

MD ASIF KHAN TONAY

AI Engineer | LLMs, RAG & Agentic AI | Computer Vision | MLOps

Dhaka, Bangladesh | (+880) 1829312850 (WhatsApp) | aktonay63@gmail.com
linkedin.com/in/aktonay | researchgate.net/profile/Mohammad-Asif-Khan-Tonay

PROFESSIONAL SUMMARY

AI Engineer building production-grade AI systems, LLM-powered applications, computer vision solutions and ML pipelines. Experienced in Generative AI, RAG, prompt engineering, AI agents, agentic workflows, OCR and fine-tuning. Currently developing enterprise AI solutions including Odoo-integrated sales intelligence agents, multimodal AI applications and scalable ML deployment pipelines.

WORK EXPERIENCE

Software Engineer Level 1 — Betopia Limited

Apr 2026 – Present

- Led end-to-end design, development and deployment of production-grade LLM-powered AI agents and ML systems for enterprise clients across multiple verticals.
- Engineered the **Sales Intelligence Agent**: Odoo-integrated, FastAPI-based, with an LLM fallback chain (vLLM / GPT-4 / Claude Sonnet), RAG, tool calling and automated quotation generation.
- Built computer vision solutions (OCR, image processing, PyTorch, TensorFlow) for document intelligence and multimodal workflows; designed cloud-native infrastructure with Docker, PostgreSQL, Redis, Qdrant, Celery and NATS.
- Applied stacked ensemble modeling, feature engineering and AIOps-driven model monitoring to keep prediction pipelines high-precision in production.

Junior AI Developer — SparkTech Agency, Betopia Group

May 2025 – Apr 2026

- Designed and deployed ML/DL models for production apps: receipt OCR (GastRonomiq), AI scheduling (Grafic Master), AI family companion (ChatFam), menu management (Menu Sidekick).
- Delivered GenAI features (prompt engineering, OpenAI API, LLM conversational workflows) across six apps shipped to App Store and Google Play.
- Managed big data pipelines (Python, Pandas, NumPy, PySpark, SQL), built CV models, and implemented MLOps workflows with Docker, FastAPI and AIOps practices.

SELECTED PROJECTS

Sales Desk — Sales Intelligence Agent · Betopia / Odoo Platform. Sales.betopia.ai

Enterprise AI sales agent on FastAPI with LLM fallback chain (vLLM / Self Hosted), RAG, requirements extraction, automated quotation generation, tool calling and cross-profile intelligence for 500+ Fiverr profiles.

Betopia Limited AI Chatbot · betopialimited.com

RAG knowledge-base chatbot with two-tier RBAC (public vs. authenticated via Odoo/ERP SSO), Qdrant hybrid search, self-hosted LLM, source-linked answers, tiered fallback with human PSM handoff, and an admin dashboard; public and ERP-integration APIs documented in Swagger.

AkBORIA SR — AI Field Service App · Odoo ERP · Flutter.akboria.com

AI service app for a distribution network with real-time Odoo data via JSON-RPC, a nightly six-parameter shop-priority scoring model, and a routing engine (self-hosted OSRM + Nearest-Neighbour, 2-opt, Or-opt) for daily SR visit sequencing; Bengali localization.

Task Mama · taskmama.app · App Store · Google Play

AI household management assistant that helps families assign tasks, organize chores and streamline routines through smart AI suggestions.

LearninGPT · learnin-gpt.com

AI-powered educational backend platform delivering intelligent learning assistance and adaptive content.

TECHNICAL SKILLS

AI / ML: AI, Machine Learning, Deep Learning, Data Science, Big Data Analytics, PyTorch, TensorFlow, Scikit-learn, XGBoost, Pandas, NumPy, PySpark

LLM & GenAI: LLMs, Generative AI, RAG, AI Agents, Agentic AI, Prompt Engineering, OpenAI API, Hugging Face Transformers, vLLM, Fine-Tuning, Tool Calling, Embedding Models, Vector DBs (Qdrant), NLP

Computer Vision: OCR, Digital Image Processing, OpenCV, Multimodal AI

MLOps / Deploy: FastAPI, Docker, REST APIs, OpenAPI / Swagger, Cloud Deployment, CI/CD, DevOps, MLOps, Model Monitoring, Inference Optimization

Data & Languages: Python, SQL | PostgreSQL, Redis, Qdrant, MongoDB

EDUCATION

Professional Masters in Computer Science — Jahangirnagar University

May 2026 – Present

Focus: Advanced AI, ML systems, algorithm design, applied research

B.Sc. in Computer Science and Engineering — East West University

May 2020 – Dec 2024

Thesis: Laryngeal Cancer Detection and Classification Using Deep Learning on Histopathological Images

PUBLICATIONS

- Laryngeal Cancer Detection and Classification Using Deep Learning on Histopathological Images (2025)** — IEEE IT 2025, 29th Int'l Conf. on Information Technology · doi.org/10.1109/IT64745.2025.10930285
- Automated Laryngeal Cancer Detection Using Endoscopy Images with Deep Learning (2024)** — Springer Nature, Data Mining and Information Security, Vol. 2, pp. 361–380 · doi.org/10.1007/978-981-96-6053-7_24
- Enhanced Nano-Scale Material Property Prediction in Li-Ion Batteries with Stacked Ensemble ML Models (2024)** — Springer Nature, Data Mining and Information Security, pp. 379–392 · doi.org/10.1007/978-981-96-6046-9_24

HONOURS, AWARDS & LANGUAGES

- Galactic Problem Solver** — NASA International Space Apps Challenge (Oct 2024)
- Anthropic Partnership Learn Claude** — Anthropic (June 2026)
- Languages:** Bengali (Native) · English (Fluent — reading, speaking, writing)