

Muhammad Umair Danish, PhD

London, Ontario, Canada

APPLIED AI RESEARCH | XAI | MECHANISTIC INTERPRETABILITY

mdanish3@uwo.ca

GitHub | ORCID | Google Scholar
| LinkedIn

157

CITATIONS

5

H-INDEX

5

I10-INDEX

9

JOURNALS

9

CONFERENCES

1

BEST PAPER

Google Scholar and CV snapshot, 4 August 2026.

Research Statement

Applied AI researcher with a PhD in Electrical and Computer Engineering and a publication-centered record of developing original machine-learning methodologies, rigorous experimental studies, reproducible software, and peer-reviewed scholarship across IEEE, Elsevier, Springer Nature, Frontiers, and CVPR. Expertise spans **mechanistic interpretability, explainable and trustworthy AI, time-series forecasting, physics-guided learning, generative-AI evaluation, computer vision, and human-AI interaction**. Research interests extend to transformer representation analysis, retrieval-augmented generation, parameter-efficient adaptation, faithfulness, hallucination evaluation, and responsible deployment. Experienced in directing publication-grade investigations and communicating technical evidence to interdisciplinary collaborators and research stakeholders.

Education

PhD, Electrical and Computer Engineering

Dec 2025

Western University, London, Ontario, Canada

Specialization: Applied AI and Explainable AI

Dissertation: *Neural Networks and Their Interpretability for Building Energy Modeling*

Advisor: Dr. Katarina Grolinger

Journal Publications

9

- ^{J1} M. U. Danish, K. Ali, K. Siddiqui, and K. Grolinger. "Physics-Guided Memory Network for Building Energy Modeling." *Energy & AI*, 21, 100538, 2025. doi:10.1016/j.egyai.2025.100538 3 citations
- ^{J2} M. U. Danish and K. Grolinger. "Leveraging Hypernetworks and Learnable Kernels for Consumer Energy Forecasting Across Diverse Consumer Types." *IEEE Transactions on Power Delivery*, 40(1), 75–87, 2024. doi:10.1109/TPWRD.2024.3486010 19 citations
- ^{J3} M. U. Danish and K. Grolinger. "Kolmogorov–Arnold Recurrent Network for Short-Term Load Forecasting Across Diverse Consumers." *Energy Reports*, 13, 713–727, 2025. doi:10.1016/j.egyr.2024.12.038 54 citations
- ^{J4} M. Aziz, U. Rehman, M. U. Danish, and K. Grolinger. "Global–Local Image Perceptual Score (GLIPS): Evaluating Photorealistic Quality of AI-Generated Images." *IEEE Transactions on Human–Machine Systems*, 55(2), 223–233, 2025. doi:10.1109/THMS.2025.3527397 35 citations
- ^{J5} M. U. Danish, U. Rehman, and K. Grolinger. "Monotone Delta: An Order-Theoretic Tournament Graph Approach for Internal Consistency Assessment." *Frontiers in Applied Mathematics and Statistics*, 12, 1687194, 2026. doi:10.3389/fams.2026.1687194 3 citations

6. ^{J6} M. U. Danish, M. Aziz, U. Rehman, and K. Grolinger. "Linear Lens: A Human-Centered, Non-Interventional Mechanistic Approach to Explainable AI." *Machine Learning with Applications*, 100897, 2026. doi:10.1016/j.mlwa.2026.100897
7. ^{J7} M. Buwaneswaran, M. U. Danish, and K. Grolinger. "Conditional Denoising Diffusion Model for Missing Value Imputation in Smart Meter Data." *Engineering Applications of Artificial Intelligence*, 181, 115403, 2026.
8. ^{J8} M. Aziz, U. Rehman, M. U. Danish, S. Ali, and A. Z. Abbasi. "Towards a Unified Evaluation Framework: Integrating Human Perception and Metrics for AI-Generated Images." *Multimedia Systems*, 31(3), 180, 2025. doi:10.1007/s00530-025-01769-7 21 citations
9. ^{J9} D. Kogutec and M. U. Danish. "MidiPy: A Python Application for Processing MIDI Data in Improvised Active Music Therapy." *Journal of Creative Music Systems*. Accepted.

