Sarah **Alnegheimish**

32 Vassar Street, 32-D712, MIT LIDS, Cambridge, MA 02139
smish@mit.edu | sarahmish.github.io | sarahalnegheimish | sarahmish

Research Interests: Machine Learning Frameworks, Deep Learning, Time Series, Natural Language Processing.

Education

2022 - 2025 **Massachusetts Institute of Technology** | GPA: 5.0/5.0
Ph.D. in Electrical Engineering and Computer Science (EECS)
Supervised by Dr. Kalyan Veeramachaneni

2019 - 2022 **Massachusetts Institute of Technology** | GPA: 5.0/5.0
S.M. in Electrical Engineering and Computer Science (EECS)
S.M. in Computational Science and Engineering (CSE)
Supervised by Dr. Kalyan Veeramachaneni

2013 - 2017 **King Saud University** | GPA: 5.0/5.0
B.Sc. in Computer Science
First Class Honors

Experience

Research Assistant | DAI Lab, MIT *Cambridge, MA*

Machine Learning Frameworks for Time Series Tasks (PhD Thesis) *Sep 2022 - present*

- Leading *Sintel* ecosystem for time series modeling and analysis.
- Designed and implemented universal representations of *primitives* and *pipelines* for time series tasks including: classification, forecasting, and anomaly detection.
- Conducted real-world case studies featuring two predictive maintenance studies on wind turbine data and two studies on anomaly detection for satellite data and electric vehicle data.
- Collaborated in the development of *Zephyr*, a data-centric framework for predictive maintenance of wind turbine energy data, with generalized abstractions to define various prediction problems.

Unsupervised Time Series Anomaly Detection (Master Thesis) *Sep 2019 - May 2022*

- Developed *Orion*, an open-source library for unsupervised time series anomaly detection (github stars: 842).
- Created a thorough benchmark suite for continuous quality and computational evaluations.
- Designed metrics for time-based anomaly scoring that prioritize the anomalous class.
- Collaborated with students on a number of anomaly detection pipelines including: an autoencoder with regression model (AER) and a generative adversarial model (TadGAN) for reconstruction-based anomaly detection.

Gender Bias in Large Language Models *Sep 2021 - Feb 2022*

- Created a new dataset to examine model components that elicit biased output.
- Proposed mitigation strategies to reduce the discrimination severity of the model output.

Intuitive Physics for Game Play Bots *Sep 2019 - May 2020*

- Built a probabilistic program to infer actions performed by an agent in Newtonian settings.
- Complemented the program with a model-free approach for efficient sampling.

Software Engineer Intern | DataCebo *Boston, MA*

Modeling Generative Time Series *Jun 2021 - Aug 2021*

- Worked on improving the modularization of the PARModel which is a generative model for synthetic time series.
- Worked closely with the SDV team on reversable data transforms and conditional sampling.

Research Specialist | Center for Complex Systems at KACST and MIT *Riyadh, Saudi Arabia*

Predictive Framework for Healthcare Data *Dec 2017 - Aug 2019*

Collaborated with Dr. Kalyan Veeramachaneni

- Developed *Cardea*, an automated machine learning library to solve health related prediction problems.
- Integrated HL7's *Fast Healthcare Interoperability Resources* standard as a representation for data.

Occupation-Skill Space *Dec 2017 - Aug 2019*

Collaborated with Professor Iyad Rahwan

- Constructed an occupation network by discovering job-job relationships through their underlying skills.
- Utilized the job space as a function to predict individual career trajectories.
- Analyzed the impact of automation on job mobility and the education hierarchy.

Data Analyst | Mozn *Riyadh, Saudi Arabia*
Fraud Detection and Optical Character Recognition *May 2017 - Dec 2017*

- Worked closely with a team of experts to develop a system that detects fraudulent shipments.
- Curated a dataset of Arabic characters and trained a convolutional model for automated character recognition.

Publications

- **Alnegheimish, S.**, Liu, D., Sala, C., Berti-Equille, L. and Veeramachaneni, K., *Sintel: A Machine Learning Framework to Extract Insights from Signals.* ACM SIGMOD International Conference on Management of Data, 2022.
- **Alnegheimish, S.***, Guo, A.*, Sun, Y.*, *Using Natural Sentence Prompts for Understanding Biases in Language Models.* Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022.
- Liu, D., **Alnegheimish, S.**, Zytek, A. and Veeramachaneni, K., *MTV: Visual Analytics for Detecting, Investigating, and Annotating Anomalies in Multivariate Time Series.* ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW), 2022.
- Alhasoun, F. and **Alnegheimish, S.**, *Probabilistic Programming Bots in Intuitive Physics Game Play. 35th AAAI Conference on Artificial Intelligence*, 2021.
- Geiger, A., Liu, D., **Alnegheimish, S.**, Cuesta-Infante, A., Veeramachaneni, K., *TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks.* IEEE Conference on Big Data, 2020. (citations: 187).
- **Alnegheimish, S.**, Alrashed, N., Aleissa, F., Althobaiti, S., Liu, D., Alsaleh, M. and Veeramachaneni, K., *Cardea: An Open Automated Machine Learning Framework for Electronic Health Records.* IEEE Conference on Data Science and Advanced Analytics (DSAA), 2020.

Honors & Awards

- MIT's EECS Graduate Alumni Fellowship, 2022-2023.
- Graduate Scholarship, King Abdulaziz City for Science and Technology (KACST), 2019 - 2022.
- Google's CS Research Mentorship Program (CSRMP), 2020 - 2021.
- Graduate Fellow, MiSK, 2019 - 2022.
- 2nd place at AEC annual best graduation project, 2017.
- Best graduation project in the college of Computer and Information Sciences, King Saud University, 2017.
- Best poster at the 9th Undergraduate Research Conference, Zayed University, 2017.
- Dean's list for outstanding students at King Saud University, 2015 and 2016.

Skills

Programming	Python, Tensorflow, Pytorch, MATLAB, Java, R, C/C++, JavaScript, HTML, SQL
Languages	English (native), Arabic (native)

Media

MIT News *April 9, 2021*

- One-stop machine learning platform turns health care data into insights.

MIT News *December 17, 2020*

- Method finds hidden warning signals in measurements collected over time by *Daniel Ackerman.*