Marwan AbdElhameed

linkedin.com/in/marwan-abdelhameed | marwan@nyu.edu | +971565377128 | github.com/MarwanWalid2

EDUCATION

New York University New York, NY

***Bachelor of Science in Computer Science**, Minor in Interactive Media* *GPA: **3.887** (Cum Laude)*

- **Relevant Coursework:** Deep learning systems (Graduate level), Database Systems, Computer Vision
- **Languages:** Arabic (Native), English (Fluent), German (Intermediate)
- Graduate-level coursework: Deep Learning Systems, Operating Systems, and Open Source Development

Languages: Arabic (Native), English (Fluent), German (Intermediate)

RESEARCH PUBLICATIONS

Accepted Papers

- Kassab, M., **AbdElhameed, M.** (2025). Mapping the Landscape of Requirements Completeness: Definitions, Techniques, Tools, and the Emerging Role of AI. *Systematic Literature Review* (Accepted: Euromicro SEAA 2025)

Preprints & Under Review

- Kassab, M., **AbdElhameed, M.** (2025). Automated Analysis and Correction of Incomplete Software Requirements using a Multi-LLM Approach. *Under Review*
- **AbdElhameed, M.**, Halim, P. (2024). Inference Scaling vs Reasoning: An Empirical Analysis of Compute-Optimal LLM Problem-Solving. *New York University*

PROFESSIONAL EXPERIENCE

NYUAD, Center for Interacting Urban Networks (CITIES) Abu Dhabi, UAE

Research Assistant *February 2024 – August 2024*

- Engineered data pipeline processing **1.4M** audio spectrograms using **NumPy**, and custom **CUDA** kernels, boosting efficiency by **35%** and reducing processing time by **50%**
- Deployed scalable ML models with **Singularity** in HPC environments, reducing deployment time by **30%**
- Led data analysis with **Numpy** and **Matplotlib**, improving research insights by **25%**

AIREV HOLDING LIMITED Remote

Software Engineer Intern *June 2024 – July 2024*

- Developed and maintained robust RESTful APIs using **JavaScript** to enable seamless data communication between microservices
- Leveraged **Git** and **GitHub** for version control, collaborating with senior developers to ensure code quality
- Conducted comprehensive API testing and debugging, enhancing backend system reliability and performance

NYU Tandon, High-Performance Computing Group New York, NY

Research Assistant *August 2023 – Dec 2023*

- Optimized 3DMHD benchmark simulation performance by **40%** through integration of **Parallel processing**, using **C++**, and **MPI**
- Improved **GCP** on-prem cluster utilization and job scheduling efficiency by **25%** by implementing comprehensive performance tuning and intelligent workload distribution.
- Led code optimization, GPU utilization, and parallel computing practices, reducing error rates by **20%**

DigitalEnergy.ai Remote

Project Manager Intern *May 2023 – August 2023*

- Implemented **Agile methodologies**, increasing productivity by **30%** across business units
- Conducted AI trend research, delivering data-driven reports that improved operational efficiency by **40%**
- Led international collaboration on AI tool implementation, reducing task completion time by **45%**

SOFTWARE ENGINEERING & DATA ANALYSIS PROJECTS

AI-Powered Requirements Analysis Platform ***Python**, Flask, JavaScript, Multi-LLM APIs*

- Architected intelligent SRS analysis platform integrating **3 LLM APIs** with advanced aggregation strategies, achieving **95%** consensus accuracy through meta-analysis
- Engineered document processing pipeline supporting PDF/DOCX/TXT with Mistral AI OCR integration, achieving **99%** text accuracy and automatic requirements extraction from **100%** document types
- Developed UML domain model generator creating class diagrams from natural language requirements, reducing manual modeling time by **80%** with **100%** PlantUML compatibility
- Implemented requirements completeness analyzer detecting missing requirements and inconsistencies, improving SRS quality by **60%** through automated issue detection and intelligent suggestions

- Built interactive web interface with session management, enabling users to accept/reject AI suggestions with real-time updates and persistent state tracking

