Abdulrahman Alabdulkareem (Abdul)

Portfolio: arkareem.com
Email: arkareem@mit.edu
Phone: (765) 772-6961

Education

- **Massachusetts Institute of Technology (MIT)** — Cambridge, MA
  Dual M.S. in EECS and CSE (Computational & Computer Science) — *2022 - 2024*
  - GPA: **5.00/5.00**
  - Relevant Courses:
    - ML: 6.S986 Large Language Models and Beyond, 6.7830 Bayesian Modeling Inference, 6.869 Advances in Computer Vision, 9.660 Computational Cognitive Science, 6.7910 Statistical Learning Theory, 6.S898 Deep Learning, 6.8630 Natural Language, 6.9320 Ethics for Engineers, 6.8200 Sensorimotor Learning
    - Math: 6.7330 Num Methods for Part Diff Eq, 6.7300 Modeling & Simulation, 6.337 Num Methods
- **Purdue University** — West Lafayette, IN
  Double Major in Computer Science and Data Science — *2016 - 2020*
  - GPA: **4.00/4.00**
  - Majors: Computer Science, Data Science; Minors: Mathematics, Statistics

Publications

- **BrainBits: Quantifying Images Reconstruction from Brain Activity** — Cambridge, MA
  Research at MIT InfoLab — 2023 - 2024
  - Worked with a large team on developing BrainBits, a method for evaluating the efficacy of image reconstruction from brain activity by using a bottleneck to quantify the information flow needed
  - Worked on decoding brain data into images with minimum bits using VQ-VAE, demonstrating that neural reconstruction requires surprisingly little neural information, suggesting the dominance of generative model priors over actual neural signal decoding
  - **Accepted to NeurIPS 2024** (Co-author)
- **Novel Unsupervised Anomaly Detection using Secure-LLM** — Cambridge, MA
  Research at MIT InfoLab — 2024 - 2024
  - Researching the use of Secure-LLM (my previous research) for novel unsupervised anomaly detection
  - The framework enables Secure-LLM to detect anomalies in the textual domain along with detecting information leakages for data leak prevention whenever data security is mandatory
  - Paper (as **main author**) pending public access; Patent being filed in cooperation with MIT TLO
- **Secure-LLM: Provably Secure Large Language Models** — Cambridge, MA
  Research at MIT InfoLab — 2023 - 2024
  - Researching and developing a novel framework to enable provably secure LLMs that are resistant to data exfiltration attacks. The framework is the first to enable LLMs to be utilized in large and segmented data silos where data security is mandatory such as within the U.S. Government
  - Paper submitted describes the framework in practice, and provides SOTA comparisons with other methods, as well as demonstrating a working version of the framework in SQL data silos
  - Paper (as **main author**) in peer-review; Patent filed in cooperation with MIT TLO
- **Defensive Poisoning for Enhanced LLM Safety** — Cambridge, MA
  Research at MIT InfoLab — 2024 - 2024
  - Researched and developed a novel approach for LLM safety by poisoning an LLM to neutralize undesirable abilities within its weights used in conjunction with RLHF, preventing LLM misuse

- Results conducted on Llama-3 8B demonstrate that positive poisoning enforces safety regulations and prevents execution of harmful commands, even after potential jailbreaks
- Paper (as **main author**) pending public access; Patent being filed in cooperation with MIT TLO

- **Research in Statistical Learning Theory** West Lafayette, IN
  Research at Purdue 2019 - 2021
  - Novel research in statistical learning theory conducted under Professor Jean Honorio
  - Results prove novel information-theoretic lower bounds for zero-order stochastic gradient estimation
  - Paper (as **main author**) accepted at IEEE 2021 (ISIT)

Projects

- **Multimodal Aviation Model for KMASS, DARPA** Cambridge, MA
  Project at MIT InfoLab 2022 - 2023
  - Was the lead developer of a Multimodal Deep Learning model for contextual clip retrieval to assist aircraft pilots as part of the Knowledge Management at Scale and Speed (KMASS) program by DARPA at the MIT InfoLab
  - Our system was capable of answering natural language queries posed by pilots in real time by providing relevant video clips from an extensive database to answer their particular query
  - Our system passed DARPA's evaluation, deployed, and shown to surpass DARPA's previous models
- **LLM Causal Understanding of NASA's ASRS Stories** Cambridge, MA
  Research at MIT InfoLab 2023 - 2023
  - Research into assessing the capabilities of state-of-the-art LLM models of understanding reports from NASA's Aviation Safety Reporting System
  - Our results demonstrate a novel lack in LLM capabilities of causally understanding stories as opposed to the typical flawed measurements of story understanding through summarization
- View other projects on my website: arkareem.com

Work Experience

- **Intelmatix** Riyadh, Saudi Arabia / Boston, Massachusetts
  Full-time AI Scientist 2022 - 2022 / 2024 - Current
  - An AI company leading Decision Intelligence for enterprises founded by MIT graduates
  - Worked on multiple Machine Learning projects such as a forecasting system for the SA government to predict the time frame and closure of requests in an internal government system using Machine Learning
- **ARAMCO, Department of PET Exploration Application Services** Dhahran, Saudi Arabia
  Full-time 2020 - 2022
  - Practical research and application of AI in Geoscience for use by field geologists in oil rigs
  - Developing scientific applications for Geologists and Geophysicists to analyze underground surfaces and the discover relevant patterns using historical patterns and numerous past drilled structures
  - Worked on projects such as predicting surface changes from coarse data like lithology and the rate of penetration of a drill bit, as well as forecasting the lithology composition of underground layers based on photographs of on-site field samples
- **Center for Complex Engineering Systems (CCES) at MIT** Cambridge, Massachusetts
  Internship 2019 - 2019
  - Worked on developing novel supervised models for dimensionality reduction / RL models
  - Our team's goal was to detect anomalies in electrical grid data and detect any potential loss of profit for the SEC (Electric Company). Our work was applied to real-world data provided by the SEC

Programming Languages

Languages: Python, Java, JavaScript, MATLAB, Julia, Gen, C/C++
Frameworks: PyTorch, TensorFlow, Scikit-learn, Numpy, Pandas, Gen, Git, React, Angular.