Ahsan Bilal

Ahsan.Bilal-1@ou.edu
+1 (405) 371-8541
git/AhsanBilal7
ahsanbilal7.github.io

RESEARCH INTERESTS

Deep Learning Optimization and LLM Inference: Optimization theory for deep learning and reasoning-oriented LLMs, including post-training methods, preference optimization, alignment, and test-time compute scaling for autoregressive and diffusion language models. Analytical foundations span statistics, probabilistic modelling, and algorithms for scalable NLP and ML systems.

Agentic and Adaptive AI Systems: Design of evolving multi-agent systems with adaptive behaviors, dynamic reputation modelling, and social graph–based coordination. Focus on building scalable, distributed system architectures with strong compliance and connectivity technologies for cross-functional AI product deployment.

Reinforcement Learning for Reasoning and Decision Systems: RL methods for reasoning-oriented LLMs, process reward models (PRMs), and agentic test-time compute allocation, with applications in complex non-stationary environments, signal processing, and ML-driven wireless communication systems (including 5G).

EDUCATION

University of Oklahoma, Oklahoma, USA 2024–Present
Ph.D. in Computer Science (Deep Learning, Reinforcement Learning, & LLMs)

- CGPA: 3.86/4.0
- Advisor: Dr. Dean Hougen: Research on deep learning optimization, robust LLM reasoning, and RL-based agentic systems.

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

- CGPA: 3.66/4.0 | Specialization GPA: 4.0/4.0 | Merit Scholarship 2020
- Senior Design Thesis: Person identification using gait with fused graph and 3D-convolutional architectures (Presentation)

RESEARCH EXPERIENCE

REAL Lab, University of Oklahoma Graduate Researcher
Advisor: Dr. Dean Hougen *Aug 2024 – Present*

- Research on adaptive test-time compute for LLMs, including verifier/PRM-guided reasoning, agentic RL controllers, trajectory-level search, and dynamic compute allocation across autoregressive (ICML'26) and diffusion language models (COLM'26). Written technical specifications and rigorous statistical analysis for each system; communicated findings cross-functionally across collaborating institutions following collaborative teamwork best practices.
- Work on structured generative models (discrete diffusion), test-time scaling during denoising, and continuous learning diffusion models for wireless channel prediction under distribution shift (ICASSP'26). Maintained well-documented, modular codebases; produced written technical documentation for all research deliverables.

Stanford AI Lab (SAIL), Stanford University Research Collaborator
Supervisors: Prof. John M. Cioffi, Dr. Emily Fox | Collaborator: M. A. Mohsin Aug 2024 – Present

- Developed diffusion-based wireless estimation and neural Gaussian radio fields (nGRF), including verifier-guided trajectory search for diffusion test-time scaling and adaptive compute allocation. Applied signal processing techniques and

machine learning (ML) pipelines using pandas and scikit-learn for data preprocessing, classification tasks, and large-scale analysis.

- Worked on continuous learning under distribution shift and multimodal RAG pipelines for wireless systems. Collaborated with stakeholders from Stanford Statistics, Glasgow, and NUST in a cross-functional, collaborative research environment; publications in NeurIPS'25, ICML'25, AAAI'25, ICC'25 (Best Paper), KDD'26 (under review).

Conoid.ai (Internet of Agents Platform) Research / Engineering
Ongoing

- Building a decentralized Internet of Agents platform for self-evolving AI systems, enabling dynamic agent teams based on trust, memory, and prior performance. Designing scalable distributed system architecture and HTTP-based agent communication protocols with security built into all transactions; written system specifications covering all key use cases and acting as a self-starter driving product development across the full engineering flow.

Undergraduate Researcher Optimal ML Lab, NUST
Supervisor: Dr. Ahmad Salman *Jan 2023 – June 2024*

- Built robust CV/biometric models: boosted-attention ViT for shadow removal and fused GCN + 3D-CNN for gait recognition. Applied analytical methods with attention to detail to achieve state-of-the-art performance on benchmark datasets; supervised experimental pipelines from problem formulation through evaluation.
- Papers under review: Shadow Removal with Boosted Attention, Gait ID using Fused Graph + 3D-CNN.

