Mhd Saria Allahham

13433336614 | @ saria.allahham@proton.me | Website | Google Scholar

Education

Queen's University — Kingston, ON, Canada
Ph.D. in Computer Science; ***GPA: 4.23/4.30*** — *May 2022 – May 2026 (Expected)*

Research Topic: Assessing Computational Reliability in Extreme Edge Computing for Distributed Learning and Inference

Queen's University — Kingston, ON, Canada
M.Sc. in Computer Science; ***GPA: 4.30/4.30*** — *Jan 2021 – Apr 2022*

Thesis: Multi-Orchestrator Mobile Edge Learning: Designing Energy-Efficient Task Allocation and Incentive Schemes

Qatar University — Doha, Qatar
B.Sc. in Computer Engineering; ***GPA: 3.90/4.00*** — *Jan 2016 – Apr 2020*
Graduated with High Order of Excellence

Senior Project: Designing a Smart Home Controller for Smart Home Devices using Hand Gestures

Job Experience

Alexa Translations — Montreal, QC, Canada
AI R&D Lead — *May 2025 – Present*

- Designed Retrieval-Augmented Translation system, a state-of-the-art translation engine powered by in-house professional translation data.
- Driving AI innovation by monitoring and evaluating cutting-edge academic and industry advances.
- Communicate and collaborate with the Product Team to implement and ship new and innovative AI features in the company platform.
- Collaborate with the Linguistic Operations Team to evaluate the developed systems and harness the data needed for development.
- Lead design of scalable LLM training (PyTorch, Unsloth) and inference pipelines (vLLM, SGLang).
- Mentor and review work of junior engineers; establish best-practice coding, reproducibility, and MLOps standards.

Alexa Translations — Montreal, QC, Canada
AI Engineer — *June 2024 – Apr 2025*

- Research and implement state-of-the-art LLM techniques including continued pre-training, instruction fine-tuning, preference alignment, and LLM deployment.
- Contribute to technological innovations by staying current to the cutting-edge achievements of GenAI and LLM from industry and academia.
- Documented code, experiments, and results, contributing to internal knowledge-sharing initiatives.

Samsung Research America, AI Center — Montreal, QC, Canada
AI/ML Research Engineer — *Oct 2022 – Feb 2024*

- Developing, deploying, and testing AI algorithms on real hardware and simulations.
- Leveraging the deep research work and findings to develop and program integrated software algorithms to solve real-world problems.
- Translating mathematical and algorithmic problem specifications into efficient deployable code.
- Developing and proposing new project ideas.
- Writing scientific papers for publication and patents.
- Engaging with Samsung business units to develop new ideas that can have business impact.

Queen's University, Telecommunications Research Lab — Kingston, ON, Canada
Graduate Research Fellow - Part Time — *Jan 2021 – Sep 2022*

- Modeling and analyzing Federated Learning at the network edge for resource-limited smart devices.
- Developing energy-efficient protocols for Federated Learning.
- Writing and reviewing research articles.

Qatar's University, Smart IoT Systems Lab Doha, Qatar

Research Assistant - Full Time *May 2020 – Sep 2021*

- Designing and implementing smart protocols using AI for Ultra Reliable Low Latency Communication (URLLC) in smart health systems.
- Reviewing and employing state-of-the-art smart algorithms for protocols.
- Writing and reviewing research articles.

Projects

Local 3GPP Chat

Self-developed

- **Summary:** A chatbot powered by Retrieval Augmented Generation (RAG) and a local LLM that gives information about ETSI and 3GPP standards.
- **Programming Languages:** Python
- **Relevant Software & Libraries:** PyTorch, Ollama, llama index, Transformers
- **Open-source version:** https://github.com/saria-lh/3GPP-RAG-chat

5G Mobile Wireless Networks Simulator with AI-based Load Balancing Algorithms

Samsung Research America

- **Summary:** A proprietary practical simulation software designed to simulate 5G Networks and AI load balancing algorithms.
- **Programming Languages:** Java, Python, MATLAB
- **Relevant Software & Libraries:** PyTorch, CVX/CVXPY, Stable Baselines3
- **Open-source version:** https://github.com/saria-lh/MERLIN

AI-based Indoor Localization and Human State Estimation using Ultra-Wideband Protocol

Samsung Research America

- **Summary:** A proprietary framework that enables the estimation of location, activities, and the number of people in an indoor environment without requiring them to carry specific devices.
- **Programming Languages:** C/C++, Python, Java
- **Relevant Software & Libraries:** PyTorch, OpenCV, ROS, Docker

Skills & Expertise

Software Development

- **Programming Languages:** Python, MATLAB, C/C++, Java, and Bash scripting.
- **Machine Learning:** SciPy, Pandas, NumPy, SciKit-Learn, Pandas, XGBoost, Matplotlib and Seaborn.
- **Deep Learning:** PyTorch, Tensorflow/Keras, and OpenCV.
- **Generative AI**: SGLang, vLLM, FAISS, HuggingFace, Unsloth, Axolotl, OpenAI/Gemini APIs.
- **MLOps**: LightningAI, Weights & Biases, Docker, AWS Cloud, OpenSearch and Azure Cloud.
- Nvidia Sionna, Blender

Artificial Intelligence

- Data Science and Machine Learning.
- Deep Learning.
- Computer Vision.
- Natural Language Processing (NLP).
- Large Language Models (LLMs).
- Reinforcement Learning.
- Multi-Agent Systems.

Telecommunications and Computer Networks

- Digital Signal Processing.
- Edge networks.
- Internet of Things (IoT).
- TCP/IP network stack.
- Wireless and Cellular Networks.
- Reading 3GPP standards and implementing functionalities.
- Digital twins, simulation and experimental analysis for telecommunication systems.

