ARUN VIGNESH MALARKKAN

arunvignesh28.github.io • malarkkanarunvignesh@gmail.com • (+1)4802552954 • linkedin.com/in/arunvignesh28

ABOUT ME

An enthusiastic Ph.D. Candidate in Computer Science specializing in Causal Machine Learning, Reinforcement Learning, and Data-centric AI, with a strong track record of developing interpretable, robust AI solutions and advancing LLM reasoning capabilities. Actively seeking an Applied Scientist/Machine Learning Research Internship to apply research expertise to real-world challenges.

EDUCATION

Ph.D., Computer Science *August 2023 - Present*
Arizona State University, Tempe, AZ

Masters in Computer Science *May 2020*
Arizona State University, Tempe, AZ **CGPA: 3.8/4.0**

TECHNICAL SKILLS

PROGRAMMING:	Python, Java, Pytorch, Tensorflow, Huggingface Transformers, CUDA, Type Script
MACHINE LEARNING:	Causal Inference, Multi-modal Knowledge Representation, LLM Fine-Tuning, Transformers, RAG
DATABASES:	MYSQL, AWS RDS, DynamoDB, MongoDB
FRAMEWORKS:	Docker, Flask, LLM - Langchain, AutoGluon, SparkML, Sagemaker
CERTIFICATIONS:	**AWS Certified Developer Associate, Amazon MLU Machine Learning and NLP**

PUBLICATIONS

- Malarkkan, A.V., Wang, D. and Fu, Y., 2024, October. Multi-view Causal Graph Fusion Based Anomaly Detection in Cyber-Physical Infrastructures. In Proceedings of the 33rd ACM CIKM (pp. 4760-4767).
- Malarkkan, A.V., Bai, H., Wang, D. and Fu, Y., 2025. Causal Graph Profiling via Structural Divergence for Robust Anomaly Detection in Cyber-Physical Systems. arXiv preprint arXiv:2508.09504 ***(Under Review – ACM TKDD)***.
- Malarkkan, A.V., Wang, D., Bai, H. and Fu, Y., 2025. Incremental Causal Graph Learning for Online Cyberattack Detection in Cyber-Physical Infrastructures. arXiv preprint arXiv:2507.14387 ***(Under Review – IEEE TBD)***.
- Malarkkan, A.V., Bai, H., Wang, X., Kaushik, A., Wang, D. and Fu, Y., 2025. Rethinking spatio-temporal anomaly detection: A vision for causality-driven cybersecurity. *arXiv preprint arXiv:2507.08177*.
- Ying, W., Wei, C., Gong, N., Wang, X., Bai, H., Malarkkan, A.V., Dong, S., Wang, D., Zhang, D. and Fu, Y., 2025. A survey on data-centric ai: Tabular learning from reinforcement learning and generative ai perspective. *arXiv preprint arXiv:2502.08828*.
- Malarkkan, A.V. and Fu, Y., 2025. DELTA: Privacy-Preserving Generative Data Reprogramming ***(Under Review – ICDM 2025)***.
- Malarkkan, A.V. and Fu, Y., 2025. CAFE: Causally-Guided Automated Feature Engineering with Multi-Agent Reinforcement Learning ***(Under Review – AAAI 2025)***.

PROFESSIONAL EXPERIENCE

DOW Chemicals – Doctoral Researcher ***JUNE 2025 – Present***

- Research collaboration under National Academy of Engineering (NAE) Frontiers of Engineering grant.
- Developing causal-aware multi-agent reinforcement learning frameworks for AI-driven material science simulations.
- Advancing computational models for materials discovery and optimization (work subject to confidentiality agreements).

Fidelity Investments - Senior Software Engineer ***JANUARY 2023 – JUNE 2023***

- Tech-lead for developing end to end Retirement goal financial projections.
- Orchestrated and developed project retirement goal savings framework based on Monte Carlo Simulations.
- Developed ensemble outlier detection tools to determine the factors impacting the projected savings goal of the customers.

Amazon - Software Development Engineer at Alexa AI ***JUNE 2021 – JANUARY 2023***

- Implemented **end to end voice (VUI) purchase flow of Skills** that require Parental Consents enabling HIPAA compliance.
- Implemented **Purchase Likelihood Score model** which drove the Alexa Purchase Recommender system and efficiently improved the performance by 8% increase in Voice Skill purchases of that quarter.
- Implemented and integrated **localization module to the Voice Integrated Purchase Experience** for alexa skill developers.
- Designed and Implemented end to end **Voice Enablement** of **1p(Amazon-built) Skill Promotional Discounts** to Alexa Skills.

