Tanmoy Sen

sent@uwstout.edu +1 (434) 466-2915 Website Google Scholar

Appointment

Assistant Professor, Department of Mathematics, Statistics and Computer Science, University of Wisconsin-Stout Polytechnic Fall 2026 – Present

Research Interests

Distributed and Machine Learning Systems, Large Language Model (LLM) Inference & Serving, Edge Computing, and Performance Optimization for Real-Time AI.

Education

PhD in Computer Science December 2024
University of Virginia CGPA: 3.97
Thesis: *Enhancing Performance of Machine Learning Training and Inference on the Resource-constrained Edge Devices*
Advisor: Prof. Haiying Shen

BSc in Computer Science & Engineering July 2014
Bangladesh University of Engineering & Technology (BUET) CGPA: 3.86
Thesis: *Simplified Processing of Video Data for Street Sensing Applications*
Advisor: Prof. Yusuf Sarwar

Research Experience

Information Systems Assistant (Project), **Research Foundation for SUNY, SUNY Polytechnic Institute** July 2026 – August 2026
Evaluated SUNY Poly's high-performance computing infrastructure, including GPU cluster configurations, multi-node communication, workload scheduling, and system performance. Developed an HPC improvement roadmap and user onboarding documentation.

Postdoctoral Research Associate, **University of Virginia** December 2024 – August 2025
Researched and designed systems to achieve Service Level Objectives (SLOs) for LLM inference, particularly for long-sequence processing, in collaboration with Connor Holmes and Masahiro Tanaka (Microsoft). Resulting work published at ICPP 2026 and ICCCN 2026. Also directed junior PhD students.

Research Intern, **Microsoft** June 2024 – September 2024
Implemented and evaluated a chunk-caching mechanism for the Phi-3-mini-128k-instruct language model, assessing the impact of caching different input-chunk percentages on model performance across multiple clients.

Research Assistant, **University of Virginia** August 2017 – May 2024

- *Distributed ML Training and Inference on Edge Devices.* Analyzed ML models to find optimal model partitioning and data assignment on heterogeneous edge devices. Published/accepted at ICDCS, MASS, ICCCN, and EuroSys.
- *Performance Analysis of Anycast.* Analyzed the AKAMAI DNS dataset to identify optimal anycast prefix pairs. Published at SIGCOMM 2021.
- *Detecting Memory DoS Attacks on the Cloud.* Studied cache hit/miss patterns under Memory DoS attacks and devised heuristic and ML-based detection algorithms. Published at ICPP 2020.

Research Intern, **IBM & University of Virginia** December 2020 – June 2021
Collected smart-building data for Rice Hall (UVA), analyzed it, and trained ML models to forecast temperature, shading, and lighting conditions.

Teaching Experience

Assistant Professor, **University of Wisconsin-Stout** Fall 2026 – Present
Fall 2026: CS 144 Computer Science I; CS 244 Data Structures.

Responsible for course design, lecture delivery, assessments, and student mentorship.

Visiting Assistant Professor, **Eastern Kentucky University** August 2025 – May 2026
Fall 2025: Cyber Networking and Security, Object-Oriented Programming II (in-person); Programming and Data Structures, Database Administration and Security (online).
Spring 2026: Machine Learning, Data Structures and Algorithms, Theory of Database Systems, Object-Oriented Programming I.
Responsible for curriculum design, lecture delivery, assessments, and student mentorship.

Teaching Assistant, **University of Virginia** August 2018 – May 2023
Graded assignments, delivered guest lectures, and introduced TensorFlow in graduate and undergraduate Cloud Computing courses.

Lecturer, **Bangladesh University of Engineering & Technology (BUET)** February 2015 – August 2017
Primary instructor for Computer Networks, Simulation & Modeling, Object-Oriented Programming, and Structured Language Programming. Also instructed labs for Assembly Language, Digital Logic Design, and Software Development.

Lecturer, **Ahsanullah University of Engineering & Technology (AUST)** October 2014 – February 2015
Primary instructor for Concrete Mathematics and Probability Theory; lab instructor for Assembly Language Programming (x86).

