Walid Abdelrahman Hanafy

(+1) 413 345-9453 • whanafy@cs.umass.edu • /whanafy • whanafy

Research Interests

AI Systems, Energy Efficiency, Failure Resilience, Cloud/Edge Computing, and IoT Systems.

Academic Appointments

2025–Now **University of Massachusetts Amherst**, *Manning College of Information and Computer Sciences.*
Postdoctoral Research Associate
Advisor: Prashant Shenoy

Education

2025 **MS'23/PhD'25, Computer Science**, *University of Massachusetts Amherst*, MA, USA.
Thesis: Carbon-aware Resource Management for Cloud Computing Platforms.
Advisor: Prashant Shenoy, LASS research lab.

2018 **Master of Computer Engineering**, *Faculty of Engineering*, Helwan University, Egypt.
Thesis: Container Type Virtualization Management in Cloud Computing.
Advisors: Sameh Salem, Amr Mohamed.

2012 **Diploma in Software Engineering**, *Minor in Software Architecture*, Information Technology Institute (ITI), Egypt.

2011 **BS in Computer Engineering (Class Valedictorian)**, Helwan University, Egypt.
Thesis: Driver Behavior Monitoring using computer vision and ultrasonic sensors.
Advisor: Sameh Salem

Honors and Awards

2026 **Best Paper Award** for SwapLess in *IEEE DCOSS-IoT 2026*.
2025 **UMass Spotlight**: How One Computer Scientist Is Rethinking Sustainable Computing at Scale ().
2025 **Krithi Ramamritham Computer Science Scholarship** for Outstanding Graduate Students.
2024 **Best Student Paper Award** for CarbonScaler in *ACM SIGMETRICS/IFIP Performance 2024*.
2021 **Outstanding Innovation Award** for my summer internship at *Nokia Bell Labs*, USA.
2012 **ICT4Change – 1st place** for my ITI graduation project from *ICT Ministry*, Egypt.
2011 **Young Innovators Award** for my BS graduation project from *Nahdet El Mahrosa*, Egypt.

Research Experience

2025–Now **Postdoctoral Research Associate**, *University of Massachusetts Amherst*, MA, USA.
Projects AI-based resource management for cloud and IoT systems.
Failure-resilient AI systems for resource-constrained edge environments.
Foundation-model-as-a-service (FMaaS) cloud platform.
Power and energy management for data centers and green AI systems.

2020–2025 **Research Assistant**, *University of Massachusetts Amherst*, MA, USA.
Projects Sustainable computing systems.
Model-serving platforms for resource-constrained edge environments.

2021, 2022 **Summer Intern**, *Nokia Bell Labs, Network Systems Research Department*, NJ, USA.

Projects Ultra-low-latency model-serving framework using RDMA and GPUDirect RDMA.
Telemetry system for RDMA network traffic.

2018 **Visiting Scholar**, *Public University of Navarra, Spain*, Funded by ERASMUS+.

Project Cleanliness detection of solar panels using computer vision and machine learning.

2015–2019 **Research Assistant**, *Helwan University*, Egypt.

Project Elasticity resource management for container-as-a-service (CaaS) clouds.

Publications and Patents

Selected Publications:

[29] Li Wu, **Walid A. Hanafy**, Tarek Abdelzaher, David Irwin, Jesse Milzman, and Prashant Shenoy. FailLite: Failure-Resilient Model Serving for Resource-Constrained Edge Environments. In *Proceedings of the 2025 ACM Symposium on Cloud Computing*, SoCC'25, 2025.

[28] Hetvi Shastri, **Walid A. Hanafy**, Li Wu, David Irwin, Mani Srivastava, and Prashant Shenoy. LLM-Driven Auto Configuration for Transient IoT Device Collaboration. In *Proceedings of the Tenth ACM/IEEE Symposium on Edge Computing (SEC)*, 2025.

[27] **Walid A. Hanafy**, Qianlin Liang, Noman Bashir, Abel Souza, David Irwin, and Prashant Shenoy. Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*, ASPLOS '24, page 479–496, 2024.