ASU Decision Center for Educational Excellence - Data Scientist/ Software Developer ***JUNE 2020 – MAY 2021***

- Developed outlier prediction models for financial plans, school going population, juvenile crime rates, and infrastructure.
- Designed and engineered end-to-end ETL pipelines and AWS Lambdas for data ingestion, modeling, and analysis.
- Implemented prediction models to optimize school resources and analyze factors influencing graduation rates, population demographics, community factors, and funding in Arizona.

ACADEMIC RESEARCH EXPERIENCE

Enhancing LLM Interpretability by leveraging Probabilistic Causal Modeling on Knowledge Graphs (*In Progress*):

- Extensive research driven towards enhancing the capabilities of Large Language Models to understand complex relationships by simulating counterfactual interventions through probabilistic causal modeling.
- Developing methodologies to construct causal knowledge graphs from LLMs and improve the robustness of LLM reasoning.
- Addressing challenges in LLM causal knowledge graph construction and reasoning paradoxes.

Causally-Guided Automated Feature Engineering with Multi-Agent Reinforcement Learning *(Under review - AAAI)*:

- Built a **multi-agent reinforcement learning framework** guided by **causal graph discovery** for automated feature generation.
- Designed **causally-aligned reward shaping** and pruning strategies to improve interpretability and efficiency.
- Outperformed state-of-the-art baselines with **higher accuracy and robustness** across benchmark datasets.

Privacy-Preserving Generative Data Reprogramming *(Under review - ICDM)*:

- Developed a two-phase framework combining **reinforcement learning** and **variational generative modeling** for feature transformation.
- Integrated **causal and adversarial disentanglement** to protect sensitive attributes while boosting model utility.
- Achieved **~9% performance gain** and **~35% reduction in privacy leakage** across 8 benchmark datasets.

Prior-Knowledge Induced Incremental Causal Graph Learning Framework for Cyberattack Detection (*Under review – ACM TBD*):

- Developed a **continual causal graph learning** framework incorporating the spectral properties to effectively represent cyberattacks in temporal streaming data.
- Developed an enhanced **experience replay** and **edge reinforcement** techniques to address catastrophic forgetting.
- Our model efficiently addresses the challenges of distribution shifts and data imbalance in real-time critical infrastructures.

Causal Graph Profiling via Structural Divergence for Robust Anomaly Detection in Cyber-Physical Systems *(Under review - TKDD)*:

- Developed a cyber-attack detection framework incorporating **causal graph learning** and **structural analysis** of the graphs representing the attack status.
- Achieved a **12% significant improvement** on the performance when compared to the best-performing standard baseline anomaly detection models on the Water Treatment Network infrastructure.

Multi-view Causal Graph Fusion Based Anomaly Detection in Cyber-Physical Infrastructures *(PUBLISHED CIKM 2024)*:

- Developed a deep learning anomaly detection framework incorporating **multi-view causal graph learning** and **spectral fusion representation** of the graphs to detect anomalies in temporal data.
- Achieved a **9% improvement** on the ROC score in determining anomalies in a large time-series dataset.

BERT based Event Extraction from clinical discharge summaries – NLP/Biomedical Data Mining:

- Developed an **Event Extraction framework** using **Bio-Bert** on the clinical discharge summaries from NLP i2b2 medical data.
- The system showed results for 29 labels, more than the labels in the state-of-the-art system and achieved 80% accuracy.

OTHER EXPERIENCE

Teaching Assistant: ***AUGUST 2023 – Present***

- Senior Teaching Assistant/ Lab Instructor - Object Oriented Programming Concepts, Data Mining, Semantic Web Mining.
- Responsible for designing course structure and developing interactive course modules .

Research Mentor: ***JANUARY 2024 – Present***

- Mentoring an Undergrad Honors thesis student on the topic of Data Mining on Temporal Systems.
- Mentoring a Graduate Student on the topic of Optimized Fine-Tuning of LLMs in Reasoning Tasks.

Research Workshops:

- Hands-on workshop, “Getting started with Machine Learning” for Freshmen and Sophomore students to get them kick-started with their AI journey at Pondicherry University, India.
- Hands-on workshop, “Basics of Graph Neural Networks and its applications” at Pondicherry University, India

Miscellaneous Experience:

- **PC member** – AAAI 2026, IEEE International conference on System, Computation, Automation and Networking 2024, IEEE BigData’24 Undergraduate and High School Research Symposium.
- **Conference Reviewer** – KDD, ICDM, Urban AI, IEEE BigData, ICLR, Neurips.
- **Journal Reviewer** – ACM Transactions on Knowledge Discovery from Data.

ACHIEVEMENTS

- ***Academic Scholarship – Arizona State University, Fall 2018***
- ***Best Under-Graduate Researcher Award 2017-2018.***