List of Publications

[1] Alaa Awad Abdellatif, **Allahham, Mhd Saria**, Noor Khial, Amr Mohamed, Aiman Erbad, and Khaled Shaban. Reliable federated learning for age sensitive mobile edge computing systems. In *ICC 2023-IEEE International Conference on Communications*, pages 1622–1627. IEEE, 2023.

[2] Alaa Awad Abdellatif, **Allahham, Mhd Saria**, Amr Mohamed, Aiman Erbad, and Mohsen Guizani. Onsra: An optimal network selection and resource allocation framework in multi-rat systems. In *ICC 2021-IEEE International Conference on Communications*, pages 1–6. IEEE, 2021.

[3] Sherif B. Azmy, **Allahham, Mhd Saria**, Nizar Zorba, and Hossam S. Hassanein. Quantifying the impact of incentives on service availability at the extreme edge. In *GLOBECOM 2024 - 2024 IEEE Global Communications Conference*, pages 4162–4167, 2024.

[4] Emna Baccour, **Allahham, Mhd Saria**, Aiman Erbad, Amr Mohamed, Ahmed Refaey Hussein, and Mounir Hamdi. Zero touch realization of pervasive artificial intelligence as a service in 6g networks. *IEEE Communications Magazine*, 61(2):110–116, 2023.

[5] Yahuza Bello, Alaa Awad Abdellatif, **Allahham, Mhd Saria**, Ahmed Refaey Hussein, Aiman Erbad, Amr Mohamed, and Mohsen Guizani. B5g: Predictive container auto-scaling for cellular evolved packet core. *IEEE Access*, 9:158204–158214, 2021.

[6] Heba DM Dawoud, **Allahham, Mhd Saria**, Alaa Awad Abdellatif, Amr Mohamed, Aiman Erbad, and Mohsen Guizani. Patient-driven network selection in multi-rat health systems using deep reinforcement learning. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2021.

[7] Naram Mhaisen, **Allahham, Mhd Saria**, Amr Mohamed, Aiman Erbad, and Mohsen Guizani. On designing smart agents for service provisioning in blockchain-powered systems. *IEEE Transactions on Network Science and Engineering*, 9(2):401–415, 2021.

[8] Hassan Saadat, **Allahham, Mhd Saria**, Alaa Awad Abdellatif, Aiman Erbad, and Amr Mohamed. Rl-assisted energy-aware user-edge association for iot-based hierarchical federated learning. In *2022 International Wireless Communications and Mobile Computing (IWCMC)*, pages 548–553. IEEE, 2022.

[9] **Al Laham, Saria**, Bobak H Baghi, Pierre-Yves Lajoie, Amal Feriani, Sachini Herath, Steve Liu, and Gregory Dudek. Device-free human state estimation using uwb multi-static radios. *arXiv preprint arXiv:2401.05410*, 2023.

[10] **Allahham, Mhd Saria**, Alaa Awad Abdellatif, Naram Mhaisen, Amr Mohamed, Aiman Erbad, and Mohsen Guizani. Multi-agent reinforcement learning for network selection and resource allocation in heterogeneous multi-rat networks. *IEEE Transactions on Cognitive Communications and Networking*, 8(2):1287–1300, 2022.

[11] **Allahham, Mhd Saria**, Alaa Awad Abdellatif, Amr Mohamed, Aiman Erbad, Elias Yaacoub, and Mohsen Guizani. I-see: Intelligent, secure, and energy-efficient techniques for medical data transmission using deep reinforcement learning. *IEEE Internet of Things Journal*, 8(8):6454–6468, 2020.

[12] **Allahham, Mhd Saria**, Mohammad F Al-Sa'd, Abdulla Al-Ali, Amr Mohamed, Tamer Khattab, and Aiman Erbad. Dronerf dataset: A dataset of drones for rf-based detection, classification and identification. *Data in brief*, 26:104313, 2019.

[13] **Allahham, Mhd Saria**, Salimur Choudhury, and Hossam S. Hassanein. Reliable federated learning with auction-based incentives at the extreme edge. In *GLOBECOM 2024 - 2024 IEEE Global Communications Conference*, pages 3134–3139, 2024.

[14] **Allahham, Mhd Saria**, Tamer Khattab, and Amr Mohamed. Deep learning for rf-based drone detection and identification: A multi-channel 1-d convolutional neural networks approach. In *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*, pages 112–117. IEEE, 2020.

[15] **Allahham, Mhd Saria**, Amr Mohamed, Aiman Erbad, and Mohsen Guizani. Motivating learners in multiorchestrator mobile edge learning: A stackelberg game approach. *IEEE Canadian Journal of Electrical and Computer Engineering*, 46(1):69–76, 2022.

[16] **Allahham, Mhd Saria**, Amr Mohamed, Aiman Erbad, and Hossam Hassanein. On the modeling of reliability in extreme edge computing systems. In *2022 5th International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, pages 1–6. IEEE, 2022.

[17] **Allahham, Mhd Saria**, Amr Mohamed, and Hossam Hassanein. Incentive-based resource allocation for mobile edge learning. In *2022 IEEE 47th Conference on Local Computer Networks (LCN)*, pages 157–164. IEEE, 2022.

[18] **Allahham, Mhd Saria**, Sameh Sorour, Amr Mohamed, Aiman Erbad, and Mohsen Guizani. Energy-efficient multi-orchestrator mobile edge learning. *arXiv preprint arXiv:2109.00757*, 2021.

[19] **Lahham, Saria Al**, Di Wu, Ekram Hossain, Xue Liu, and Gregory Dudek. Probabilistic mobility load balancing for multi-band 5g and beyond networks. In *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1673–1678, 2024.