[26] **Walid A. Hanafy**, Qianlin Liang, Noman Bashir, David Irwin, and Prashant Shenoy. CarbonScaler: Leveraging Cloud Workload Elasticity for Optimizing Carbon-Efficiency. *SIGMETRICS Performance Evaluation Review*, volume 52, page 49–50, June 2024. **Best Student Paper Award**.

Conference and Journal Articles:

[25] Kang Yang, **Walid A. Hanafy**, Prashant Shenoy, and Mani Srivastava. FM-CAC: Carbon-Aware Control for Battery-Buffered Edge AI via Time-Series Foundation Models. In *2026 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, 2026.

[24] **Walid A. Hanafy**, Thanathorn Sukprasert, Abel Souza, David Irwin, and Prashant Shenoy. Untangling the Carbon-Cost Tradeoffs and Stampede Effect Challenges in Cloud Computing. *IEEE Transactions on Computers*, pages 1–14, 2026.

[23] Nathan Ng, **Walid A. Hanafy**, et al. Collaborative Processing for Multi-Tenant Inference on Memory-Constrained Edge TPUs. In *2026 22nd International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, 2026. **Best Paper Award**.

[22] Jorge Murillo, **Walid A. Hanafy**, David Irwin, Ramesh K. Sitaraman, and Prashant Shenoy. Go with the Flow: Analyzing the Carbon Footprint of Green Streaming. In *The 17th ACM International Conference on Future and Sustainable Energy Systems (e-Energy '26), June 22–25, 2026, Banff, Canada*, 2026.

[21] Fabian Francis Mkocheko, **Walid A. Hanafy**, Adam Lechowicz, David Irwin, and Prashant Shenoy. City-Scale Decarbonization of Residential Heating under Transformer Constraints. In *Proceedings of the 13th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys '26, 2026.

[20] Ahmed S. Abouhend, **Walid A. Hanafy**, Christopher Watt, Raphael C. Nnachi, Prashant Shenoy, and Caitlyn S. Butler. Making waves: Toward systems-level interpretation of hormonal and endogenous biomarkers in wastewater-based epidemiology. *Water Research*, page 126269, 2026.

[19] Li Wu, **Walid A. Hanafy**, Abel Souza, Khai Nguyen, Jan Harkes, David Irwin, Mahadev Satyanarayanan, and Prashant Shenoy. CarbonEdge: Leveraging Mesoscale Spatial Carbon-Intensity Variations for Low Carbon Edge Computing. In *Proceedings of the 34th International Symposium on High-Performance Parallel and Distributed Computing (HPDC '25)*, 2025.

[18] Diptyaroop Maji, **Walid A. Hanafy**, Li Wu, David Irwin, Prashant Shenoy, and Ramesh K. Sitaraman. Data Centers Carbon Emissions at Crossroads: An Empirical Study. *ACM SIGEnergy Energy Informatics Review*, volume 5, page 48–55, 2025. Presented in HotCarbon'25.

[17] David Irwin, Prashant Shenoy, Mohammad Hajiesmaili, **Walid A. Hanafy**, Jimi Oke, Ramesh Sitaraman, Yuvraj Agarwal, Geoff Gordon, Zico Kolter, Deepak Rajagopal, Mani Srivastava, Vivienne Sze, Priya Donti, Andrew Chien, John Birge, Ali Hortacsu, and Line Roald. A Vision for Computational Decarbonization of Societal Infrastructure. *IEEE Internet Computing*, pages 1–7, 2025.

[16] Phuthipong Bovornkeeratiroj, **Walid A. Hanafy**, David Irwin, and Prashant Shenoy. GreenThrift: Optimizing Carbon and Cost for Flexible Residential Loads. *ACM Journal on Computing and Sustainable Societies*, March 2025.

[15] **Walid A. Hanafy**, Roozbeh Bostandoost, Noman Bashir, David Irwin, Mohammad Hajiesmaili, and Prashant Shenoy. The War of the Efficiencies: Understanding the Tension between Carbon and Energy Optimization. *ACM SIGEnergy Energy Informatics Review*, volume 4, page 87–93, September 2024. Presented in HotCarbon'23.

