Ismail Alshofi

AI Engineer · Agentic AI, Voice & Multimodal Systems

ialshofi@gmail.com | +963 988 026 431 | +971 50 484 9581
linkedin.com/in/ismail-alshofi-323965168 | github.com/ismailalshofi | ismail-alshofi.vercel.app

PROFILE

AI Engineer with 3+ years in ML and AI, focused on production agentic and real-time voice systems. Built voice pipelines reaching ~900 ms end-to-end response time on live phone calls and LLM orchestration layers with 40+ tools.

SKILLS

Agentic AI: Multi-Agent Systems, Tool Calling, MCP, RAG, Memory, Context Engineering, LLM Evaluation

LLMs & Voice: Claude, GPT, Gemini, LangGraph, LangChain, Real-Time Voice Agents, Speech-to-Text, Text-to-Speech

Backend & ML: Python, FastAPI, NestJS, WebSockets, PostgreSQL, pgvector, Qdrant, Milvus, Docker, AWS, PyTorch, TensorFlow, OpenCV

Languages: Arabic (native), English (fluent)

PROFESSIONAL EXPERIENCE

AI Engineer
Sep 2024 – Oct 2026

Skew (USA) · Remote

- Built a **multi-agent, real-time voice pipeline** for live phone calls (Deepgram, ElevenLabs, Chatterbox), cutting latency by **~60%** to **~900 ms from end of caller speech to first reply audio**.
- Extended voice agents to inbound and outbound telephony and to a **physical robot** controlled through LLM tool calling.
- Built an LLM orchestration layer with **40+ tools** for retrieval, actions, and multi-step workflows, using **mode scoping** and context engineering to keep responses accurate and consistent.
- Designed agent architectures with memory, **RAG** (pgvector, Milvus), and **multiple personas** that stay consistent across chat and voice, with evaluation workflows to monitor response quality.
- Built an AI matching system over **10,000+ specification documents**, making a **2–3 hour manual lookup** instant.

Machine Learning Researcher
Dec 2023 – Jul 2024

Damascus University × Ministry of Health, Syria

- Designed a **two-stage deep learning pipeline** (U-Net segmentation, then ensemble localization) for cephalometric landmark detection, placing **52.6%** of predictions across **19 landmarks** within 2 mm and a further **26.3%** within 2–3 mm.

Machine Learning Engineer · Mobile Developer until Mar 2023
Feb 2022 – Dec 2023

Smart Software Solutions, Qatar

- Developed an anomaly detection model that flags unusual shipments and transactions in a heavily imbalanced dataset; applied NER and POS tagging to extract entities from messages and documents.

OPEN SOURCE

- **Beyond Words**: real-time voice agent where a scenario-tested policy decides turn-taking (barge-in, backchannels, adaptive pauses).
- **rag-structure**: RAG API on Chroma, Qdrant or pgvector behind one interface, with checked citations and retrieval evals.
- **Tokenization study**: char vs BPE Transformer on a Shakespeare dataset I published; bits-per-character reverses the per-token result.

TECHNICAL WRITING

How to Plan a Premium AI Agent: A Practical Guide to Designing AI Applications | *Medium*

Mode-Scoping: One AI Agent Instead of Many | *Medium*

EDUCATION & AWARDS

Damascus University
2018 – 2024

B.Eng. in Information Technology Engineering — specialization in Artificial Intelligence & NLP

5th Place, Syrian Smart Systems Competition | **Kaggle Notebook Expert** | *Top ~5% (3,302 of 62,341)*