Sina Mahdipour Saravani

CONTACT — *E-mail:* sina.mpsaravani@gmail.com & sina@cs.utah.edu *Homepage:* sinamps.github.io

INTERESTS & EXPERTISE

- High-Performance Computing for Deep Learning; Machine Learning Compilers
- Efficient Natural Language Processing (NLP)

EDUCATION

University of Utah, Salt Lake City, United States

- Ph.D., Computer Science, In Progress, GPA: **4.0/4.0** 2022 - Present
  - ◇ Coursework done. Research on Efficient and High-Performance Deep Learning.

Colorado State University, Fort Collins, United States

- M.S., Computer Science, GPA: **4.0/4.0** 2020 - 2022
  - ◇ Thesis: *Redundant Complexity in Deep Learning: An Efficacy Analysis of NeXtVLAD in NLP*

Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

- B.Sc., Computer Engineering, GPA: **16.61/20.0** 2014 - 2019
  - ◇ Thesis: *Implementation of FPGA Accelerators for Convolutional and Pooling Layers of a Convolutional Neural Network (CNN)*

PUBLICATIONS & MANUSCRIPTS

Google Scholar Link: https://scholar.google.com/citations?user=-32OFf8AAAAJ&hl=en

- Accelerated Auto-tuning of GPU Kernels for Tensor Computations. *ICS 2024*.
  Chendi Li*, Yufan Xu*, **Sina Mahdipour Saravani**, Saday Sadayappan
- Empirical Analysis of Matrix Factorization Methods for Pre-trained Transformers. arXiv.
  Ashim Gupta, **Sina Mahdipour Saravani**, Saday Sadayappan, Vivek Srikumar
- An Investigation into the Contribution of Locally Aggregated Descriptors to Figurative Language Identification. *EMNLP 2021 Workshop*.
  Sina Mahdipour Saravani, Ritwik Banerjee, and Indrakshi Ray
- Automated Identification of Social Media Bots using Deepfake Text Detection. *ICISS 2021*.
  Sina Mahdipour Saravani, Indrajit Ray, and Indrakshi Ray
- A Generalized Method for Automated Multilingual Loanword Detection. *COLING 2022*.
  A. Nath*, **Sina Mahdipour Saravani***, I. Khebour, S. Mannan, Z. Li, and N. Krishnaswamy.
- Automated Code Extraction from Discussion Board Text Dataset. *ICQE 2022*.
  Sina Mahdipour Saravani, Sadaf Ghaffari, Yanye Luther, James Folkestad, and Marcia Moraes.

RESEARCH & WORK EXPERIENCES

- **Research Assistant, University of Utah** 2022 - Present
  - ⋆ *Supervisor:* **Dr. Saday Sadayappan**
  - ◇ **Fast Auto-tuning for Matmul and Convolution GPU Kernels for Deep Learning**
    Working on a coordinate descent search algorithm for fast auto-tuning with proxy performance modeling and characterization of the problem spaces.
  - ◇ **Scalable Tiled Processing of Huge Inputs through Convolutional Neural Networks**
    Enabling processing of whole scale imaging scans for cancer detection with CNNs.
  - ◇ **Automated Distribution of Compute and Data on multi-Node, multi-GPU Systems**
    Working on automated discovery of efficient and optimized distribution strategies.
  - ◇ **Snake-Pattern Development of Multiple Matmul Blocks for Cerebras CS-2 Chip**
    Achieved performance improvements compared to SUMMA in simulations.
  - ◇ **Low-Rank Factorization Methods for Distributed Large Language Models (LLMs)**
    Empirically evaluated low-rank factorization methods, including Monarch matrices vs. SVD.
- **Data Science AI-CV-NLP, Intern, Ancestry.com Operations Inc.** Summer 2024
  - ◇ **Multilingual Coreference and Entity Resolution**
    Researched, designed and deployed an state-of-the-art encoder-based mixture-of-experts coreference resolution model working on unannotated input text with innovative expert design suitable for internal data types.

