

# Ammar Ahmed

ahme0599@umn.edu

## Education

---

### University of Minnesota - Twin Cities

PhD Candidate in Computer Science

Advisor: Dr. Ali Anwar

Minneapolis, MN

Fall 2022 - Present

### Information Technology University (ITU)

Bachelor's degree in Computer Science

Thesis project: Predicting Movie's Success Through its Production Features and Social Media

Lahore - Pakistan

Fall 2013 - Spring 2017

## Publications

---

- **A. Ahmed**, A. A. Khan, A. Ahmad, S. Di, Z. Liu, and A. Anwar, "Retrieval-of-thought: Efficient reasoning via reusing thoughts," in *The Fourteenth International Conference on Learning Representations*, 2026
- X. Wang, Q. Le, **A. Ahmed**, E. Diao, Y. Zhou, N. Baracaldo, J. Ding, and A. Anwar, "MAP: Multi-human-value alignment palette," in *The Thirteenth International Conference on Learning Representations*, 2025. Oral presentation — Top 1.82% of submissions
- W. Gill, M. Elidrissi, P. Kalapatapu, **A. Ahmed**, A. Anwar, and M. A. Gulzar, "Meancache: User-centric semantic cache for large language model based web services," *IEEE 39th International Parallel and Distributed Processing Symposium*, 2025
- Z. Su, **A. Ahmed**, Z. Wang, A. Anwar, and Y. Cheng, "Everything you always wanted to know about storage compressibility of pre-trained ml models but were afraid to ask," *Proceedings of the VLDB Endowment*, vol. 17, pp. 2036–2049, 2024
- A. Said, M. Shabbir, S.-U. Hassan, Z. R. Hassan, **A. Ahmed**, and X. Koutsoukos, "On augmenting topological graph representations for attributed graphs," *Applied Soft Computing*, vol. 136, 2023
- N. Zaheer, O. U. Ahmad, **A. Ahmed**, M. S. Khan, and M. Shabbir, *SEMOUR: A Scripted Emotional Speech Repository for Urdu*. New York, NY, USA: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021
- **A. Ahmed**, Z. R. Hassan, and M. Shabbir, "Interpretable multi-scale graph descriptors via structural compression," *Information Sciences*, vol. 533, pp. 169 – 180, 2020
- A. A. Khan, **A. Ahmed**, Z. Fayyaz, S. Di, M. Hong, and A. Anwar, "Faster synchronous on-policy rl via straggler-aware group sizing," in *EMNLP*, 2026 (under revision)
- "Systematic evaluation of optimization techniques for long-context language models: From individual methods to strategic combinations," *IEEE Cluster*, 2026 (under revision)

## Experience

---

### Argonne National Laboratory

Research Aide

Lemont, IL

Summer Internship 2026, 2024

- Conducted rigorous analysis of inference optimization algorithms for LLM

### University of Minnesota - Twin Cities (UMN)

Graduate Teaching Assistant

Minneapolis, MN

- Machine Architecture and Organization (CSCI 2021) Spring 25
- Introduction to Operating Systems (CSCI 4061) Fall 2022, 23 & year 25

- Discrete Structures of Computer Science (CSCI 2011) *Spring 23 & Fall 24*

*Research Assistant @ DMLSys Lab — Advisor: Dr. Ali Anwar* *Summer 2023 & 25*

- Designed framework for chain of thought reuse for large reasoning models
- Build compression framework to reduce storage consumption of deep learning models and LLMs

**Information Technology University**

Lahore - Pakistan

*Research Assistant*

*Dec 2017 - July 2022*

- Constructed Urdu speech emotion-based dataset and prediction system with 15,040 instances.
- Designed *graph descriptors* for graph classification based on their substructures
- Designed a machine learning-based solution to *predict unreported costs of medical services*

## *Academic Projects*

---

- Developed a cost efficient framework for robotic task decomposition using Large Language Models
- Trained deep neural networks and LSTMs to identify similar questions on Quora
- Implemented the particle swarm genetic algorithm to train neural networks
- Developed an Instagram-esque social media app for Android. Integrated with Google's Firebase

## *Awards & Scholarships*

---

**Class Valedictorian - Information Technology University** *2013 - 2017*

Awarded for achieving the highest CGPA in the Class of 2017

**Dean's Honour - Information Technology University** *Fall 2015 – Spring 2017*

This honor is reserved for students achieving a minimum CGPA of 3.70 in a semester

**Punjab Edu Endowment Fund - Information Technology University** *2013 – 2017*

Scholarship for talented students governed by the *State of Pakistan*

## *Technical skills*

---

<b>Areas of expertise</b>	Deep learning, Federated learning, Large language models, Distributed machine learning, Graph theory, and Data representation
<b>Tools</b>	Python (PyTorch, Transformers, Networkx, Scikit-learn, Sympy, igraph), C/C++, Java for Android Development

## *Volunteer*

---

<b>Event Coordinator</b>	Computer Science Grad Student Association (UMN)	<i>2024 - 2025</i>
<b>Treasurer</b>	ITU Funventure Society	<i>2014 – 2016</i>

## *References*

---

**Dr. Ali Anwar**

Assistant Professor

Department of Computer Science

& Engineering

University of Minnesota - Twin Cities

aanwar@umn.edu

**Dr. Mudassir Shabbir**

Associate Professor and Chairperson

Department of Computer Science

Information Technology University

mudassir@rutgers.edu