Conference Publications

9

1. ^{C1} D. Tissera, O. Awadallah, M. U. Danish, A. Sadhu, and K. Grolinger. "Any-Class Presence Likelihood for Robust Multi-Label Classification with Abundant Negative Data." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026. CVPR record 2 citations
2. ^{C2} M. U. Danish, M. Aziz, U. Rehman, and K. Grolinger. "Cognitive Effects of Textual, Graphical, and Interactive Explanations in Explainable Artificial Intelligence." *18th IEEE International Conference on Human System Interaction (HSI)*, 2026. Accepted.
3. ^{C3} M. Aziz, M. U. Danish, and U. Rehman. "Distributional Feature Separability for Explainability of Neural Networks for Financial Credit Evaluation." *IEEE Conference on AI, Science, Engineering, and Technology (AIXSET)*, pp. 127–133, 2025. Best Paper Award.
4. ^{C4} M. U. Danish, M. Buwaneswaran, T. Fonseka, and K. Grolinger. "Graph Attention Convolutional U-Net: A Semantic Segmentation Model for Identifying Flooded Areas." *50th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, 2024. doi:10.1109/IECON55916.2024.10905482 16 citations
5. ^{C5} M. U. Danish, M. Noorchenarboo, A. Narayan, and K. Grolinger. "ChebyRegNet: An Unsupervised Deep Learning Technique for Deformable Medical Image Registration." *50th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, 2024. doi:10.1109/IECON55916.2024.10905705 2 citations
6. ^{C6} M. U. Danish, A. A. Yousaf, M. Aziz, K. Grolinger, and U. Rehman. "RootSolver: A Cognitive-Inspired Approach to Polynomial Root-Finding with Multiplicity Detection." *17th International Conference on Advanced Computational Intelligence*, 2025. IEEE Xplore
7. ^{C7} M. Aziz, M. U. Danish, F. Iqbal, and U. Rehman. "Finite-Difference Approximation for Explaining Neural Network Predictions in Financial Credit Evaluation." *International Conference on Advanced Machine Learning and Data Science (AMLDS)*, Tokyo, 2025. IEEE Xplore
8. ^{C8} A. A. Yousaf, M. Aziz, M. U. Danish, S. A. Safi, and U. Rehman. "Statistically-Guided Kernel Regression: Embedding Feature Relevance into Kernel Learning." *2nd International Conference on Advanced Machine Learning and Data Science (AMLDS)*, Osaka, 2026. Accepted for oral presentation and publication.
9. ^{C9} A. A. Yousaf, U. Rehman, and M. U. Danish. "Seeing Through Risk: A Symbolic Approximation of Prospect Theory." *22nd IEEE International Colloquium on Signal Processing & Its Applications (CSPA)*, 2026. Accepted.

Manuscripts Under Review

3

1. ^{M1} M. U. Danish, U. Rehman, and K. Grolinger. "Understanding the Inner Workings of Memory Networks with Multi-Stage Mechanistic Interpretability." In revision at *IEEE Transactions on Artificial Intelligence*.
2. ^{M2} M. U. Danish, U. Rehman, and K. Grolinger. "Behavioral Mechanistic Interpretability for Enhancing Human Cognition in Explainable AI." Under review at *IEEE Transactions on Human-Machine Systems*.
3. ^{M3} M. U. Danish, U. Rehman, and K. Grolinger. "AI-SSIM: Human-Centric Image Assessment Through Pseudo-Reference Generation." Under review at *Computer Vision and Image Understanding*.

Research Appointments

Postdoctoral Research Fellow

Jan 2026–Present

Western University, London, ON

- Lead research on mechanistic interpretability and concept-based explainability, developing faithful, non-interventional methods for understanding neural representations and decision pathways.
- Design publication-grade studies spanning behavioral XAI, human cognition, model reliability, internal representation analysis, and AI-generated-content evaluation.
- Build reproducible PyTorch research pipelines with controlled baselines, ablations, statistical validation, visualization, and decision-ready technical communication.

Graduate Research Assistant

2023–2025

Western University, London, ON

- Created neural forecasting methods for building energy and diverse consumer loads, including physics-guided memory, hypernetwork/kernel, and Kolmogorov–Arnold recurrent architectures.
- Developed end-to-end experimental workflows for preprocessing, model training, validation, cross-consumer generalization, interpretability, performance comparison, and failure analysis.
- Converted applied research into three first-author journal articles in Energy & AI, IEEE Transactions on Power Delivery, and Energy Reports.

