

Muhammad Arbab Arshad

linkedin.com/in/arbab-arshad

Email: arbab.arshad97@gmail.com

Mobile: 848-313-9857

U.S. Permanent Resident (EB-1A Extraordinary Ability) — No Sponsorship Required

TECHNICAL SKILLS

- **Large Language Models:** Hugging Face, LangChain, vLLM (inference engine for LLMs), TorchTune (for LLM fine-tuning), MLflow (for model tracking), Retrieval Augmented Generation, LLM Quantization, Vision Language Models
- **Machine Learning & AI:** PyTorch, TensorFlow, CUDA, MLOps, Computer Vision, Distributed Systems, NLP
- **Programming Languages:** Python, Java, C++, JavaScript, R, SQL, Scala, Go
- **Development:** React, Node.js, Express, REST APIs, Gradio, MongoDB, PostgreSQL
- **Infrastructure & Tools:** AWS, Docker, CI/CD, Terraform, Git, Linux, Agile, kanban, Jira, Bash

EXPERIENCE

- **AI Institute for Resilient Agriculture (AIIRA), Iowa State University** Iowa, United States of America
Ph.D. Candidate and Research Assistant Jan 2023 – Present
 - **Agricultural LLM:** Developed Large Language Model for precise agricultural recommendations using expert-verified data on 90 species. Implemented Retrieval Augmented Generation with 82% recall and 65% precision. Integrated multi-LLM support for enhanced performance. (AgLLMs.github.io)
 - **LLM Evaluation:** Developed 12-task agricultural benchmark to evaluate multimodal LLMs. Established baseline metrics, achieving up to 73.37% F1 score with few-shot learning. (AgLLMs.github.io/AgEval)
 - **3D Modeling:** Evaluated Neural Radiance Fields for detailed 3D plant reconstruction in various environments. Achieved 74.6% accuracy in challenging outdoor scenarios, demonstrating potential for complex modeling. Developed optimization technique reducing training time by 50% with minimal accuracy loss.
- **Intel** California, United States of America
Graduate Software Engineering Intern Feb 2025 – Aug 2025
 - **Computer Vision-based UI Automation via Large Language Models (LLMs):** Architected autonomous agent framework leveraging multimodal LLMs to generate cross-platform automation scripts, reducing manual testing workflows. Implemented adaptive vision-to-code generation for complex UI interactions.
 - **Visual Anomaly Detection for Graphics Validation:** Contributing to initiative addressing Intel's \$6.5M annual manual testing costs for hardware-game compatibility validation. Developing real-time anomaly detection system using Vision-Language Models across diverse rendering scenarios.
 - **LLM Benchmarking & Hardware Enablement:** Implementing production deployment of cutting-edge LLMs (DeepSeek, Llama, Qwen, Janus) on Intel GPUs through OpenVINO and PyTorch optimizations as part of Intel's AI acceleration team. Executing systematic performance analysis across Intel and NVIDIA hardware platforms for strategic product positioning.
- **Kingland Systems** Iowa, United States of America
Software Engineering Intern May 2023 – Aug 2023
 - **Cloud Deployment:** Implemented auto-scaling AWS Fargate deployment and developed end-to-end stress testing pipeline using JMeter and Blazemeter, optimizing resource usage and enabling comprehensive load testing
 - **CI/CD Enhancement:** Customized GitLab CI/CD pipeline to seamlessly integrate testing, ensuring uninterrupted development workflow
 - **Recognition:** Received formal recognition in two sprint retrospectives for establishing comprehensive load testing baseline
- **Laboratory for Software Design, Iowa State University** Iowa, United States of America
Research Assistant Jan 2022 – Aug 2022
 - **Program Repair:** Executed 5 automated repair tools on GPU clusters for empirical study. Optimized parallel execution, reducing time by 16x across 40 clusters.
 - **Research Impact:** Publication received Distinguished Paper Award at 38th IEEE/ACM Conference on Automated Software Engineering.

EDUCATION

- **Iowa State University** Iowa, United States of America
Doctor of Philosophy (Ph.D.) in Computer Science; GPA: 3.96 Jan 2022 – Dec 2025
- **American University of Sharjah** Sharjah, United Arab Emirates
Master of Science (M.S.) in Computer Engineering; GPA: 3.52 Aug 2019 – Aug 2021

PROJECTS

- **MeditateGPT**: Designed and developed an application for customized guided meditations using GPT-4 API, MERN Stack, Amazon Polly, and AWS S3.
- **Bat Behavior Analysis**: Adapted unsupervised ML image clustering algorithms to audio data for bat behavior analysis, achieving 88.28% accuracy in classifying bats.
- **Emotional Melody**: Developed a text-to-audio generation system for poetry-to-melody using GANs, generating melodies with 68% perceived similarity to real melodies.
- **Intrusion Detection**: Utilized Amazon EI to remotely detect SSH and FTP brute-force attacks, achieving 99% F1 score and 8x speed increase compared to local inference.

SELECTED RESEARCH PUBLICATIONS

- **Leveraging Vision Language Models for Specialized Agricultural Tasks**: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2025.
- **Assisted Few-Shot Learning for Vision-Language Models in Agricultural Stress Phenotype Identification**: Advances in Neural Information Processing Systems (NeurIPS), Workshop on Adaptive Foundation Models: Evolving AI for Personalized and Efficient Learning, 2024.
- **Mutation-based Fault Localization of Deep Neural Networks**: 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2023.
- **Putting GPT-4o to the Sword: A Comprehensive Evaluation of Language, Vision, Speech, and Multimodal Proficiency**: Applied Sciences, 2024.
- **Comparison of deep learning algorithms for site detection of false data injection attacks in smart grids**: Energy Informatics, 2024.
- **Evaluating Neural Radiance Fields (NeRFs) for 3D Plant Geometry Reconstruction in Field Conditions**: Plant Phenomics, 2024.
- **Forecasting highly fluctuating electricity load using machine learning models based on multimillion observations**: Advanced Engineering Informatics, 2022.