[14] Jorge Murillo, **Walid A. Hanafy**, David Irwin, Ramesh Sitaraman, and Prashant Shenoy. CDN-Shifter: Leveraging Spatial Workload Shifting to Decarbonize Content Delivery Networks. In *Proceedings of the 2024 ACM Symposium on Cloud Computing*, SoCC '24, page 505–521, 2024.

[13] Jinyang Li, Yizhuo Chen, Tomoyoshi Kimura, Tianshi Wang, Ruijie Wang, Denizhan Kara, Yigong Hu, Li Wu, **Walid A. Hanafy**, Abel Souza, Prashant Shenoy, Maggie Wigness, Joydeep Bhattacharyya, Jae Kim, Guijun Wang, Greg Kimberly, Josh Eckhardt, Denis Osipychev, and Tarek Abdelzaher. Acies-OS: A Content-Centric Platform for Edge AI Twinning and Orchestration. In *2024 33rd International Conference on Computer Communications and Networks (ICCCN)*, pages 1–9, 2024.

[12] Hyunseok Chang, **Walid A. Hanafy**, Sarit Mukherjee, and Limin Wang. INSERT: In-Network Stateful End-to-End RDMA Telemetry. In *IEEE Conference on Computer Communications*, INFOCOM'24, 2024.

[11] Roozbeh Bostandoost, **Walid A. Hanafy**, Adam Lechowicz, Noman Bashir, Prashant Shenoy, and Mohammad Hajiesmaili. Data-driven algorithm selection for carbon-aware scheduling. *ACM SIGEnergy Energy Informatics Review*, volume 4, page 148–153, April 2024. Presented in HotCarbon'24.

[10] Roozbeh Bostandoost, Adam Lechowicz, **Walid A. Hanafy**, Noman Bashir, Prashant Shenoy, and Mohammad Hajiesmaili. LACS: Learning-Augmented Algorithms for Carbon-Aware Resource Scaling with Uncertain Demand. In *The 15th ACM International Conference on Future and Sustainable Energy Systems (e-Energy '24), June 4–7, 2024, Singapore*, 2024.

[9] **Walid A. Hanafy**, Limin Wang, Hyunseok Chang, Sarit Mukherjee, T.V. Lakshman, and Prashant Shenoy. Understanding the Benefits of Hardware-Accelerated Communication in Model-Serving Applications. In *2023 IEEE/ACM 31st International Symposium on Quality of Service (IWQoS)*, IWQoS'23, pages 1–10, 2023.

[8] Abel Souza, Noman Bashir, Jorge Murillo, **Walid A. Hanafy**, Qianlin Liang, David Irwin, and Prashant Shenoy. Ecovisor: A Virtual Energy System for Carbon-Efficient Applications. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*, volume 2 of *ASPLOS 2023*, page 252–265, 2023.

[7] Qianlin Liang, **Walid A. Hanafy**, Noman Bashir, David Irwin, and Prashant Shenoy. Energy Time Fairness: Balancing Fair Allocation of Energy and Time for GPU Workloads. In *Proceedings of the 8th ACM/IEEE Symposium on Edge Computing (SEC)*, 12 2023.

[6] Qianlin Liang, **Walid A. Hanafy**, Noman Bashir, Ahmed Ali-Eldin, David Irwin, and Prashant Shenoy. Dělen: Enabling Flexible and Adaptive Model-serving for Multi-tenant Edge AI. In *Proceedings of IEEE/ACM Eighth International Conference on Internet-of-Things Design and Implementation (IoTDI), San Antonio*, 5 2023.

[5] Qianlin Liang, **Walid A. Hanafy**, Ahmed Ali-Eldin, and Prashant Shenoy. Model-driven Cluster Resource Management for AI Workloads in Edge Clouds. *ACM Transactions on Autonomous and Adaptive Systems*, volume 18, pages 1–26, 2023.

