Muhammad Adnan

6335 Thunderbird Crescent, V6T 2G9, Vancouver, BC, Canada
☎ +1(236) 978-1115
✉ adnan@ece.ubc.ca
☞http://people.ece.ubc.ca/adnan/

EDUCATION

Ph.D., Electrical and Computer Engineering Aug 2026 (Expected)
University of British Columbia, Vancouver, BC, Canada
Research Area: System Design Optimizations for large Machine Learning Models
Advisor: Prashant J.Nair

M.A.Sc., Electrical and Computer Engineering Nov 2021
University of British Columbia, Vancouver, BC, Canada
Thesis: Accelerating input dispatching for deep learning recommendation models training

B.E., Electrical Engineering July 2013
National University of Sciences and Technology, Islamabad, Pakistan

RESEARCH

Artificial intelligence and machine learning is an essential part of any technology - from self driving cars to E-commerce - any technology is using AI and ML in some way. Currently most of the machine learning jobs are executed in data centers using state of art CPU's and GPU's in a distributed fashion with software frameworks not optimized for such workloads.

Vision

The landscape of deep learning models has undergone significant changes in the last few years, characterized by a steady increase in the size of these models. Leading the way in this trend are recommendation models and large language models, which have become increasingly prevalent and influential in the field of artificial intelligence. The emergence of such large models has brought with it a host of system-level challenges, particularly in relation to their training and inference. Training these models across distributed computational resources remains an open problem, while the deployment of large models for inference poses its own set of challenges. My research interest centers at the intersection of architecture and systems, with a particular focus on addressing the challenges posed by large deep-learning models.

Interests

Recommender Systems, Ranking, Large Language Models (LLMs), Systems for ML, Hardware Accelerators, Machine Learning, High Performance Computing.

CONFERENCE Publications

1. "Heterogeneous Acceleration Pipeline for Recommendation System Training", *51st International Symposium on Computer Architecture* **(ISCA'24)**
   Muhammad Adnan, Yassaman Ebrahimzadeh Maboud, Divya Mahajan, Prashant J.Nair

2. "Keyformer: KV Cache reduction through key tokens selection for Efficient Generative Inference", *7th Annual Conference on Machine Learning and Systems* **(MLSys'24)**
   Muhammad Adnan, Akhil Arunkumar, Gaurav Jain, Prashant J.Nair, Ilya Soloveychik, Purushotham Kamath

3. "Accelerating Recommendation System Training by Leveraging Popular Choices", *48th International Conference on Very Large Data Bases* **(VLDB'22)**
Muhammad Adnan, Yassaman Ebrahimzadeh Maboud, Divya Mahajan, Prashant J.Nair

4. "The Mindful Recommender: Adaptive Feature Interaction Selection for Deep Sparse Networks", *In submission at* **(KDD'24)**
Muhammad Adnan, Yassaman Ebrahimzadeh Maboud, Divya Mahajan, Prashant J.Nair

5. "Gliding the Slipstream: Popularity-Based Embedding Skipping for Recommender Training", *In submission at* **(MLSys'24)**
Yassaman Ebrahimzadeh Maboud, Muhammad Adnan, Divya Mahajan, Prashant J.Nair

Workshops

1. "Accelerating Recommendation System Training by Leveraging Popular Choices", *presented at Personalized Recommendation Systems and Algorithms Workshops at Fourth Conference on Machine Learning and Systems* **(MLSys'21)**
Muhammad Adnan, Yassaman Ebrahimzadeh Maboud, Divya Mahajan, Prashant J.Nair
2. "Ad-Rec: Advanced Feature Interactions to Address Covariate-Shifts in Recommendation Networks", *presented at ML for Systems Workshop at Thirty-Sixth Conference on Neural Information Processing Systems* **(NeurIPS'23)**
Muhammad Adnan, Yassaman Ebrahimzadeh Maboud, Divya Mahajan, Prashant J.Nair

PATENTS AND THESIS

1. "Workload-Aware Hardware Architecture Recommendations"
USSN # 17/965,681, *In Submission*
Inventors: Muhammad Adnan, Amar Phanishayee, Miguel Castro, Janardhan Kulkarni, Divya Mahajan

2. "Accelerating input dispatching for deep learning recommendation models training"
M.A.Sc. Thesis at University of British Columbia
Advisor: Prashant J.Nair