INDUSTRY EXPERIENCE

Machine Learning Engineer Islamabad, Pakistan
Cowlar Design Studio (Y Combinator 21) — Based in USA *Feb 2024 – Aug 2024*

- Developed Action Recognition system for Smart Carts with 95% accuracy; built dual inference deployment for edge devices and Nvidia GPU clusters for scalable, real-time inference. Implemented clean, modular C++ and Python scripting components for high-throughput data processing pipelines, leveraging rapid prototyping to meet fast-paced development cycles.
- Automated fiber cable alignment using computer vision and machine learning with 96% success, improving precision to 5 micrometers and scaling production 40x. Communicated technical status and analytical findings clearly to cross-functional teams and key stakeholders.
- Contributed to AI-powered retail solutions deployed in collaboration with Al Meera Consumer Goods Company (Qatar), a ∼QAR 2.9B+ annual revenue company with ∼QAR 2.7B market capitalization. Delivered innovative automation addressing a core business problem; worked collaboratively with product and infrastructure teams, demonstrating strong teamwork across all phases of the project. Implemented security controls for high-volume retail transactions and used Dash-based dashboards for internal monitoring and analysis.

SELECTED PUBLICATIONS

Google Scholar

[PC1] What If We Allocate Test-Time Compute Adaptively? ICML'26
A. Bilal*, A. Mohsin*, M. Umer, A. Subhan, H. Rizwan, A. Mohsin, D.F. Hougen

[PC2] Distilling Disagreement at Test Time Reasoning NeurIPS'26
(Submitted)
A. Bilal*, M.A. Mohsin*, M. Umer, E. Fox, D.F. Hougen

* indicates equal contribution.

[PC3] General Preference Reinforcement Learning NeurIPS'26 (Submitted)
M. Umer, M.A. Mohsin, A. Bilal, E. Fox

[PC4] S^3: Stratified Scaling Search for Test-Time in Diffusion Language Models COLM'26 (Submitted)
A. Bilal*, M.A. Mohsin*, M. Umer, D.F. Hougen

[PC5] Pressure, What Pressure? Sycophancy Disentanglement in Language Models via Reward Decomposition COLM'26 (Submitted)
M.A. Mohsin*, A. Bilal*, M. Umer, E. Fox

[PC6] Continuous-Utility Direct Preference Optimization EMNLP'26 (In Progress)
M.A. Mohsin*, M. Umer*, A. Bilal, Z. He, M.U. Rafique, A. Aali, M.A. Jamshed, J.M. Cioffi, E. Fox

[PC7] ITDPDM: Information-Theoretic Discrete Poisson Diffusion Model NeurIPS'25
S. Bhattacharya, A.R. Gorle, A. Bilal, C. Ding, A.K.S. Yadav, T. Weissman

[PC8] On the Fundamental Limits of LLMs at Scale TMLR'26 (Submitted)
A. Mohsin, A. Bilal, W. Zhao, M. Umer, Researchers from DeepMind and Meta

[PC9] Neural Gaussian Radio Fields for Channel Estimation KDD'26
M. Umer*, M.A. Mohsin*, A. Bilal*, J.M. Cioffi

[PC10] Channel Prediction Under Network Distribution Shift Using Continual Learning-Based Loss Regularization ICASSP'26
M.A. Mohsin, M. Umer, A. Bilal, M.I. Qadir, M.A. Jamshed, D.F. Hougen, J.M. Cioffi

[PC11] Conditional Prior-Based Non-Stationary Channel Estimation Using Accelerated Diffusion Models ICASSP'26
M.A. Mohsin*, A. Bilal*, M. Umer, A. Ali, M.A. Jamshed, D.F. Hougen, J.M. Cioffi

[PC12] Transformer-Based Sparse CSI Estimation for Non Stationary Channels ICC'26
M.A. Mohsin, M. Umer, A. Bilal, H. Rizwan, S. Bhattacharya, M.A. Jamshed, J.M. Cioffi

[PC13] Continual Learning for Wireless Channel Prediction ICML'25
M.A. Mohsin, M. Umer, A. Bilal, M.A. Jamshed, J.M. Cioffi

[PC14] Task Aware Distributed Source Coding for Correlated Audio Signals Using Perceptual Loss AAAI'25
S. Bhattacharya, M.A. Mohsin, A. Bilal, J.M. Cioffi

[PC15] Retrieval Augmented Generation with Multi-Modal LLM Framework for Wireless Environments ICC'25
M.A. Mohsin, A. Bilal, S. Bhattacharya, J.M. Cioffi