[4] **Walid A. Hanafy**, Alfredo Pina, and Sameh A. Salem. Machine Learning Approach for Photovoltaic Panels Cleanliness Detection. In *2019 15th International Computer Engineering Conference (ICENCO)*, pages 72–77, 2019.

[3] **Walid A. Hanafy**, Amr E. Mohamed, and Sameh A. Salem. A New Infrastructure Elasticity Control Algorithm for Containerized Cloud. *IEEE Access*, volume 7, pages 39731–39741, 2019.

[2] **Walid A. Hanafy**, Amr E. Mohamed, and Sameh A. Salem. Novel Selection Policies for Container-based Cloud Deployment Models. In *2017 13th International Computer Engineering Conference (ICENCO)*, pages 237–242, 2017.

[1] **Walid A. Hanafy**, Amr E. Mohamed, and Sameh A. Salem. A Load Balancing with Power Optimization Algorithm for Container-based Infrastructure Management. In *2017 12th International Conference on Computer Engineering and Systems (ICCES)*, pages 161–166, 2017.

Patents:

[1] Hyunseok Chang, Limin Wang, Sarit Mukherjee, and **Walid Abdelrahman**. End-to-End RDMA Telemetry System, 2023. Patent US11876691B2.

Workshops and Posters:

[8] Roozbeh Bostandoost, Adam Lechowicz, **Walid A. Hanafy**, Prashant Shenoy, and Mohammad Hajiesmaili. Quantifying the carbon reduction of dag workloads: A job shop scheduling perspective. In *Proceedings of the 2nd International Workshop on Systems and Methods for Sustainable Large-Scale AI (GreenSys)*, GreenSys '26, pages 1–6, 2026.

[7] Hetvi Shastri, **Walid A. Hanafy**, Li Wu, David Irwin, Mani Srivastava, and Prashant Shenoy. Poster Abstract: Rethinking Collaboration Among Mobile Devices in IoT Environments. In *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems*, pages 676–677, 2025.

[6] Hetvi Shastri, Pragya Sharma, **Walid A. Hanafy**, Mani Srivastava, and Prashant Shenoy. FMTK: A Modular Toolkit for Composable Time Series Foundation Model Pipelines. In *The NeurIPS workshop on Recent Advances on Time Series Foundation Models: Have We Reached the BERT Moment? (BERT2S)*, 2025.

[5] Krishna Praneet Gudipaty, **Walid A. Hanafy**, Li Wu, Jeffrey Twigg, Benjamin M. Marlin, Jesse Milzman, Suhas Diggavi, Tarek Abdelzaher, and Prashant Shenoy. Practical Considerations for Failure Resilient ML Systems at the Edge. In *MILCOM 2025 - 2025 IEEE Military Communications Conference (MILCOM)*, pages 1–6, 2025.

[4] Li Wu, **Walid A. Hanafy**, Abel Souza, Tarek Abdelzaher, Gunjan Verma, and Prashant Shenoy. Enhancing Resilience in Distributed ML Inference Pipelines for Edge Computing. In *MILCOM 2024 - 2024 IEEE Military Communications Conference (MILCOM)*, pages 1–6, 2024.

[3] Mihir Shenoy and **Walid A. Hanafy**. Home Decarbonizer: Greening Household Energy Consumption Using Temporal Shifting. In *2024 IEEE MIT Undergraduate Research Technology Conference (URTC)*, pages 1–5, 2024.

[2] **Walid A. Hanafy**, Li Wu, Tarek Abdelzaher, Suhas Diggavi, and Prashant Shenoy. Failure-Resilient ML Inference at the Edge through Graceful Service Degradation. In *Proceedings of the 41st IEEE Military Communications Conference (MILCOM) workshop on Internet of Things for Adversarial Environments*, 10 2023.

[1] **Walid A. Hanafy**, Tergel Molom-Ochir, and Rohan Shenoy. Design Considerations for Energy-Efficient Inference on Edge Devices. In *Proceedings of International Workshop on Energy-Efficient Learning at the Edge*, WEEE'21, page 302–308, 2021.

