# Momin Ahmad Khan

momin-ahmad-khan.github.io

Email : makhan@umass.edu
LinkedIn: LinkedIn portfolio


## Education

**Doctor of Philosophy, Electrical and Computer Engineering** Sep. 2021 – Dec. 2025 (Expected)
*University of Massachusetts Amherst, USA* *CGPA 3.95/4.00*

- PhD student supervised by: Fatima Muhammad Anwar
- Research areas: Mechanistic Interpretability, Federated Learning, Adversarial Machine Learning
- Major courses: Security and Privacy for GenAI, Advanced Information Assurance, Probability and Random Processes

**Bachelor of Electrical Engineering** Sep. 2017 – Jun. 2021
*School of Electrical Engineering and Computer Science (SEECS)* *CGPA 3.97/4.00*
*National University of Sciences & Technology (NUST), Islamabad, Pakistan*

- Undergraduate student supervised by Dr. Hassan Aqeel Khan and Dr. Faisal Shafait
- Final year thesis: "Low-Cost Whole Slide Image Scanner with Deep Learning Applications" (Awarded Rector's Gold medal)
- Major courses: Machine Learning, Computer Vision, Signal Processing, Embedded Systems, Microprocessor Systems

## Research Interests

- Mechanistic interpretability of multimodal foundation models via sparse autoencoders and latent-space steering
- Security and robustness in federated learning under data heterogeneity and adversarial manipulation
- Systematic analysis of FL defenses, uncovering evaluation pitfalls and proposing remedies for reliable benchmarking

## Publications

Controlling Vision–Language–Action Policies through Sparse Latent Directions (Paper) *MechInterp @ NeurIPS 2025*
***Momin Ahmad Khan**, Novak Boskov, Fatima Muhammad Anwar, Manzoor Ahmad Khan*

HYDRA-FL: Hybrid Knowledge Distillation for Robust and Accurate Federated Learning (Paper) (Code) *NeurIPS 2024*
***Momin Ahmad Khan**, Yasra Chandio, Fatima Muhammad Anwar*

SABRE-FL: Selective and Accurate Backdoor Rejection for Federated Prompt Learning (Paper) *Under Review @ ICLR 2026*
***Momin Ahmad Khan**, Yasra Chandio, Fatima Muhammad Anwar*

Decoding FL Defenses: Systemization, Pitfalls, and Remedies (Paper) (Code) *Under Review @ ACM TOPS*
***Momin Ahmad Khan**, Virat Shejwalkar, Yasra Chandio, Amir Houmansadr, Fatima Muhammad Anwar*

Compromising Federated Medical AI-Backdoor Risks in Prompt Learning (Paper) *Poster Abstract @ SenSys 2025*
***Momin Ahmad Khan**, Yasra Chandio, Eugene Bagdasarian, Fatima Muhammad Anwar*

A Neurosymbolic Approach to Adaptive Feature Extraction in SLAM (Paper) *IROS 2024*
*Yasra Chandio, **Momin Ahmad Khan**, Khotso Selialia, Luis Garcia, Joseph DeGol, and Fatima M. Anwar*

On the Pitfalls of Security Evaluation of Robust Federated Learning (Paper) (Code) *DLSP @ IEEE S&P 2023*
***Momin Ahmad Khan**, Virat Shejwalkar, Amir Houmansadr, Fatima Muhammad Anwar*

Security Analysis of SplitFed Learning (Paper) (Code) *AIChallengeIoT @ SenSys 2022*
***Momin Ahmad Khan**, Virat Shejwalkar, Amir Houmansadr, Fatima Muhammad Anwar*

Universal Timestamping with Ambient Sensing (Paper) *SECON 2022*
*Adeel Nasrullah, **Momin Ahmad Khan**, Fatima Muhammad Anwar*

## Experience

**Research Scientist Intern** Jun. 2025 – Sep. 2025
*Nokia Bell Labs, Murray Hill, NJ, USA*
*Research topics:* Mechanistic Interpretability, Multimodal Foundational Models, Representation Engineering in LLMs/VLMs

- Development of interpretable control pipeline for embodied AI agent (Microsoft MAGMA).
- Exploring Sparse Autoencoders (SAEs) for mechanistic interpretability in multimodal LLMs.
- Proposed inference-time framework for steering multimodal LLMs in the action-space

**Research Scientist Intern** Jun. 2024 – Aug. 2024
*Nokia Bell Labs, Murray Hill, NJ, USA*
*Research topics:* AI agents, image-generation models, speech-to-text and text-to-speech models

- Developed end-to-end autonomous model selection pipeline for in-house Bell Labs infrastructure.
- Used LLama 3-based LLM agents for designing data curation pipelines.
- Prototyped MeetingMate, a smart multimodal meeting assistant for Bell Labs.