- **Research Assistant, Colorado State University** 2020 - 2022
  - ⋆ *Supervisors:* **Dr. Indrakshi Ray, Dr. Ritwik Banerjee, Dr. Nikhil Krishnaswamy**
  - ◇ **Machine Translation for Similar Low-Resource Language Pairs with Loan Words**
    Studied the potential benefits of using loan words, both as a knowledge base and as insights to architecture design, for automated machine translation between similar language pairs.
  - ◇ **An Investigation into the Contribution of VLAD to Figurative Language Identification**
    Investigated the application and effectiveness of vector of locally aggregated descriptors on top of Transformer layers. Studied sarcasm detection in Twitter as a use case.
  - ◇ **Deepfake Text Detection for Social Media Bot Identification**
    Implemented Transformer-based models to detect bot-generated text on a deepfake dataset resulting in performance improvements by using domain-specific pre-trained models.
  - ◇ **Automated Code Extraction from Discussion Board Text Dataset**
    Developed algorithms to extract topic codes from course discussion datasets.
- **Research Assistant, Amirkabir University of Technology** 2018 - 2019
  - ⋆ *Supervisor:* **Dr. Reza Safabakhsh**
  - ◇ **FPGA Accelerators for Convolutional and Pooling Layers of a CNN**
    Researched and implemented an FPGA accelerator for the convolutional and max pooling functions of CNNs using High-Level Synthesis. It was deployed on a ZYBO SoC board and achieved up to 30× speedup compared to the equivalent software code on a CPU.
- **NLP Research & Development Intern, CommentMiner** 2017 - 2018
  - ◇ **NLP Microservices for the Persian Language**
    Implemented text-processing microservices (topic classification, profanity detection, NER, and sentiment analysis) and a question-answering chat bot for the Persian language.

Teaching & Mentoring Experiences

- **Temporary Teaching Faculty**, University of Nevada, Las Vegas Summer 2020
  - ◇ Computer Science II (CS 202) course on C++, Primary Instructor
- **Teaching Assistantship and Student Mentorship**
  - ◇ Fault-Tolerant Computing (CS 530) course, Colorado State University, Spring 2022
  - ◇ Data Mining (CS 458/658) course, University of Nevada, Las Vegas, Spring 2020
  - ◇ Embedded & Real-Time Systems course, Amirkabir University of Technology, Fall 2018
  - ◇ Mentored two graduate, five undergraduate and two high school students for research in NLP and one student in the i-STEM Scholars program.

Professional Services

- **Reviewer** for the following conferences:
  - ◇ COLING 2025
  - ◇ LREC-COLING 2024
  - ◇ WebConf 2021 & 2022
  - ◇ ICQE 2022
  - ◇ ICDCS 2021
  - ◇ ACISP 2021
  - ◇ IEEE TPS 2021 & 2020
  - ◇ IEEE S&P 2020
- **Industry Relations Officer**, Scientific Association and Olympiad Affairs Office of AUT 2015

Honors & Awards

- **UNLV Access Grant**, University of Nevada, Las Vegas 2020
- **Top 50 start-ups** in GITEX start-ups competition, UAE (CommentMiner) 2017
- **3rd place** in ElecomStars start-ups competition, Iran (CommentMiner) 2017
- **1st place grant** in Sharif VC Cup start-ups competition, Iran (CommentMiner) 2017
- Ranked top **0.2**% in Nationwide University Entrance Exam in Math. & Physics, Iran 2014

Relevant Skills

- Programming: Python, C/C++, Java
- Tools and Frameworks: TVM/Ansor, DeepSpeed, PyTorch, HuggingFace Transformers, TensorFlow, spaCy, scikit-learn, MALLET, Stanford NLP, polyglot, NLTK, OpenMP, CUDA, Docker
- Others: LaTeX, Bash, Vivado and Hardware Design Softwares