

# Omar Adel

## Machine Learning Engineer

[omarxadel21@gmail.com](mailto:omarxadel21@gmail.com)

+20 120 422 7071

Alexandria, Egypt | **Open to Relocation**

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

[github.com/omarxadel](https://github.com/omarxadel)

[huggingface.co/omarxadel](https://huggingface.co/omarxadel)

## RESEARCH & OPEN-SOURCE

---

### Arabic LLM Tokenization & Morphology | MSc Research Project 2024 – 2025 · Alexandria University

Investigated sub-word tokenization failures in Arabic LLMs. Developed and published three fine-tuned model variants:

- **Arabic-Tokenskip-Qwen2-7B** – novel token-skipping technique reducing Arabic tokenization fragmentation, measurably improving model Arabic comprehension
- **Arabic-Morph-Qwen2-7B** – morphology-aware fine-tuning improving Arabic instruction following and reasoning
- **Arabic-Morph-DeepSeek-R1-Distill-Llama-8B** – applied same morphology approach to DeepSeek-R1 distilled model
- Published **compressed\_arabic\_reasoning\_dataset** (9.2k rows, LLM-generated) and **MaWPS-ar** (5.1k rows, crowd-annotated, cited) – Arabic math reasoning datasets

### Breast Cancer Detection via Quantized CNN | MSc Course Project 2024 · Alexandria University

- Applied model quantization and EdgeML techniques to a CNN classifier for breast cancer detection; achieved ~94% reduction in model size with no significant drop in accuracy
- Demonstrated practical viability of deploying diagnostic ML models in resource-constrained environments

### Hearsome: Egyptian Arabic Speech Recognition | Lead Researcher 2021 – 2022 · Alexandria University

- Fine-tuned **HuBERT** and **wav2vec2-large-xlsr** for Egyptian Arabic ASR; achieved <25% WER. Deployed model reached **14,000+ Hugging Face downloads**
- Designed full system architecture: mobile app, server, admin panel, and model deployment pipeline

## EXPERIENCE

---

### AI R&D Engineering Manager | Turing Aug 2025 – Present · Remote

Lead a core R&D team of 25–40 engineers and researchers delivering LLM training data and evaluation infrastructure to **Anthropic, Microsoft, ServiceNow, and Amazon**.

- Operate as an **internal company within Turing** – own research direction, roadmap, revenue, and all client relationships end-to-end
- Oversee full data lifecycle: generation, benchmarking, fine-tuning supervision, and model evaluation across **Instruction Following** and **Agentic Intelligence** domains
- Fine-tuned **Qwen3**, **Nemotron** and other frontier models for client-specific alignment objectives
- Built internal tooling and dashboards for data validation, automation, and eval benchmarking
- Conduct enterprise technical pitches and SOW preparation to secure new client engagements

### R&D Team Lead | Turing Oct 2024 – Jul 2025 · Remote

- Established and scaled the Reasoning R&D Division for a Fortune 500 client, growing from **5 to 40 core researchers**
- Built evaluation frameworks for **chain-of-thought and reasoning benchmarks** enabling faster model iteration
- Integrated eval insights directly into client model improvement cycles with ML Science teams

### Machine Learning Research Engineer | Turing May 2024 – Sep 2024 · Remote

- Joined data curation, task design, and large-scale generation pipelines for function calling and coding domains
- Improved data generation efficiency by **30%** via workflow automation and optimised trainer notebooks
- Built proof-of-concept eval pipelines measuring function-calling accuracy and code execution reliability

### Software Engineer, Machine Learning | Pleny Sep 2021 – May 2022 · Alexandria, Egypt

- Researched and prototyped recommendation engines; developed 4 backend microservices serving thousands of users

## EDUCATION

---

### M.Sc. Computer and Systems Engineering | Alexandria University

2023 – Present

Focus: NLP & Language Models. Research areas: Arabic reasoning, morphology-aware fine-tuning, hate speech recognition, and guided dataset construction.

### B.Sc. Computer and Communications Engineering | Alexandria University

2017 – 2022

## TECHNICAL SKILLS

---

<b>ML / Research</b>	PyTorch, HuggingFace Transformers, SFT, RLHF/RLAIF, embedding models, RAG, LangChain, LangGraph, PydanticAI, NLP, ASR
<b>Fine-tuning</b>	LoRA, QLoRA, Unsloth, PEFT, SFTTrainer, full fine-tune, prefix tuning
<b>Models trained</b>	Qwen2-7B, DeepSeek-R1-Distill-Llama-8B, Flan-T5-XL, mTO, HuBERT, wav2vec2
<b>Languages</b>	Python (primary), TypeScript, JavaScript, C/C++
<b>Infrastructure</b>	AWS, GCP, PostgreSQL, MongoDB, Git, FastAPI, Flask

## PUBLISHED MODELS & DATASETS

---

**Hugging Face**     [huggingface.co/omarxadel](https://huggingface.co/omarxadel) – 8 models, 2 datasets, 20,000+ total downloads

**Key artifacts**     Arabic-Tokenskip-Qwen2-7B, Arabic-Morph-DeepSeek-R1-Distill-Llama-8B, compressed\_arabic\_reasoning\_dataset, MaWPS-ar

## RESEARCH PAPERS

---

3 research papers written across NLP and ML topics — currently unpublished. Available upon request.