- Tools used: OpenAI Whisper, Meta's ImageBind

**Research Assistant and PhD Candidate** Sep. 2021 – current
*Emerging Embedded Technologies Lab, UMass Amherst, USA*
*Research topics:* Federated Learning, ML Security and Privacy, Interpretable AI

- Designing secure and robust federated learning systems under data heterogeneity and adversarial threats
- Exploring mechanistic interpretability in multimodal foundation models using sparse representations and latent space steering
- Analyzing and improving evaluation practices in FL defenses to enable reliable benchmarking and reproducible research
- Teaching Assistant for ECE-635 and ECE-231

**Undergraduate Research Assistant** May 2019 – May 2021
*Signal Processing and Machine Learning (SIGMA) Lab in collaboration with TUKL*
*R&D Center, SEECS, NUST, Islamabad, Pakistan*

*Main research topics:* Medical Imaging, Signal processing, AI on Edge devices, Embedded systems

- Optimization of AI models (YOLOv4, ResNets) with TensorRT on edge devices, including Jetson Nano and Jetson TX2
- Real-time auto-focusing with Laplacian filter as focus metric. Integrated algorithm with stepper motors circuitry
- Segmentation of keratin pearl and epithelium tissues using UNET in Whole Slide images
- Teaching assistant for Digital Logic Design course (EE-221)

**Embedded Systems Engineer** Aug. 2020 – Dec 2020
*Sedenius Technologies (subsidiary of Sedenius Engineering GmbH) and NUST, Pakistan*

- Explored Embedded Systems research aspects of Self Driving Cars
- Proposed Embedded and Electronic circuit design for "Mobile Data-logger for Test-bed Scale Car"
- Developed schematics and PCB design for "RC Autonomous Car" in EAGLE

**AI Teaching Assistant, part time** Jun. 2020 – May 2021
*AI Lounge, Islamabad, Pakistan*

- Created tailored course content on deep learning and computer vision for school children.
- Simplified complex topics using intuitive methods and engaging visuals

## Technical Skills

**Languages**: Python, C/C++, MATLAB, Assembly and embedded C
**Programming**: PyTorch, Tensorflow, JAX, MXNet, fast.ai, Keras, Latex, TensorRT, Scikit-learn, Hugging-Face, LangChain
**Deep Learning**: SFT, DPO, GRPO, LLMs, VLMs, VLAs, Classification, Object Detection
**Tools**: Linux, Slurm, PyCharm, VS code, Git, Google Cloud Platform, AUTOCAD

## Honors and Awards

- Chancellor's Silver Medal for second-highest CGPA in undergraduate degree, NUST
- Rector's Gold Medal for best Undergraduate Final Year Project, NUST
- Recipient of NUST scholarship for outstanding entrance examination score
- NUST High Achievers Award and merit scholarship recipient in 8/8 semesters
- 1st Position in PIEAS Entrance Examination all over Pakistan, 2017
- 1st Position in GIKI Entrance Examination all over Pakistan, 2017

## Academic Services

- Reviewer for NeurIPS 2024/2024, ICLR 2026, SenSys Posters 2025, FLCA@AAAI 2026
- Member of organizing team for AI Security Seminar at UMass Amherst
- Part of the organizing team for "AI at the Edge" workshop organized by AI Lounge, Pakistan
- Mentor and project evaluator for SIGMA Lab interns at NUST
- Participated in NeurIPS 2024, SenSys 2022, and IEEE S&P 2023 in-person

## Extracurricular Activities

- Runner-up in Dota 2 at NUST E-Sports Gala (2020 & 2021)
- Led door-to-door clothes collection drives for Akhuwat Clothes Bank and distributed them to underserved communities
- Organized donation drive for Master Ayub's School, an open-air school for underprivileged students

## Interests and Hobbies

Multiplayer computer games (Top 10% globally in Dota 2), photography, cooking, traveling, playing guitar, and table tennis.