Conference Presentations & Talks

June 2026 **Sustainability Week 2026**, Banff, Alberta, Canada.
City-Scale Decarbonization of Residential Heating under Transformer Constraints

June 2026 **DCOSS-IoT 2026**, Reykjavik, Iceland.
Collaborative Processing for Multi-Tenant Inference on Memory-Constrained Edge TPUs

June 2026 **19th Cloud Control Workshop**, Stockholm, Sweden.
Lightweight Resilience for Edge AI Systems

Oct. 2025 **INFORMS Annual Meeting 2025**, Atlanta, GA, USA.
Carbon-aware Scheduling: A Framework for Integrating Intermittent Renewable Energy into Grid Operations

July 2025 **HotCarbon'25**, Cambridge, MA, USA.
Data Centers Carbon Emissions at Crossroads: An Empirical Study

Mar. 2025 **Guest Lecturer**, UMass COMPSCI 677: Distributed and Operating Systems.
A Hitchhiker's Guide to Sustainable Computing.

Feb. 2025 **NESysDay**, Northeastern University, Boston, MA, USA.
Carbon-aware Scheduling: Approaches and Trade-offs.

Dec. 2024 **Guest Lecturer**, UMass COMPSCI 590X: Decarbonization and Data Science.
Introduction to Carbon-aware Computing.

July 2024 **HotCarbon'24**, Santa Cruz, CA, USA.
Data-driven Algorithm Selection for Carbon-Aware Scheduling.

June 2024 **SIGMETRICS'24**, Venice, Italy.
CarbonScaler: Leveraging Cloud Workload Elasticity for Optimizing Carbon-Efficiency.

May 2024 **ASPLOS'24**, San Diego, CA, USA.
Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions.

July 2023 **MILCOM'23**, Boston, MA, USA.
Failure-Resilient ML Inference at the Edge through Graceful Service Degradation.

July 2023 **HotCarbon'23**, Boston, MA, USA.
The War of the Efficiencies: Understanding the Tension between Carbon and Energy Optimization.

June 2023 **IWQoS'23**, Orlando, FL, USA.
Understanding the Benefits of Hardware-Accelerated Communication in Model-Serving.

June 2021 **e-Energy WEEE'21**, Online.
Design Considerations for Energy-Efficient Inference on Edge Devices.

Dec. 2019 **ICENCO 2019**, Cairo, Egypt.
Machine Learning Approach for Photovoltaic Panels Cleanliness Detection.

Dec. 2017 **ICENCO 2017**, Cairo, Egypt.
Novel Selection Policies for Container-based Cloud Deployment Models.

Dec. 2017 **ICCES 2017**, Cairo, Egypt.
A Load Balancing with Power Optimization Algorithm for Container-based Infrastructure.

Teaching Experience

2020–2023 **University of Massachusetts Amherst**, *Teaching Assistant*.
Spring 2023 COMPSCI 677: Distributed and Operating Systems
Spring 2020 COMPSCI 577: OS Design and Implementation

2015–2019 **Information Technology Institute (ITI)**, *Lecturer*.

Designed and taught courses in Software Design, Cloud Computing, and Software Architecture

Technologies: Microsoft Stack, Microsoft Azure, and Docker *multiple recurrent courses*

2015–2019 **Helwan University**, *Teaching Assistant.*

Software Engineering, Database Systems, Operating Systems, and Distributed Systems

Mentoring Experience

University of Massachusetts, Amherst, USA

2025–2026	Early Research Scholars Program	*A Closer Look at the World's Energy Supply*
2024–Now	Hetvi Shastri, PhD Student	*Foundation Models and IoT*
2024–Now	Krishna Praneet Gudipaty, PhD Student	*Failure Resilient AI*
2024–Now	Fabian Francis Mkocheko, PhD Student	*LLM Serving*
2024–Now	Thanathorn Sukprasert, PhD Student	*Grid-responsive Data Centers*
Spring 2024	Prachiti Parkar, MSc Student (Now at Microsoft)	*Sustainable Computing*
2021–2022	Tergel Molom-Ochir, Undergraduate Student (Now PhD Student at Duke)	*Edge AI*

Helwan University, Egypt

2021–Now	Undergraduate Students – Graduation and Course Projects	*Cloud/Web Systems*

Other Work Experience

2017–2019 **Research Engineer**, *Testicide (start-up)*, Egypt.

