Asif Khan

mdasifkhan.github.io
Harvard University

Email: asif.khan@hms.harvard.edu
GitHub: github.com/MdAsifKhan
Mobile: (617) 555-1234

Professional Summary

AI researcher specializing in deep generative modeling and representation learning with applications in biomedicine. My current research involves building deep learning models to predict the risk of multiple cancers using large-scale electronic health record (EHR) datasets. This includes training foundation models on EHR data to learn patient embeddings that accurately capture individual health states and risk profiles while minimizing demographic biases and ensuring interpretability, providing transparent predictions at both the patient and population levels. Additionally, I implement transfer learning and domain-adaptive pretraining to enhance model generalization across diverse healthcare systems. In parallel, I develop computational models that predict cellular responses to drug treatments or CRISPR interventions. This research deepens our understanding of cancer resistance mechanisms and informs the design of more effective combination therapies.

Education

- **The University of Edinburgh** — Edinburgh, UK
  Ph.D. in Machine Learning — *Oct. 2019 – Oct. 2023*
- **University of Bonn** — Bonn, Germany
  MSc. in Computer Science, Grade: 1.1 ("ausgezeichnet" - "excellent") — *Oct. 2017 – Sep. 2019*

Research Experience

- **Harvard Medical School** — Boston, USA
  Postdoctoral Fellow in AI & Healthcare (Advisor: Prof. Chris Sander) — *Jan. 2024 – Present*
  - Leading development of **interpretable deep learning models** for cancer risk prediction from large-scale EHR databases.
  - Designing large language models for domain-adaptive training on EHR data.
  - Developing interpretability metrics to pinpoint key features driving cancer risk predictions at both individual and population levels, and analyzing patient embedding spaces to provide clinicians with comparative insights.
  - Collaborating with clinicians to ensure model predictions are robust, well-calibrated, and reliable for clinical applications.

- **The University of Edinburgh** — Edinburgh, UK
  Ph.D. in Machine Learning (Advisor: Prof. Amos Storkey) — *Oct. 2019 – Oct. 2023*
  - Developed **robustness** techniques for deep generative models to reduce adversarial vulnerabilities.
  - Designed Hamiltonian-based frameworks for modeling complex systems with conservation guarantees.
  - Created novel geometric representation learning methods for graphs, enhancing modeling of heterophilous interactions.

- **Huawei Noah's Ark Lab** — London, UK
  Research Scientist Intern — *Sept. 2021 – Dec. 2021*
  - Led development of a combinatorial Bayesian optimization framework for antibody design, incorporating uncertainty quantification.
  - Validated the framework on multiple antigens, culminating in a publication in *Cell Reports Methods*.
  - Developed modular code to generalize the method across diverse antigens.

- **Sony** — Stuttgart, Germany
  Research Intern — *Mar. 2019 – Aug. 2019*
  - Developed a generative adversarial network (GAN) for unsupervised speech-to-speech conversion with a focus on model robustness.
  - Implemented approaches to improve model stability and reduce performance variance across different speech domains.

- **Bio-Ontology Research Group, KAUST** — Jeddah, Saudi Arabia
  Research Assistant — *Jan. 2016 – May 2017*
  - Developed ontology-aware neural networks for predicting Gene Ontology functions.
  - Implemented an interpretable model that integrates symbolic reasoning with deep learning to derive meaningful representations of biological entities, including drugs and genes.

Technical Skills

- **Machine Learning:** Deep generative models, representation learning, self-supervised learning, geometric deep learning, large language models
- **Healthcare AI:** Large-scale EHR data analysis, single-cell perturbation modeling, drug discovery
- **Programming:** Python (PyTorch, NumPy, Scikit-learn), SQL, Bash, Git, Hugging Face, GCP

Publications

[1] **A. Khan***, D. Ritter*, C. Zheng*, D. Marks, N. V. Do, N. R. Fillmore, C. Sander. *Pancreatic cancer risk prediction using deep sequential modeling of longitudinal diagnostic and medication records.* **Cell Reports Medicine**, 2025. [*Joint First Author]

[2] **A. Khan***, D. Ritter*, C. Zheng, D. T. Forster, M. Harsh, A. Gazizov, D. S. Marks, N. R. Fillmore, C. Sander. *Multi-cancer risk prediction using transformers trained on large-scale longitudinal EHR data.* NeurIPS Workshop on Learning from Time Series for Health, 2025. [*Joint First Author]

[3] **A. Khan**, D. T. Forster, M. Harsh, D. Ritter, C. Zheng, R. Orenbuch, A. Kuzeiz, D. S. Marks, N. R. Fillmore, C. Sander. *AI-driven risk stratification is essential for affordable early detection of cancer.* NeurIPS Workshop on GenAI for Health: Potential, Trust, and Policy Compliance, 2025.

[4] L. M. Luque, **A. Khan**, G. Torrisi, T. D. Green, D. Hardman, C. Owczarek, T. A. Phillips, D. S. Marks, M. Parsons, C. Sander, L. J. Schumacher. *Identifying tissue states by spatial protein patterns related to chemotherapy response in triple-negative breast cancer.* bioRxiv, 2025.

