

Abdullah Al Mamun, PhD

Fremont, CA, 94539

xxxxx@gmail.com, +1-786-451-xxxx

YOE: 7 years, Visa: Green Card

[Linkedin.com/in/newAbdullah](https://www.linkedin.com/in/newAbdullah)

<https://newabdullah.com>

Google Scholar

SKILLS

Python, CI/CD, Docker/container, Kubernetes, AWS, GPU, CUDA, PyTorch, ONNX, Spark, Ads ranking, RecSys, Transformer, LLM, GPT, LLaMA, Hugging Face, Embedding retrieval, Vector DB, RAG, LangChain, Agentic AI, Fine-tuning, LoRA, QLoRA, RLHF, PPO, DPO, Quantization, Pruning, 4D Parallelism, KV-Cache, vLLM

EXPERIENCES

- Aisera Inc** Palo Alto, USA
Sr. Research Scientist 06/2024–present
 - Building end-to-end scalable and reusable Agentic AI for employee onboarding (HR), driving ~\$xxM ARR.
 - Helped finetuning LLaMA-3 70B and deployed it on AWS, reducing CAPEX significantly compared to GPT-4.
 - Enhanced RAG pipeline that improved real-time semantic search experience by ~93%.
- Meta** Menlo Park, USA
Research Scientist 05/2022–06/2024
 - LLM in Ads creation process leading to a 0.09% incremental revenue (iRev) growth (\$xxM).
 - Fine-tuned LLaMA 3 for ad creation workflows and inference optimization (leveraging SFT, RLHF, KV cache, 4D parallelism, quantization, and distillation), achieving a 13% increase in CTR and a 6% increase in CVR by enabling efficient AI-driven text suggestions.
 - Drove a 0.031% iRev growth (\$xxM) by enhancing Personalized Ads Recommendations.
 - Created automated custom audience (AutoCA) clusters to fix miscalibrations in ad ranking. Leveraging targeting relaxation technology, AutoCA efficiently captures user and ad interactions, significantly improving model predictions by 0.1% irev.
- Florida International University** Miami, USA
Graduate Research Assistant (AI/ML) 09/2018–04/2022
 - End-to-end development of an interpretable deep learning models for early cancer detection and drug recommendation.
- King Fahd University of Petroleum & Minerals** KSA
Machine Learning Engineer 01/2015–12/2016
 - Implemented LSTM-based sentiment analysis tool to analyze customer feedback with 98% accuracy.
- Softwindtech Ltd.** Bangladesh
AI/ML Engineer 07/2012–07/2014
 - Design and developed personalized web assistant using semantic search.
 - Collaborated with other engineers to identify and alleviate several bugs and errors in different software.

EDUCATIONS

- Florida International University** Miami, USA
PhD in Computer Science (Applied ML) 08/2018–04/2022
- King Fahd University of Petroleum & Minerals** KSA
Masters in Computer Engineering (ML in Computer Vision)
- Dhaka University of Engineering and Technology** Bangladesh
Bachelor in Computer Science and Engineering (ML in NLP)

CERTIFICATIONS

- Google:** End-to-End Machine Learning with TensorFlow on Google Cloud Platform (Issued: March, 2020)
- University of Illinois Urbana-Champaign:** Cloud Computing Concepts: Part 1 & 2 (Issued: April 2015)

AWARDS RECEIVED

- Champion:** 2nd GCC Robotics Challenge at Doha, Qatar, 2017. Organized by IEEE and IET
- Conference Travel Fellowships:** 10th ACM BCB, NYC (2019); IEEE BIBM; San Diego (2019 and 2020)