Muhammad Ahmed Mohsin

Ph.D. Candidate, Stanford Electrical Engineering

muahmed@stanford.edu | ahmd-mohsin.github.io | building alexein.ai

Research Interests

Preference Optimization and Test-Time Scaling. Designing sample-efficient preference-learning and reinforcement-learning methods for aligning language-model reasoning, and adaptive test-time compute methods that dynamically control inference depth, tool invocation, and verification under strict budgets. I am interested in how post-training and inference-time search jointly shape reasoning, and in principled trade-offs among accuracy, latency, and reliability.

RL Infrastructure for Agentic Systems and Coding Agents. I build Crucible, an open reinforcement-learning infrastructure. It post-trains agents on long, multi-step tasks from real enterprise work. The agents operate in stateful worlds of databases, workspaces, and browsers. The runtime records each token, limits staleness, and gives credit to each step. Sandboxes isolate model-written code, and a grader-integrity protocol stops a policy from changing its own verifier. At Amazon AGI, I study how coding agents turn test-time compute into solved tasks, and how to train these behaviors with supervised and reinforcement post-training.

AI Alignment, Benchmarking, and Recursive Self-Improvement. Developing persona-based benchmarks and user-simulation frameworks that evaluate how models behave across diverse populations and interactive settings. I am broadly interested in AI alignment, including calibrated uncertainty and reducing sycophancy, and in measurement methodology that makes alignment progress observable and reproducible. I also study recursive self-improvement, measuring whether models genuinely compound when trained on their own outputs and separating real self-improvement from search and coverage effects under matched compute.

Education

Stanford University, Bay Area, California 2024–Present
Ph.D. in Electrical Engineering (LLM Post-Training, Reinforcement Learning, Agentic Architectures)

- **CGPA** 4.0/4.0
- **Advisors** Dr. Emily Fox and Dr. John Cioffi. Research collaboration with Google DeepMind, Meta MSL, Amazon AGI and Microsoft Core AI.
- **Fellowship** Qualcomm Fellowship.

National University of Sciences & Technology (NUST), Islamabad, Pakistan Aug. 2020 – June 2024
Bachelor of Engineering in Electrical Engineering

- **CGPA** 3.87/4.0 | **Specialization GPA** 4.0/4.0
- **Senior Design Thesis** Reinforcement learning for resource allocation optimization in dynamic environments
- **Rector's Gold Medal** Awarded Rector's Gold Medal for best final year project cohort (2024)

Research Experience

Amazon AGI, Applied Research Scientist June 15, 2026 – September 18, 2026
Advisor: Myeongsoo Kim | Project: Teaching Agents to Code Reliably (ICLR 2027)

- Ran a controlled study of six test-time strategies on a fixed Qwen3-Coder-30B agent over SWE-bench Verified, isolating diversity of exploration and diversity of edits as the skills that convert compute into solved bugs and showing that self-authored verification sits near chance.

- Distilled diverse, reactive search into the policy via weighted supervised fine-tuning (pass@1 $31.8 \rightarrow 35.1$, pass@8 $46.8 \rightarrow 51.2$), then trained a discriminative self-verifier with GRPO using a mutation-gauntlet reward with step-level, threshold-free credit.

Stanford Artificial Intelligence Laboratory (SAIL), Ph.D. Researcher September 2025 – Present
Advisor: Prof. Emily Fox

- **Crucible: RL Infrastructure for Enterprise Agents.** Built an open RL stack for post-training agents on long, multi-step tasks drawn from real work: an asynchronous rollout runtime with staleness bounds, token-level off-policy corrections, exact loss aggregation under FSDP2/HSDP training, and dense turn-level rewards. Agents act in stateful *Worlds* over databases, workspaces, and browsers with snapshot/restore, sandboxed execution, and a grader-integrity protocol that blocks verifier tampering.
- **General Preference Reinforcement Learning (GPRL).** Preserved multi-dimensional preference structure through the policy update by computing and independently normalizing per-dimension advantages under a General Preference Model, with a drift monitor that detects single-axis exploitation to resist reward hacking, reaching a 56.5% length-controlled win rate on AlpacaEval 2.0.
- **Coverage Preservation in RL as an Off-Policy Constraint.** Showed that on-policy RLVR (e.g. GRPO) narrows reasoning coverage because the gradient routed to a correct mode scales as $O(p_\theta(M))$, then preserved the recoverability of verified reasoning modes with an off-policy, one-sided, base-anchored support floor enforced by teacher forcing, cutting reasoning-mode collapse from 39% to 2.3% and raising the large-k continued-RL ceiling.

Samsung Semiconductors, Ph.D. Researcher June 2024 – September 2025
Advisor: Dr. John M. Cioffi | Project: Graph Neural Networks for Accelerating Low-Rank SDP Solvers (JMLR 2026)

- Developed a constraint-graph representation of SDPs with a Graph Attention Network encoder to predict rank trajectories, integrating learned rank schedules into low-rank solvers to eliminate manual heuristics and achieve up to $3\times$ speedups on large-scale benchmarks.

Selected Publications

* denotes equal contribution. Author name in **bold**.