Peer-Reviewed Conference Publications

- H. Dev, M. Basak, **T. Sen**, P. Shome and B. Islam, "Hard Rules, Soft Preferences: Bridging Reasoning, Learning, and Optimization for Personalized Packing Checklist Generation". Proc. of the 35th ACM International Conference on Information and Knowledge Management (CIKM 2026), November 7 - November 11, 2026, Italy. *(To Appear)*
- H. Shen, **T. Sen** and Y. He, "Heterogeneous SLO Guaranteed Multi-Resource-Aware Batching in LLM Serving". Proc. of the 55th International Conference on Parallel Processing (ICPP 2026), September 28 - October 1, 2026, Singapore.
- H. Shen, **T. Sen** and M. Tanaka, "Managing KV Cache for Coordinated Waiting and Execution Time in LLM Serving". Proc. of the 35th International Conference on Computer Communications and Networks (ICCCN 2026), July 27 - 30, 2026, Honolulu, Hawaii, USA.
- A.Z. Sadiq, H. Shen, **T. Sen**, N. Deng, S. Xiang. "Improving Resource Overcommitment with Granular and Pattern-Based Machine Learning Predictions." *Proc. IEEE BigData*, 2025.
- A. Namazi, H. Shen, **T. Sen**, M. Zhang. "Degree-Based Scheduling and Memory Management for Large-Scale Exact Online GNN Inference." *Proc. IEEE BigData*, 2025.
- **T. Sen**, H. Shen, A. Iyer. "Flex: Fast, Accurate DNN Inference on Low-Cost Edges Using Heterogeneous Accelerator Execution." *Proc. EuroSys*, 2025.
- **T. Sen**, H. Shen. "Fault Tolerant Data and Model Parallel Deep Learning in Edge Computing Networks." *Proc. IEEE MASS*, 2024.
- **T. Sen**, H. Shen. "Elastically Scheduling Multiple DNN Inference Jobs on an Edge for High Efficiency and Accuracy." *Proc. IEEE MASS*, 2024.
- **T. Sen**, H. Shen. "Adversarial Attack Detection for Deep Learning Driving Maneuver Classifiers in Connected Autonomous Vehicles." *Proc. IEEE ICCCN*, 2024.
- **T. Sen**, H. Shen. "A Data and Model Parallelism based Distributed Deep Learning System in a Network of Edge Devices." *Proc. IEEE ICCCN*, 2023.
- Q. Liu, H. Shen, **T. Sen**, R. Ning. "Time-series Misalignment Aware DNN Adversarial Attacks for Connected Autonomous Vehicles." *Proc. IEEE MASS*, 2023.
- **T. Sen**, H. Shen. "Distributed Training for Deep Learning Models On An Edge Computing Network Using Shielded Reinforcement Learning." *Proc. IEEE ICDCS*, 2022.
- S. Liu, S. Bavikadi, **T. Sen**, H. Shen, P. Sutradhar, A. Ganguly, S. Dinakarrao, B. Smith. "Accelerating Adversarial Attack Using Process-in-Memory Architecture." *Proc. MSN*, 2022.
- Q. Liu, **T. Sen**, H. Shen, S. Dinakarrao. "A Universal Adversarial Attack on Time-Series DNNs for Connected Autonomous Vehicles." *Proc. IEEE MASS*, 2022.
- H. Shen, F. Yao, **T. Sen**, G. Moreira, A. Sarker. "Query-Efficient and Imperceptible Attacks on Multivariate Time Series DNN Models." *Proc. IFIP Networking*, 2022.