TALKS

- "Training Big Sparse Recommendation Models on Commodity Servers" tutorial at **HPCA 2023**.
- "Life in Grad school panelist" at Undergrad Architecture Mentoring (uArch) Workshop in conjunction with **ISCA 2021**.
- "IEEE Advanced Embedded Systems Workshop: Commercial Off the Shelf embedded systems", at Lahore University of Management Sciences (LUMS) in Dec. 2017
- "Real Time Systems Workshop: Using Real Time Operating System", at Bahrain Polytechnic in Jan. 2016
- "Embedded Systems for Communications", at Misr International University (MIU) in Nov. 2015

- "Using Real time systems for communication systems", at American University of Sharjah (AUS) in March. 2015

SERVICE

- PhD Student Reviewer, 55th IEEE/ACM International Symposium on Microarchitecture (MICRO'22).
- Reviewer, IEEE Computer Architecture Letters (IEEE CAL).
- Reviewer, IEEE Transactions on Computers (IEEE TC).

ACADEMIC EXPERIENCE

- **Graduate Research Assistant** 2019-Present
  Systems and Architectures (STAR) Lab.
- **Graduate Teaching Assistant** Fall-2019, Fall-2020
  Introduction to Computer Architecture: Undergraduate-Level
- **Graduate Teaching Assistant** Fall-2020, Summer-2020, Fall 2021
  Digital Systems Design: Undergraduate-Level
- **Graduate Teaching Assistant** Fall-2022
  Introduction to Microcomputers: Undergraduate-Level
- **Graduate Teaching Assistant** Winter-2023
  Digital Systems and Microcomputers : Undergraduate-Level

INDUSTRIAL EXPERIENCE

- **d-Matrix** May 2023 - Dec 2023
  ML Performance Intern
  Investigating performance of generative language models for inference.
  Mentor: Akhil Arunkumar
- **Microsoft Research** June 2021 - Sep 2021
  Research Intern (ML Systems)
  Investigated design space exploration of domain specific training accelerators for DNN workloads.
  Mentor: Amar Phanishayee
- **National Instruments (NI)** 2013 - 2019
  Account Manager
  Responsible for providing technical support to the customers in Middle East, preparing demo's for Sales & Marketing and managing the team of applications engineers in Pakistan.

PROJECTS
Graduate

- **Optimizing memory hierarchy for Deep Neural Networks**
  A Real Time Machine Learning project (RTML) with DARPA.
- **Optimizing the search algorithm for finding efficient mappings for DNN accelerators.**
- **Avoiding Cache Pollution from Mis-speculated Loads for efficient Cache Management.**
  Course project for Advanced Computer Architecture.
- **MAC-ECC: An Approach for an Optimized Memory Reliability**
  Course project for Security and Reliability.

HONORS AND AWARDS

- Selected as **Machine Learning and Systems Rising Star** in the 2023 cohort by **MLCommons**.
- Recipient of **Graduate Support Initiative (GSI) - 2023-2024** Award.
- Recipient of **VLDB Endowment Travel SPEND Award** for attending VLDB 2022 conference
- Nominated for **Rookie of the year at EMEIA level** at NI Week 2017
- **Certified LabVIEW Developer (CLD)**
- Nominated for **President's gold medal** for best senior year project
- Received **NUST merit based scholarship** for 4 years for academic achievements during undergraduate
- Received **Commandant's plaque of Excellence Award** being high achiever
- Finalist of **Lipton Talent Hunt (LTH)** in 2013

LEADERSHIP

- President of **Pakistani Students Association in Canada** for helping present and future Pakistani students studying in Canada.
- Lead for **Planet NI STEM Program** for promoting STEM Education in Pakistan
- Event lead for **EME Olympiad 2011 & 2012**

SKILLS

C, C++, Python, Pytorch, Bash scripting, Cadence Virtuoso, Architecture Simulators, Deep Neural Network Performance Simulators

PROFESSIONAL MEMBERSHIP

ACM

REFEREES

Prof. Prashant J.Nair
University of British Columbia

KAIS 4014
2332 Main Mall
Vancouver, BC, Canada
✉ prashantnair@ece.ubc.ca

Dr. Amar Phanishayee
Microsoft Research

One Microsoft Way
Redmond
WA 98052, USA
✉ amar@microsoft.com

Prof. Mieszko Lis
University of British Columbia

KAIS 4050
2332 Main Mall
Vancouver, BC, Canada
✉ mieszko@ece.ubc.ca

Dr. Divya Mahajan
Microsoft

One Microsoft Way
Redmond
WA 98052, USA
✉ divya.mahajan@microsoft.com