LLM Alignment, Preference Optimization, and Reasoning

[C1] General Preference Reinforcement Learning NeurIPS 2026
M. Umer, **M. A. Mohsin***, *A. Bilal*, E. Fox*

[C2] Continuous-Utility Direct Preference Optimization EMNLP 2026
***M. A. Mohsin**, M. Umer, A. Bilal, E. Vitercik, J. M. Cioffi*

[C3] Sycophancy Disentanglement in LLMs via Reward Decomposition AAAI 2027
***M. A. Mohsin**, A. Bilal*, M. Umer, E. Fox*

[C4] Reliability via Verified Preference In Progress
***M. A. Mohsin**, M. Umer, A. Bilal, J. M. Cioffi, E. Fox*

Test-Time Scaling and Coding Agents

[C5] Teaching Agents to Code Reliably ICLR 2027
***M. A. Mohsin**, M. Kim, K. Ruan, S. Garg, V. Kumar, M. K. Ramanathan*

[C6] LoRe-KV: Lookahead Output-Perturbation with Response-Entropy Weighting for KV Cache Eviction ICLR 2027
*A. Bilal, **M. A. Mohsin**, M. Umer, D. F. Hougen*

[C7] Epistemic Uncertainty for Test-Time Discovery NeurIPS 2026
***M. A. Mohsin**, K. Riaz*, M. Umer, A. Bilal, J. M. Cioffi, E. Fox*

[C8] S^3: Stratified Scaling Search for Test-Time in Diffusion Language Models CoLM 2026
A. Bilal, **M. A. Mohsin***, M. Umer, D. F. Hougen*

[C9] What If We Allocate Test-Time Compute Adaptively? ICML 2026
A. Bilal, **M. A. Mohsin***, M. Umer, D. F. Hougen, J. M. Cioffi*

Multi-Agent Systems, Benchmarking, and Causal Discovery

[C10] MatrAIx: Simulating the World with 8.3 Billion Persona Agents Invited to Nature
*Xiaomin Li, et al., **Muhammad Ahmed Mohsin**, et al., Dawn Song*

[C11] MicroVerse: An Instrument for Measuring Self-Authored Identity Drift in Long-Horizon Multi-Agent Language-Model Simulations CoLM 2026
*Sky Ng, et al., **Muhammad Ahmed Mohsin**, et al., Yuexing Hao*

[C12] PersonaEval: Persona-Based User Simulation for Evaluating Interactive Applications CoLM 2026
*Yifan Simon Liu, et al., **Muhammad Ahmed Mohsin**, et al., Xiaomin Li*

[C13] AgentPersonaBench: Benchmarking Persona-Driven User Simulation ACL 2027
*Jintao Huang, et al., **Muhammad Ahmed Mohsin**, et al., Yuexing Hao*

[C14] When Should Semantic Priors Enter Partially-Identified Causal Discovery? AISTATS 2027
***M. A. Mohsin**, A. Bilal, S. B. Ather, Tomasso, Valentin, L. de Costa*

[C15] Automated Prompt Optimization Improves Robustness of Medical Benchmarks for LLMs ML4H 2025
*A. Aali, **M. A. Mohsin**, V. Bikia, A. Singhvi, S. Bedi, H. Cui, M. A. F. Hernandez, A. Unell, Y. Mai, M. A. Pfeffer, S. Koyejo, R. Daneshjou, E. Alsentzer, C. Potts, A. S. Chaudhari*

[C16] On the Fundamental Limits of Large Models at Scale TMLR 2025
***M. A. Mohsin**, M. Umer, A. Bilal, Z. Memon, M. I. Qadir, S. Bhattacharya, H. Rizwan, A. R. Gorle, M. Z. Kazmi, A. Mohsin, A. Subhan, M. U. Rafique, Z. He, P. Mehta, J. Han, M. A. Jamshed, D. F. Hougen, J. M. Cioffi*

Signal Processing, Wireless, and AI

[C17] Graph Neural Networks for Accelerating Low-Rank SDP Solvers JMLR 2026
***M. A. Mohsin**, M. Umer, A. Bilal, J. M. Cioffi, E. Vitercik*

[C18] Neural Gaussian Radio Fields for Channel Estimation KDD 2026
M. A. Mohsin, M. Umer*, A. Bilal, J. M. Cioffi*

[C19] Continual Learning for Wireless Channel Prediction ICML 2025
***M. A. Mohsin**, M. Umer, A. Bilal, J. M. Cioffi*

[C20] Task-Aware Distributed Source Coding for Correlated Audio Signals Using Perceptual Loss AAAI 2024
***M. A. Mohsin**, A. Bilal, S. Bhattacharya, J. M. Cioffi*

[C21] PyramidTabNet: Transformer-Based Table Recognition in Image-Based Documents ICDAR 2023
*M. Umer, **M. A. Mohsin**, A. Ul-Hasan, F. Shafait*

Invited Talks and Tutorials

Tutorial IEEE ICMLCN, Abu Dhabi, UAE.
Tutorial IEEE VTC, Boston, USA.
Panel MatrAIx AI Panel, NeurIPS.

Reviewer and Grants

Conference Reviewer AAAI, ICDAR, NeurIPS, ICML (Silver Reviewer), ICLR, KDD, CoLM
Journal Reviewer IEEE Transactions on AI, Transactions on Machine Learning Research, Applied Artificial Intelligence.
Travel Grants IEEE GLOBECOM 2024 (Cape Town, South Africa), IEEE ICC 2025 (Canada), Stanford Conference Travel Grant 2025, IEEE ICC 2026 (Glasgow, Scotland).
Best Paper/Poster Awards 6G Summit Abu Dhabi (Best Poster Nomination), IEEE ICC Canada (Best Workshop Paper Award), IEEE Communications Society Competition (honorary mention), IEEE FIT 2026 Best Main Conference Paper Award.
Technical Program Committee NeurIPS 2025, ICASSP 2026.
Workshop Co-Chair IEEE ICC 2026.
Area Chair ICASSP 2026, NeurIPS 2025.

Honors and Achievements

Fellowship Awarded Stanford Graduate Fellowship, Qualcomm Fellowship
Knight-Hennessy Fellowship Finalist.
President's Medal Third position nationwide in pre-engineering group.
Scholarships HSSC Federal Board Scholarship (PKR 200,000). STEP-ECAT Scholarship for top 10 engineering candidates nationwide (PKR 200,000).