- **T. Sen**, H. Shen. "Context-aware Data Operation Strategies in Edge Systems for High Application Performance." *Proc. ICPP*, 2021. [**Best Paper Candidate**]
- **T. Sen**, H. Shen, W. Saad, T. Doan. "A Resilient and Robust Edge-Cloud Network System Supporting CPS." *Proc. IEEE MASS*, 2021.
- A. Sarker, H. Shen, **T. Sen**, Q. Mendelson. "Efficient Black-Box Adversarial Attacks for Deep Driving Maneuver Classification Models." *Proc. IEEE MASS*, 2021.
- X. Zhang, **T. Sen**, Z. Zhang, T. April, B. Chandrasekaran, D.R. Choffnes, B.M. Maggs, H. Shen, R.K. Sitaraman, X. Yang. "AnyOpt: Predicting and Optimizing IP Anycast Performance." *Proc. ACM SIGCOMM*, 2021.
- A. Sarker, H. Shen, **T. Sen**. "A Suspicion-Free Black-Box Adversarial Attack for Deep Driving Maneuver Classification Models." *Proc. IEEE ICDCS*, 2021.
- S. Shubha, **T. Sen**, H. Shen, M. Normansell. "A Diverse Noise-Resilient DNN Ensemble Model on Edge Devices for Time-Series Data." *Proc. IEEE SECON*, 2021.
- **T. Sen**, H. Shen, M. Normansell. "NoiseSenseDNN: Modeling DNN for Sensor Data to Mitigate the Effect of Noise in Edge Devices." *Proc. IEEE MASS*, 2020.
- A. Sarker, H. Shen, **T. Sen**, H. Uehara. "An Advanced Black-Box Adversarial Attack for Deep Driving Maneuver Classification Models." *Proc. IEEE MASS*, 2020.
- Z. Li, **T. Sen**, H. Shen, M.C. Chuah. "Impact of Memory DoS Attacks on Cloud Applications and Real-Time Detection Schemes." *Proc. ICPP*, 2020.
- **T. Sen**, H. Shen. "Machine Learning based Timeliness-Guaranteed and Energy-Efficient Task Assignment in Edge Computing Systems." *Proc. ICFEC*, 2019.
- H. Dev, M.E. Ali, J. Mahmud, **T. Sen**, M. Basak, R. Paul. "A Real-Time Crowd-Powered Testbed for Content Assessment of Potential Social Media Posts." *Proc. SocInfo*, 2015.
- H. Dev, M.E. Ali, **T. Sen**, M. Basak. "AntiqueData: A Proxy to Maintain Computational Transparency in Cloud." *Proc. DASFAA*, 2014.
- H. Dev, M.E. Ali, **T. Sen**, M. Basak. "An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks." *Proc. Datacloud @ SC*, 2012.

Journal Publications

- Z. Li, **T. Sen**, H. Shen, M.C. Chuah. "A Study on the Impact of Memory DoS Attacks on Cloud Applications and Exploring Real-Time Detection Schemes." *IEEE/ACM Transactions on Networking*, vol. 30, no. 4, 2022.
- J. Shen, S. Ghatti, N. Levkov, H. Shen, **T. Sen**, K. Rheuban, K. Enfield, N. Facteau, G. Engel, K. Dowdell. "A Survey of COVID-19 Detection and Prediction Approaches Using Mobile Devices, AI, and Telemedicine." *Frontiers in Artificial Intelligence*, vol. 5, 2022.

Book Chapters

- S. Liu, **T. Sen**, and H. Shen. Chapter 10: Securing Autonomous Vehicles with Machine Learning. In *6G Security: New Frontiers in Mobile Network Technologies and Verticals*, ed. Luiz A. DaSilva. John Wiley & Sons, 2025.

Arxiv Preprints

- H. Dev, M. Basak, **T. Sen**, P. Shome, and B. Islam. "Hard Rules, Soft Preferences: Bridging Reasoning, Learning, and Optimization for Personalized Packing Checklist Generation." arXiv:2607.15562, 2026.
- H. Dev, **T. Sen**, M. Basak, and B. Islam. "From Feasibility to Desirability: Plan, Learn, Adapt (PLA) Framework for Personalized On-Device Itinerary Generation." arXiv:2607.15552, 2026.
- H. Shen and **T. Sen**. "EconoServe: Maximizing Multi-Resource Utilization with SLO Guarantees in LLM Serving." arXiv:2411.06364, 2024.

Demo and Workshop Papers

- **T. Sen**, M. Basak, H. Dev, M. Linden. "Demo: FlyEnJoy: An Offline-First Mobile System for Trip Planning and In-Flight Social Interaction." *ACM MobiSys*, 2025.
- H. Dev, R. Paul, **T. Sen**, M. Basak, M.E. Ali. "Cassandra: An Oracle for Content Assessment of

Potential Social Media Posts." *Workshop on Mobile Computing and Human Computer Interaction (MoHCI)*, 2014. **[1st Runner-up, Best Paper Award]**