[PC16] HDRL for Spectrum Resource Optimization in Integrated Terrestrial and Non-Terrestrial Networks AAAI'25
M.A. Mohsin, H. Rizwan, M. Umer, S. Bhattacharya, A. Bilal, J.M. Cioffi

[PC17] Abstract – LLM for Explainable AI IEEE DSAA'24
A. Bilal, B. Lin

[PC18] Meta-Thinking in LLMs via Multi-Agent Reinforcement Learning: A Survey IEEE TAI (Submitted)
A. Bilal, M.A. Mohsin, M. Umer, M.A.K. Bangash, M.A. Jamshed

[PC19] On Shadow Removal With Boosted Attention in a Vision Transformer Springer ML (Submitted)
A. Bilal, A. Salman, K. Khurshid, D.F. Hougen

[PC20] Person Identification using Gait with Fused Graph and 3D Convolutional Architectures ACM TAIS (Submitted)
A. Bilal, A. Salman, K. Khurshid

TEACHING EXPERIENCE

Teaching Assistant University of Oklahoma
CS-1313: Programming for Non-majors in C *Fall 2024 – Present*

- Supervised weekly lab sessions; designed assignments, graded with clear rubrics, and led help sessions supporting students with C programming. Collaborated with Dr. Neeman to develop supplementary written technical documentation; communicated clearly with students in a collaborative classroom environment to address individual problem areas and foster analytical thinking.

REVIEWER AND TALKS

Conference Reviewer: ICML, NeurIPS, ICASSP, PAKDD.

Journal Reviewer: TMLR, IEEE WCM, Springer MT&A, IP&M, Aquaculture Int., IJIM, IEEE Access, ICES, ISFI.

Talks: Gave a talk on "AI in Healthcare" at Norman Regional Hospital under Dr. Lubna Mirza.

HONORS AND ACHIEVEMENTS

Honorary Certificate of Appreciation: IEEE Communications Society Student Competition 2025 for "Democratizing 6G: AI-Native Wireless Digital Twin for Global Digital Equity and Sustainability."

Best Student Presentation Runner-up Award: IEEE DSAA'24 Student Forum.

Graduate Fellowship: Awarded Gallogly College of Engineering Graduate Fellowship 2025.

Best Paper Award: ICC Workshop 2025 in Montreal.

Student Travel Grant: IEEE DSAA 2024 in San Diego.

Best Adjudged Industrial Project Award: Final Year Project received 1st place at NUST Open House 2024.

UGRIP Selection: Selected as Undergraduate Research Intern for First Cohort by MBZUAI.

IEEE Recognition: Selected as an Emerging Young Researcher in the IEEE Islamabad Section.

Prime Minister Laptop Scheme: Winner of scholarship program.

CERTIFICATIONS

Google UX Professional Certificate

Deep Learning Specialization Certificate

The Advanced Communication Skills Course

SKILLS

Programming: Python, C/C++, Java, Embedded C, MATLAB. Proficient in scripting, algorithm design, and data structures.

Deep Learning & ML: PyTorch, TensorFlow (Keras), OpenCV, Docker, Mlflow, EC2 Instance. Strong machine learning background spanning neural network architectures, natural language processing (NLP), continuous learning, and scalable pipelines for large-scale data processing and dataset management. Proficient in Dash for interactive data analysis and visualization.

Statistics & Theory: Analytical and statistical learning theory, probabilistic modelling, information theory, and optimization methods applied to machine learning research; experienced in rigorous analysis and problem formulation from specification through evaluation.

Infrastructure & Systems: Distributed systems design, HTTP APIs, Docker-based infrastructure with security best practices, scalable deployment on cloud and edge. Familiar with development best practices, including version control, clean code, and written technical documentation; applied rapid prototyping to iterate quickly across diverse use cases.

Web Development: JavaScript (React.js), HTML/CSS, Next.js.

Tools: VS Code, Git, AutoCad, Figma, PyCharm, Raspberry Pi OS, NginX.

Soft Skills: Collaborative teamwork across cross-functional teams; communicate technical status and findings clearly to diverse audiences and stakeholders; supervised junior researchers and student groups; self-starter with attention to detail and innovative problem-solving in fast-paced environments.

Design: Figma, Adobe XD, Adobe Illustrator, Adobe Photoshop, Sketch, WordPress Theme Design.

Core Competencies: Signal Processing, Machine Learning (ML), Classification, Algorithms, 5G, Connectivity Technologies, Compliance, Data Analysis (pandas), Model Development (scikit-learn), Distributed AI Systems.