[5] Y. Sun, Z. Liu, R. Sun, L. Qian, S. H. Payne, W. Bittremieux, M. Ralser, C. Li, Y. Chen, Z. Dong, Y. Perez-Riverol, **A. Khan**, C. Sander, R. Aebersold, J. A. Vizcaíno, J. R. Krieger, J. Yao, H. Wen, L. Zhang, Y. Zhu, Y. Xuan, B. B. Sun, L. Qiao, H. Hermjakob, H. Tang, H. Gao, Y. Deng, Q. Zhong, C. Chang, N. Bandeira, M. Li, S. Sun, Y. Yang, G. S. Omenn, Y. Zhang, P. Xu, Y. Fu, X. Liu, C. M. Overall, Y. Wang, E. W. Deutsch, L. Chen, J. Cox, V. Demichev, F. He, J. Huang, H. Jin, C. Liu, N. Li, Z. Luan, J. Song, K. Yu, W. Wan, T. Wang, K. Zhang, L. Zhang, P. A. Bell, M. Mann, B. Zhang, T. Guo. *Strategic priorities for transformative progress in advancing biology with proteomics and artificial intelligence.* **Nature Methods**, 2025 (in press).

[6] **A. Khan***, G. Torrisi*, L. Luque*, C. Owczarek, M. Parsons, C. Sander, L. Schumacher. *Batch-effect invariant GNNs for predicting chemotherapy response in TNBC.* ICML ML4LMS Workshop, 2024. [*Joint First Author]

[7] **A. Khan**, A. Storkey. *Contrastive learning for non-local graphs with multi-resolution structural views.* arXiv:2308.10077, 2023.

[8] **A. Khan**, A. Storkey. *Adversarial robustness of VAEs through the lens of local geometry.* Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS), 2023.

[9] **A. Khan***, A. I. Cowen-Rivers*, A. Grosnit, D. Goh, P. A. Robert, V. Greiff, et al. *Toward real-world automated antibody design with combinatorial Bayesian optimisation.* **Cell Reports Methods**, 2023. [*Joint First Author]

[10] A. I. Cowen-Rivers, P. J. Gorinski, A. Sootla, **A. Khan**, et al. *Structured Q-learning for antibody design.* NeurIPS RL4RealLife Workshop, 2022.

[11] **A. Khan**, A. Storkey. *Hamiltonian latent operators for content and motion disentanglement in image sequences.* Advances in Neural Information Processing Systems (NeurIPS), 2022.

[12] **M. A. Khan**, F. Cardinaux, S. Uhlich, M. Ferras, A. Fischer. *Unsupervised cross-domain speech-to-speech conversion with time-frequency consistency.* arXiv:2005.07810, 2020.

[13] **M. A. Khan***, A. Kukleva*, H. Farazi, S. Behnke. *Utilizing temporal information in deep CNNs for efficient soccer ball detection and tracking.* RoboCup Symposium (Springer), 2019. [*Joint First Author]

[14] **M. A. Khan***, A. Kristiadi*, D. Lukovnikov, J. Lehmann, A. Fischer. *LiteralE: Incorporating literals into knowledge graph embeddings.* ISWC (Springer), 2019. [*Joint First Author]

[15] A. Agarwal, R. Gangopadhyay, S. Dubey, S. Debnath, **M. A. Khan**. *Learning-based predictive DSA framework for enhanced QoE.* IET Communications, 2018.

[16] M. Kulmanov, **M. A. Khan**, R. Hoehndorf. *DeepGO: Predicting protein functions from sequence and interactions using a deep ontology-aware classifier.* Bioinformatics, 2017.

[17] M. Alshahrani, **M. A. Khan**, O. Maddouri, A. R. Kinjo, N. Q. Rosinach, R. Hoehndorf. *Neuro-symbolic representation learning on biological knowledge graphs.* Bioinformatics, 2017.

[18] M. Tanveer, **M. A. Khan**, S.-S. Ho. *Robust energy-based least squares twin SVMs.* Applied Intelligence, 2016.

[19] A. Agarwal, S. Dubey, **M. A. Khan**, R. Gangopadhyay, S. Debnath. *Learning-based primary user activity prediction in cognitive radio.* SPCOM (IEEE), 2016.

Funding, Fellowships, & Professional Activities

- **Funding & Fellowships**
  - Fund for Innovation in Cancer Informatics Fellowship 2025–2027
  - DAAD Postdoc–Net–AI Fellowship 2024
  - NeurIPS Scholar Award 2022
- **Teaching**
  - **University of Edinburgh, UK** Oct. 2019–Oct. 2023
    - Tutor for Probabilistic Modeling & Reasoning
    - Marker for Probabilistic Modeling & Reasoning, Machine Learning Practical, Introductory Applied ML, and Data Mining & Exploration; evaluated coursework, exams, and projects
  - **University of Bonn, Germany** Oct. 2017–Feb. 2019
    - Teaching Assistant for Knowledge Graph Analysis: conducted tutorials and graded exams
    - Prepared theoretical and programming exercises: https://github.com/SmartDataAnalytics/Knowledge-Graph-Analysis-Programming-Exercises
- **Peer Reviewing**
  - *Conference & Journals:* NeurIPS 2025, ICML 2025 (Outstanding Reviewer), NeurIPS 2024 (Top Reviewer), ICML 2024 (Best Reviewer), NeurIPS 2023 (Top Reviewer), ICML 2023, AISTATS 2022–2023 (Top Reviewer), NeurIPS 2022 (Top Reviewer), ICLR 2022 (Highlighted Reviewer).
  - *Workshops:* ML4LMS Workshop, ICML (2024); SynS & ML Workshop, ICML (2023); ML4PS Workshop, NeurIPS (2021–2023)

Student Supervision

- Sophie Sharum, Undergraduate, Harvard University, 2025
- Abdullah Kuzeiz, Masters, Harvard-MIT Health Science and Technology, 2025
- Michael Dennison, Undergraduate, MIT, 2024
- Nicole Xu, Undergraduate, MIT, 2024