Hassan Ali

PhD Candidate, UNSW, Sydney

hassan.ali@unsw.edu.au | Google Scholar | Sydney, Australia
0000-0002-1701-0390 | hassanalikhatim | hassanalikhatim.github.io

How do I see myself?

I am a self-motivated machine learning engineer and researcher. As a researcher, I strive to enable real-world deployment of machine learning models that people can trust. As an engineer, My goal is to use my skills to assist people in their daily routine tasks.

Education

Sep 2023 – Ongoing
University of New South Wales (UNSW), Sydney, Australia
PhD in Computer Science and Engineering (Full-time)
- Research focuses on Trustworthy Machine Learning

Sep 2017 – Aug 2019
National University of Sciences and Technology (NUST), Islamabad, Pakistan
Master of Science in Electrical Engineering **(CGPA: 4.0/4.0)**
- Thesis Title: "Analyzing the Security Vulnerabilities of Deep Neural Networks: Attacks and Defenses"

Sep 2013 – Aug 2017
University of Engineering and Technology (UET), Lahore, Pakistan
Bachelor of Science in Electrical Engineering **(CGPA: 3.645/4.0)**

Work Experience

Feb 2024 - Present
University of New South Wales (UNSW)
Casual Research Assistant (Dr. Arash Shaghaghi)
- Large Language Models

Sep 2021 - Sep 2023
Information Technology University (ITU)
Research Assistant (Dr. Junaid Qadir)
- Human-centric Robust ML-driven IoT Smart Services

Jan 2021 - Nov 2021
Information Technology University (ITU)
Research Assistant (Dr. Junaid Qadir)
- Mitigating Anti-social Behavior through Beneficial AI

Tools and skillset

- Python, PyTorch, TensorFlow (last 5 years)
- Java, C, MATLAB, Verilog, VHDL, HTML

Publications

2025

1. Nofal, A. B., **Ali, H.**, Hadi, M., Ahmad, A., Qayyum, A., Johri, A., Al-Fuqaha, A. & Qadir, J. AI-enhanced interview simulation in the metaverse: Transforming professional skills training through VR and generative conversational AI. *Computers and Education: Artificial Intelligence* **8,** 100347. doi:10.1016/j.caeai.2024.100347 (2025).

2024

2. **Ali, H.**, Javed, R. T., Qayyum, A., AlGhadhban, A., Alazmi, M., Alzamil, A., Al-utaibi, K. & Qadir, J. Robust Encrypted Inference in Deep Learning: A Pathway to Secure Misinformation Detection. *IEEE Transactions on Dependable and Secure Computing.* doi:10.1109/tdsc.2024.3447629 (2024).
3. **Ali, H.**, Nepal, S., Kanhere, S. S. & Jha, S. Adversarially Guided Stateful Defense Against Backdoor Attacks in Federated Deep Learning. *ACSAC 2024 (Accepted for publication)*. https://arxiv.org/abs/2410.11205 (2024).
4. Butt, M. A., **Ali, H.**, Qayyum, A., Sultani, W., Al-Fuqaha, A. & Qadir, J. R^2S100K: Road-Region Segmentation Dataset for Semi-supervised Autonomous Driving in the Wild. *International Journal of Computer Vision,* 1–19. doi:10.1007/s11263-024-02207-3 (2024).
5. Al-Maliki, S., Qayyum, A., **Ali, H.**, Abdallah, M., Qadir, J., Hoang, D. T., Niyato, D. & Al-Fuqaha, A. Adversarial Machine Learning for Social Good: Reframing the Adversary as an Ally. *IEEE Transactions on Artificial Intelligence.* doi:10.1109/TAI.2024.3383407 (2024).
6. Qayyum, A., Butt, M. A., **Ali, H.**, Usman, M., Halabi, O., Al-Fuqaha, A., Abbasi, Q. H., Imran, M. A. & Qadir, J. Secure and trustworthy artificial intelligence-extended reality (AI-XR) for metaverses. *ACM Computing Surveys* **56,** 1–38. doi:10.1145/3614426 (2024).

2023

7. **Ali, H.**, Butt, M. A., Filali, F., Al-Fuqaha, A. & Qadir, J. Consistent Valid Physically-Realizable Adversarial Attack Against Crowd-Flow Prediction Models. *IEEE Transactions on Intelligent Transportation Systems,* 1–16. doi:10.1109/TITS.2023.3343971 (2023).
8. **Ali, H.**, Khan, M. S., AlGhadhban, A., Alazmi, M., Alzamil, A., Al-Utaibi, K. & Qadir, J. Con-detect: Detecting adversarially perturbed natural language inputs to deep classifiers through holistic analysis. *Computers & Security* **132,** 103367. doi:10.1016/j.cose.2023.103367 (2023).
9. Butt, M. A., Qayyum, A., **Ali, H.**, Al-Fuqaha, A. & Qadir, J. Towards secure private and trustworthy human-centric embedded machine learning: An emotion-aware facial recognition case study. *Computers & Security* **125,** 103058. doi:10.1016/j.cose.2022.103058 (2023).

2022

10. **Ali, H.**, Khan, M. S., Al-Fuqaha, A. & Qadir, J. Tamp-X: Attacking explainable natural language classifiers through tampered activations. *Computers & Security* **120,** 102791. doi:10.1016/j.cose.2022.102791 (2022).

2021

11. **Ali, H.**, Khan, M. S., AlGhadhban, A., Alazmi, M., Alzamil, A., Al-Utaibi, K. & Qadir, J. All your fake detector are belong to us: evaluating adversarial robustness of fake-news detectors under black-box settings. *IEEE Access* **9,** 81678–81692. doi:10.1109/ACCESS.2021.3085875 (2021).
12. Petrick, N., Akbar, S., Cha, K. H., Nofech-Mozes, S., Sahiner, B., Gavrielides, M. A., Kalpathy-Cramer, J., Drukker, K., Martel, A. L. & BreastPathQ Challenge Group, f. t. SPIE-AAPM-NCI BreastPathQ Challenge: an image analysis challenge for quantitative tumor cellularity assessment in breast cancer histology images following neoadjuvant treatment. *Journal of Medical Imaging* **8,** 034501–034501. doi:10.1117/1.jmi.8.3.034501 (2021).

2020

13. Khalid, F., **Ali, H.**, Hanif, M. A., Rehman, S., Ahmed, R. & Shafique, M. *FaDec: A Fast Decision-based Attack for Adversarial Machine Learning* in *2020 International Joint Conference on Neural Networks (IJCNN)* (2020), 1–8. doi:10.1109/ijcnn48605.2020.9207635.

2019

14. **Ali, H.**, Khalid, F., Tariq, H. A., Hanif, M. A., Ahmed, R. & Rehman, S. SSCNets: Robustifying DNNs using Secure Selective Convolutional Filters. *IEEE Design & Test* **37,** 58–65. doi:10.1109/mdat.2019.2961325 (2019).
15. Khalid, F., **Ali, H.**, Tariq, H., Hanif, M. A., Rehman, S., Ahmed, R. & Shafique, M. *QuSecNets: Quantization-based defense mechanism for securing deep neural network against adversarial attacks* in *2019 IEEE 25th International Symposium on On-Line Testing and Robust System Design (IOLTS)* (2019), 182–187. doi:10.1109/iolts.2019.8854377.