Adil Pervez

adilmp14@gmail.com | +92-335-2969676 | LinkedIn | GitHub

Profile

AI Engineer with **two years of industry experience in AI** at US and UAE AI companies (Turing, CNTXT AI), building **RLHF preference data and training datasets** for frontier LLMs and computer vision models. Builds production-style **LLM applications** in Python (RAG pipelines, SQL agents with guardrails) with **evaluation harnesses** and CI gates, and **fine-tunes transformer models** with PyTorch and Hugging Face.

Education

National University of Computer and Emerging Sciences (FAST NUCES) — Pakistan
Bachelor of Science in Computer Science — *Sept 2020 – June 2024*

Projects

Multilingual LLM SQL Agent with Guardrails | *Python, sqlglot, SQLite, FastAPI, Ollama, Anthropic API* GitHub

- LLM assistant answering database questions in **English, Arabic and Urdu** at **85% execution accuracy** (67% on hard held-out questions) on a local 7B model, no GPU; swappable backends (Ollama, Anthropic API) via FastAPI.
- Validates every query as a syntax tree against the real schema and runs it read-only: **0 harmful statements executed** under adversarial prompts; **439 tests** and a CI regression gate.

BookMind: RAG Pipeline with Evaluation Harness | *Python, Claude API, FastAPI, Streamlit, Docker* GitHub

- Built a **retrieval-augmented generation** pipeline: book ingestion and chunking, a hand-written BM25 retriever, and grounded answers from Claude with **chapter citations** and a refusal guardrail.
- Evaluation harness (**83% Recall@5**, 0.688 MRR, citation and hallucination checks); exposed as a FastAPI REST service with a Streamlit UI and Docker.

Video Transformer Fine-tuning for Action Recognition | *PyTorch, Hugging Face, VideoMAE, Docker* GitHub

- **Fine-tuned** a **VideoMAE** video transformer to classify **21 stroke types** at **85.8% validation accuracy**; published the model to the Hugging Face Hub.
- Built the video preprocessing pipeline (frame extraction, normalization, augmentation) and a FastAPI + Docker inference service; analyzed errors with confusion matrices and per-class precision/recall.

AI Model Calibration Audit & jevcal CLI | *Python, Pandas, NumPy, Matplotlib* GitHub

- Audited **8,000** judgments from **TypeSafe's Jev**, a commercial AI decision model, against human labels and shipped an open-source calibration CLI that removed **96%** of calibration error.

Experience

Turing — United States - Remote
LLM Evaluation Engineer (Contract) — *March 2026 – June 2026*

- Generated **RLHF preference data** in Turing's post-training pipeline used to align **frontier LLMs**.
- Scored model responses on **math, logic and domain reasoning** against rubrics, feeding the **reward signal**.
- Analyzed **model failure modes** and edge cases, documenting rationales that drove training-data fixes.

CNTXT AI — United Arab Emirates - Remote
AI Data Quality Engineer — *June 2024 – Jan 2026*

- Evaluated **medical and legal LLM outputs** for accuracy and safety, flagging hallucinations and policy breaches.
- Validated large-scale **text, image and video datasets** for **LLM and computer vision** training.
- Identified **systematic failure patterns** and edge cases, improving dataset reliability for downstream models.

Skills

LLMs & Agents: LangChain, LangGraph, AI Agents, Tool Calling, RAG, Prompt Engineering, LLM APIs
LLM Evaluation & Data: Eval Harnesses, Guardrails, Hallucination Detection, RLHF & Preference Data, CI Gates
ML & Deep Learning: PyTorch, Hugging Face Transformers, Fine-tuning, Computer Vision, NLP, Recall@k, MRR
Engineering: Python, FastAPI, REST APIs, SQL (PostgreSQL, SQLite), Pandas, Docker, Git, GitHub Actions, pytest