Abdelrahman **Diab**

Machine Learning Engineer

CONTACT INFO

📞 **(+49)15560048933**

✉ **abdrahmandiab99@gmail.com**

in **abdelrahman-diab1/**

📍 **Penckstraße 8, 80995, Munich**

EDUCATION

M. Sc. Informatics

Technical University in Munich (TUM)
2024 SS - Present

B.Sc. Computer Science and Engineering

German University in Cairo (GUC)
GPA: 1.52 (A)
2017 WS - 2022 SS

SOFTWARE SKILLS

Programming Languages
- Python, Java, Spring Boot, Swift, Javascript (RubyonRails, ReactJS, NodeJS), C++, GStreamer.

ML Libraries
- PyTorch, Tensorflow, Jax, Matplotlib, OpenCV, ONNX.

Version Control & DevOps
- Git, GitHub Actions & CI/CD tools.
- Docker, Kubernetes.
- Shell Programming (bash / zsh)

Utilities
- Microsoft Office, Canva, Figma, JWT, JSON.

LANGUAGES
- **Arabic** - Native
- **English** - Fluent
- **German** - Intermediate (B1)

PERSONAL HOBBIES

Music
- Played multiple sold-out shows across my home town Cairo as the guitarist for "Skeleton Clique" band.

PROFILE

TUM Master's in Informatics student with a strong ML background, and a passion for creative, user-focused solutions eager to merge ML expertise with Cloud Technologies to build impactful projects.

WORK EXPERIENCE

Machine Learning Engineer   Jun 2024 - Present

Logivations | Part Time Working Student
- Fine-tuned the MMPose model for keypoint detection and integrated it into an existing *GStreamer* pipeline using a *TensorRT/ONNX* deployment workflow.
- Enhanced forklift pallet pick/drop **localization accuracy by 47% (ME)** through a long data analysis of domain-informed dynamic strategies, optimizing scenario-specific decisions.

Machine Learning Engineer   May 2022 - Dec 2023

Ivy - The Unified AI Framework | Full Time (Remote)
- Contributed to building core functions of the **Ivy Transpiler**, a critical component for enabling code conversions across ML frameworks (*NumPy*, *PyTorch*, *TensorFlow*, *Jax*).
- Designed UI components and interactive demos for the company website using Figma for design and ReactJS for implementation as the **UX Team** leader.
- Assisted in developing a docker based **AWS Lambda** service for transpiling code across the four target ML frameworks.

NOTABLE COURSES

NLP Practical Course: Adaptive parameter Compression   **TUM , 2025**
- Explored **Weight Compression Techniques**, specifically:
  - *Pruning*, *Knowledge Distillation*, *Quantization*, and *Weight Initialization* strategies.
- Fine-tuned **LLaMA-3.2-1B** (decoder) and **DeBERTa-base** (encoder) with APC, achieving up to ~50% inference speed-up on LLaMA while retaining 90–95% of base model performance across **GLUE** tasks.

Architecture of Massively Scalable Applications   **GUC , 2022**
- Covered core concepts of **cloud computing** and studied server architectures as well as tools for ensuring **scalability**, **availability**, and **fault tolerance**.
- Lead a cluster of 15 students split into 3 teams to architect, develop and test a replica of the LinkedIn social media platform using **Java Spring Boot**. Employing elements of:
  - Micro-services, Load Balancing, Message Queues, Containerization, Observability through logging, Distributed Databases & Caching, and Integrated API gateways.

Deep Learning Specialization (5 courses)   **deeplearning.ai , 2021**
- Implemented & trained classical DL Networks such as **YOLOv4 Car Detection**, **Facial Recognition** and **Art Generation** using a mix of **PyTorch** and **TensorFlow**.

RELEVANT PROJECTS

Master's Thesis - Agentic regression bug testing   In Progress
- Designing an MCP + Github Bot for automated regression bug bisection.
- The goal is a platform-agnostic framework that can autonomously build and test projects from GitHub to visually identify and isolate UI regressions

TapGig - Gstreamer Migration   2025
- Partnered with Florian Dejako (Ex-CTO, Elgato) to architect and implement a high-fidelity, real-time streaming platform on a **GStreamer** and **Ruby on Rails** monolith.
- Designed the core **GStreamer media pipeline** to process audio and video.
- Implemented a control interface for live-switching video feeds, managing audio levels, and communicating with talent via a low-latency intercom channel.