, C-MAPP 2023

- Ismet Dagli, Mehmet E. Belviranli, "HaX-CoNN: Heterogeneity-aware Execution of Concurrent Deep Neural Networks", Student Research Competition (SRC) at **MICRO 2022, Finalist (selected Top-3)**
- Ismet Dagli, Mehmet E Belviranli, "Multiple Neural Network Inference on Heterogeneous SoCs", GRADS 2022
- Ismet Dagli, Mehmet E Belviranli, "Energy-Aware Execution of Neural Network Inference on Multi-Accelerator Heterogeneous SoCs", C-MAPP 2022
- Ismet Dagli, H. Levent Akin, "Increasing the Localization Performance via Semantic Segmentation" CMPE BOUN, Poster Presentation, 2019

Awards

- **MLCommons Rising Star** (selected 41 out of 170 applicants, ~24%)
- CGO'24 ACM Student Researech Competition (SRC) 2024, Finalist (selected as 3rd)
- MICRO'22 ACM Student Research Competition (SRC) 2022, Finalist (selected as 3rd)
- Travel awards (SC'24/IPDPS'25, MICRO'24, PPoPP'24, HPDC'23, STOC'23, GSG Mines'23/21)
- ScienceSlam@SC21, Full registration and all expenses and prize money for SC21 conference
- Stipend during bachelor's degree by the Turkish government, Turkey, 2015-2019.

Talks/Presentations

- Guest lecturer for a seminar talk, Ozyegin University, 2024
- ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP) Conference Paper Presentation, 2024
- International Symposium on Code Generation and Optimization (CGO) SRC, Finalist Talk, 2024
- Guest Lecturer at Heterogeneous Computing (CSCI-598), Colorado School of Mines, 2023
- Flexible Resource and Application Management on the Edge (FRAME), HPDC Workshop, Workshop Paper Presentation, 2023
- IEEE/ACM International Symposium on Microarchitecture (MICRO) SRC, Finalist Talk, 2022
- ACM/IEEE Design Automation Conference (DAC), Conference Paper Presentation, 2022
- IEEE/ACM Redefining Scalability for Diversely Heterogeneous Architectures Workshop (RSDHA), SC Workshop, Workshop Paper Presentation, 2021

Group (as mentor) & Personal Projects

I have mentored **one master's, ten undergraduates, and four high-school researchers**. I had four papers and poster publications with them, including their co-authorships on

- DATE'25 conference paper,
- HPDC'23 workshop paper,
- MICRO'24 workshop poster,
- PPoPP'24 artifact and open-sourcing the project,
- Several posters at Mines' courses and seminars

High-school research mentor (Role: Team Lead) **May 2023 - Aug 2025**

- I lead four students, including one female student, to develop efficient LLM runtime selection model mechanism. One of the mentees won

Fatima Fellowship mentor (Role: Team Lead) **Sept 2023 - Present**

- I lead a group of students, underrepresented and international three students including two female

students, to run LLM models at the edge devices to run inference by intelligently managing data communication through DRAM and storage.

Google ExploreCSR mentor (Role: Team Lead) **Jan 2023 - May 2024**

- I led two senior and junior undergrads to run DNN efficiently on Google Coral Dev Boards.

Gumus R&G Autonomous Car Team (Role: Software Team lead) **Aug 2020 - Aug 2024**

- I lead a team of up to 30 students for autonomous vehicle competition (Robotaxi). I lead the software team for any task, such as object detection and recognition, SLAM, path planning, and simulation.

Skills

- **Programming Skills:** C++ (14/17/20), Python, CUDA, C, OpenMP, MPI, Java, ROS, Verilog,
- **Tools & Technologies:** TensorRT, TensorRT-LLM (TRT-LLM), PyTorch, Tensorflow, ONNX/ONNXRuntime, Nsight System, Nsight Compute, Perf, MLIR, JAX, Jetson platforms, Linux platforms, Z3 Solver, Keras, Caffe, OpenCV
- **Grant/proposal writing:** I helped my PhD advisor with several grant/proposal writing for research institutions and collaboration proposals with industry companies

Service

- **Committee Member:**
  - AE Committee member for FAST'25
  - AE Committee member for EuroSys'25 Fall and Spring
- **Organizing Committee:**
  - RSDHA'21 web chair
  - RSDHA'22 web chair
  - RSDHA'23 web chair
- **External Reviewer for conferences:**
  - HiPC'25 Poster,
  - AI4SC'24,
  - DAC'2023,
  - ICS'2023,
  - ISC-HPC'23,
  - RSDHA'23,
  - DAC'22,
  - HIPS'22
  - ICS'22
  - RSDHA'22
- **Reviewer for journals**
  - IEEE Transactions on Parallel and Distributed Systems (TPDS)
  - ACM Transactions on Embedded Computing Systems (TECS)
  - IEEE Access
  - Transactions on Cloud Computing (TCC)
  - Parallel Computing (PARCO) Journal
  - Journal of System Architecture (JSA)
- **Grant reviewer for:**
  - US Army Grant Reviewer (2024)