- S. Chakraborty, **T. Sen**, M. Basak, M.T. Sharif, A.A. Muzaddid. "An SMS based Smart System for Remote Control of Utility Services During Earthquake." *Workshop on Design and Implementation of Emerging Computing Systems (WECS)*, 2013. **[Best Paper Award]**
- T. Ahmed, **T. Sen**, M. Basak, M.Y.S. Uddin. "A Simplified Processing of Participatory Video Data for Street Traffic Estimation." *Workshop on Advances in Data Management: Applications and Algorithms (WADM)*, 2013. **[1st Runner-up, Best Paper Award]**

Selected Presentations & Talks

- Adversarial Attack Detection for Deep Learning Driving Maneuver Classifiers in Connected Autonomous Vehicles. IEEE ICCCN 2024, Hawaii, USA. (Oral Presentation)
- A Data and Model Parallelism based Distributed Deep Learning System in a Network of Edge Devices. IEEE ICCCN 2023, Hawaii, USA. (Oral Presentation)
- Distributed Training for Deep Learning Models On An Edge Computing Network Using Shielded Reinforcement Learning. IEEE ICDCS 2022. (Virtual Oral Presentation)
- Context-aware Data Operation Strategies in Edge Systems for High Application Performance. ICPP 2021. (Virtual Oral Presentation)
- A Resilient and Robust Edge-Cloud Network System Supporting CPS. IEEE MASS 2021. (Virtual Oral Presentation)
- NoiseSenseDNN: Modeling DNN for Sensor Data to Mitigate the Effect of Noise in Edge Devices. IEEE MASS 2020. (Virtual Oral Presentation)
- Machine Learning based Timeliness-Guaranteed and Energy-Efficient Task Assignment in Edge Computing Systems. ICFEC 2019, Larnaca, Cyprus. (Oral Presentation)

Mentoring Experience

PhD Students, University of Virginia

- Mahir Ashbab
- Ali Zafar Sadique
- Zakaria Mehrab — *Current: Software Engineer, Google*
- Sudipta Saha Shubha — *Current: Senior Research Scientist, Arm*
- Fazle Rabbi Masum — *Current: Applied Scientist II, Amazon*

Undergraduate Students, University of Virginia

- Matthew Normansell — *Current: Software Engineer, Amazon*
- Brian Yu — *Current: Software Engineer, Stripe*

Technical Skills

Programming: C, C++, CUDA, Java, Python, MATLAB, SQL, Assembly (x86), Shell, LaTeX
ML Frameworks: TensorFlow, PyTorch, Keras, SciPy, NumPy
LLM/Systems: vLLM, FasterTransformer, MapReduce, Spark
IoT: Raspberry Pi, SparkFunEdge, Samsung SmartWatch
Databases: Oracle, MySQL, MongoDB
Quantum/XAI: Qiskit, PennyLane, Cirq; SHAP, Captum (PyTorch)
Web: HTML, CSS, PHP, JSP, JavaScript, Ajax, jQuery

Honors and Awards

- Computer Science Fellowship, University of Virginia — August 2017 – July 2018
- Dean's List, Bangladesh University of Engineering & Technology (BUET)
- University Merit Scholarship, BUET
- Distinguished Poster Award, 2nd Undergraduate Thesis Poster Presentation, CSE, BUET
- ITEE-FE Certification (2014) — Level 2 Information Technology Engineers Examination (Passing Rate: 15.8%)

- Finalist, Samsung R&D Contest, 2013

Professional Services

Reviewer

- IEEE Transactions on Mobile Computing
- IEEE/ACM Transactions on Networking
- Elsevier Journal on Engineering Applications of Artificial Intelligence
- Elsevier Journal on Knowledge-Based Systems
- Elsevier Journal on Physical Communication
- Elsevier Journal on Reliability Engineering & System Safety
- Journal of Systems Architecture

External Reviewer

- IEEE INFOCOM (2019–2024), IEEE BigData (2019–2022), IEEE SECON (2019–2022), IEEE MASS (2020–2022), IEEE MobiHoc (2023–2025), ACM SIGCOMM (2022), IEEE/ACM Transactions on Networking, IEEE Transactions on Green Communications and Networking

Workshops & Professional Development

- Quantum Machine Learning for Science and Engineering Workshop (2025) — modules on quantum circuits, variational algorithms, and hybrid classical–quantum workflows.

Organizing Committee: International Conference on Networking Systems and Security (NSysS), 2015 and 2016.

Treasurer: Association of Bangladeshi Students, University of Virginia.