Distributed Model Serving Platform ***Flask**, Ray.serve, TorchX*

- Built scalable REST API handling concurrent model inference requests, achieving **99%** uptime
- Integrated Ray with TorchX for distributed computing, processing **10,000+** concurrent requests across multiple nodes
- Implemented efficient task distribution and monitoring system, reducing inference latency by **20%**

Photo Sharing Web Application ***Python**, Django, SQL, PostgreSQL*

- Built full-stack app with user authentication, photo upload, and album management, increasing retention by **25%**
- Optimized data relationships and queries, improving retrieval efficiency by **40%**
- Implemented social features (friends, tagging, commenting), boosting user engagement by **30%**

Personal Finance Manager ***Python**, Flask, PyMongo, Docker*

- Developed finance management web app with interactive charts, increasing user engagement by **40%**
- Implemented efficient data management with PyMongo, reducing query response time by **50%**
- Deployed using Docker and DigitalOcean with CI/CD pipeline, ensuring **99.9%** uptime

NYT Articles Analysis Pipeline ***Python**, Data Analysis, NLP*

- Engineered scalable data pipeline processing 100,000+ NYT articles monthly, implementing advanced web scraping and NLP techniques to achieve 98% content extraction accuracy while reducing API costs by 40% through intelligent limiting
- Developed and optimized text analytics system using NLTK and Gensim, reducing processing time by 75% for 1M+ words while achieving 92% topic coherence in content classification across 15 distinct topics
- Architected end-to-end machine learning pipeline leveraging Latent Dirichlet Allocation (LDA) for topic modeling, processing 50,000+ words/minute with custom preprocessing, achieving sub-second visualization response times for 1M+ data

What to Eat Today? ***JavaScript**, Node.js, Express.js, MongoDB*

- Launched meal suggestion app, attracting **1,000+** active users in three months
- Integrated with Spoonacular API, offering **5,000+** recipes with **99.9%** uptime
- Implemented AJAX for real-time updates, achieving **98%** data integrity rate

MACHINE LEARNING & AI PROJECTS

ML Infrastructure Optimization ***Python**, Ray, TorchX, W&B*

- Architected batch prediction system using Ray actors and TorchX, processing **250** requests per batch with 0.5 GPU allocation
- Implemented multi-worker data loading system and distributed hyperparameter optimization, identifying optimal configurations
- Developed threaded execution pipeline using Python's threading and ZMQ, reducing model serving latency by **28%**

Seq2Seq Conversational AI ***PyTorch**, NLTK, Neural Networks*

- Engineered sequence-to-sequence model processing **220,000+** movie dialogue exchanges with nn.GRU and nn.Embedding
- Implemented Luong attention mechanism, improving response relevance by **30%**
- Developed custom text preprocessing pipeline handling **18,000+** unique vocabulary tokens efficiently

SSD ONNX Model Integration ***PyTorch**, ONNX, Computer Vision*

- Engineered computer vision model achieving **93.33%** accuracy in object detection using MobileNetV1 SSD
- Optimized model performance through hyperparameter tuning, achieving **94.31%** validation accuracy
- Reduced inference time by **25%** through ONNX model conversion and improved model robustness by **15%**

Distributed ML Training System ***PyTorch**, CUDA, Parallel Computing*

- Developed distributed training system using PyTorch's DataParallel, achieving **1.92x** speedup with 4 GPUs
- Optimized batch size scaling to achieve **31,063** samples/second throughput, a **7.3x** improvement
- Reduced I/O bottlenecks by **40%** through custom data loading pipeline optimization

Math Problem-Solving LLM System ***Python**, Large Language Models*

- Fine-Tuned Mistal-7b achieving **35%** accuracy on GSM8K dataset
- Enhanced solution accuracy by **25%** from base model through advanced inference scaling techniques
- Created systematic evaluation framework for optimal prompting and inference strategy selection

Multilingual RAG System ***Python**, mBART, Vector DB*

- Architected multilingual RAG system with mBART-large-50, enabling cross-lingual content generation
- Improved content relevance by **40%** through context-aware document fetching
- Scaled system to handle **100k+** document pairs with optimized memory usage