Led the design and implementation of a "codeless" GUI testing framework.

Tools: Azure Service Fabric, CosmosDB, AngularJS

2016–2017 **Software Engineer**, *IT-Bits*, UAE.

Led the successful development of software solutions for multiple pharmaceutical companies.

Tools: ASP.NET Core, MSSQL server, Angular 2

2012–2013 **Software Engineer**, *Consukorra-PES*, Egypt.

Designed and implemented a GIS-based data collection platform, leading to a 200% speedup in the collection process.

Tools: QGIS, PostgreSQL, .NET Framework

Service and Volunteer Work

2021–Now **Journal Reviewer**.

IEEE Access, Springer Automated Software Engineering, ACM Computing Surveys, IEEE Internet Computing, IEEE Transactions on Computers, IEEE Transactions on Consumer Electronics

2026 **Program Committee Member**, HotCarbon'26, Seattle, WA, USA.

2026 **Organization Committee Member**, CarbonMetrics 2026, Ann Arbor, Michigan, USA.

2026 **Program Committee Member**, IEEE IC2E 2026, Santa Clara, CA, USA.

2026 **Program Committee Member**, IEEE/ACM SEC 2026, Santa Clara, CA, USA.

2026 **Program Committee Member**, ACM CAIS 2026, San Jose, CA, USA.

2026 **Program Committee Member**, ACM SIGCOMM 2026, Denver, CO, USA.

2026 **Program Committee Member**, ACM Symposium on Cloud Computing (SoCC) 2026, Singapore.

2026 **Program Committee Member**, GreenSys'26, Edinburgh, Scotland.

2026 **Program Committee Member**, ACM e-Energy 2026, Banff, Canada.

2025 **Program Committee Member**, IEEE/ACM SEC 2025, Washington, D.C., USA.

2025 **Program Committee Member**, HotCarbon'25, Cambridge, MA, USA.

2025 **Organization Committee Member**, CarbonMetrics 2025, Stony Brook, NY, USA.

2024–2025 **Shadow PC Member**, EuroSys'25, EuroSys'26.

2024–2025 **UMass CICS Turing Program**, *Summer Camp for High School Students*, Instructor.

2020–2024 **UMass MSA**, *Graduates Branch*, Treasurer.

2021 **Artifact Evaluation Committee Member**, *EuroSys 2021.*

2017 **Docker Egypt Meetup**, *Tutorials on containers*, Funded by Microsoft and Docker, Egypt.

Technical Skills

Programming	Python, C/C++, C#, Java, Go
Web	HTML 5, .NET Stack, AngularJS
Database	MSSQL, PostgreSQL, MySQL, Cosmos DB
Cloud	VMware, AWS, Microsoft Azure, Docker, and Kubernetes
Testing	xUnit, SpecFlow, and Selenium
ML/AI	PyTorch, TensorRT, NVIDIA Triton, ONNX, vLLM

Referees

Prashant Shenoy, *University of Massachusetts Amherst* ✉ shenoy@cs.umass.edu

David Irwin, *University of Massachusetts Amherst* ✉ deirwin@umass.edu

Ramesh K. Sitaraman, *University of Massachusetts Amherst* ✉ ramesh@cs.umass.edu

Mani Srivastava, *University of California, Los Angeles* ✉ mbs@ucla.edu

Mohammad Hajiesmaili, *University of Massachusetts Amherst* ✉ hajiesmaili@cs.umass.edu

Ahmed Ali-Eldin Hassan, *Chalmers University of Technology* ✉ ahmed.hassan@chalmers.se

Updated On: June 26, 2026