Research Assistant, Part-Time

2023–Present

Wilfrid Laurier University, Waterloo, ON

- Co-developed **MidiPy**, an open-source Python toolkit that turns MIDI event streams into reproducible measures and visual analytics for improvised active music-therapy research.
- Contributed research software design, data-processing workflows, documentation, PyPI-oriented packaging, usability, and translation for interdisciplinary health-science collaborators.

Research Assistant

Jun 2022–Dec 2022

Gwangju Institute of Science and Technology (GIST), South Korea

- Implemented and evaluated deep-learning workflows for biomedical image restoration, deformable registration, and segmentation, emphasizing controlled preprocessing and reliable evaluation.

Honors and Awards

- **Best Paper Award**, IEEE Conference on Artificial Intelligence X Science, Engineering, and Technology (AIxSET), 2025.
- **Western Postdoctoral Fellowship**, Western University, 2026.
- **Western Graduate Fellowship**, Western University, 2024.
- **Western Graduate Research Scholarship**, Western University, 2023.
- **Korean Government Scholarship**, 2022.

Selected Research Contributions

— Linear Lens and Human-Centered Mechanistic Interpretability 2025–2026

- Developed a non-interventional framework for exposing internal neural-network behavior and translating representations into understandable, human-centered evidence without altering the deployed model.
- Extended the research direction to behavioral mechanistic interpretability, multi-stage memory-network analysis, and cognitive evaluation of textual, graphical, and interactive explanations.

— Generative-AI Perceptual Quality and Evaluation 2024–2026

- Contributed to GLIPS and a unified human-perception/metric framework for evaluating photorealistic AI-generated images; published in IEEE Transactions on Human–Machine Systems and Multimedia Systems.
- Investigated pseudo-reference generation and perceptual alignment for human-centric image assessment, connecting generative models with reliable evaluation and human-AI interaction.

Research Expertise

INTERPRETABILITY + AI ASSURANCE

Non-interventional mechanistic analysis, concept explanations, attribution diagnostics, robustness, human-centered evaluation, and responsible-AI evidence.

LLMs + GENERATIVE AI

Transformers, representation analysis, embeddings, RAG, vector retrieval, prompt and evaluation design, PEFT/LoRA concepts, faithfulness, tool use, and multimodal evaluation.

MODEL INNOVATION

Physics-guided memory networks, hypernetworks, learnable kernels, Kolmogorov–Arnold recurrent networks, diffusion, GNNs, and regularization.

EXPERIMENTAL LEADERSHIP

Study design, baselines, ablations, statistical analysis, reproducibility, peer-reviewed writing, research software, and mentoring.

Methods and Tools

- **Programming and platforms:** Python, PyTorch, TensorFlow/Keras, MATLAB, SQL/MySQL, Bash, Git/GitHub, Linux, Jupyter, Conda, $\text{\LaTeX}$; R (basic).
- **Machine learning:** deep neural networks, supervised/unsupervised learning, time-series forecasting, transformers, CNNs, RNNs, GNNs, diffusion models, representation and kernel learning, physics-guided ML, multimodal modeling.
- **LLMs and generative AI:** working knowledge of tokenization, embeddings, self-attention, prompt design, RAG, vector retrieval, PEFT/LoRA, instruction-tuning concepts, tool use, hallucination, faithfulness, safety, evaluation, and responsible AI.
- **XAI and experimentation:** mechanistic interpretability, concept-based XAI, feature attribution, counterfactual/diagnostic reasoning, model auditing, ablation studies, experiment tracking, statistical analysis, visualization, and reproducible research.

Teaching

Teaching Assistant — Western University

2023–2025

ES1036 Programming Fundamentals for Engineers; SE3309 Database Management Systems; SE2251 Software Design; SE3350 Software Engineering Design I. Guided students in programming, debugging, databases, software design, engineering problem solving, assignment interpretation, and code quality.

Lead Teaching Assistant - SE3309 — Western University

Sep–Dec 2024

Coordinated TA workflows, grading consistency, structured rubrics, solution guides, review materials, and teaching-team communication.

Scholarly Service

Peer reviewer: Engineering Applications of Artificial Intelligence; IEEE Transactions on Industrial Informatics; IEEE Transactions on Artificial Intelligence; Energy & AI; Applied Energy.

References

Available upon request.

Citation counts are dynamic and reflect the Google Scholar snapshot supplied on 4 August 2026. Publication status should be refreshed before future applications.