ADNAN SHAMIM DAR

AI Researcher | Expert AI Reviewer and Evaluation Specialist (Medical, Multilingual) | Human-Data Quality | Research Founder | AI Systems Builder

Ancona, Italy (remote, EU time zone, flexible hours) | dar.adnan.54@gmail.com | +39 351 442 1130 | linkedin.com/in/adnan-shamim-dar-85664b230 | github.com/Adnan-shamim-dar | adnan-shamim-dar.github.io | 0009-0008-9711-0252

SUMMARY

AI researcher, fourth-year medical student (MD candidate, UNIVPM, Italy) and Trusted-tier expert reviewer on Mercor (top-performer incentive) with 11 months of continuous, paid work on frontier-model training data: 20+ AI training projects across 5 platforms (Mercor, micro1, Outlier/Scale AI, Labelbox/Alignerr, Ermis), covering medical, generalist, agentic computer-use, robotics video, voice-agent, function-calling, design-grading and Italian multilingual work. I review other experts' submissions against tolerance checklists, write and calibrate rubrics, red-team models, and turn quality judgments into scored, auditable decisions. I also founded and lead NeoResist-MD (computational oncology), and I build production-grade AI software: Tandem, an encrypted multi-device local LLM runtime (~18k lines, Windows/Android/iOS), and refusal-gate, a PubMed-grounded refusal layer for medical LLM answers. Target roles: reviewer lead, QA / quality manager, human-data project lead, medical domain-expert researcher, evaluation designer.

KEY FACTS

- 20+ AI training projects, 5 platforms, Dec 2025 to present; Mercor expert since Jul 2026 on two projects
- Qualified as Biology Subject Matter Expert (AI Content Specialist) on Alignerr (Labelbox), ID-verified, 2026
- micro1 Proton Robotics certification quizzes: 100% and 100%
- Mercor Odin Trusted reviewer, top-performer incentive (2x effective rate): 80+ reviews in first 3 weeks, under the 45 min handling target, about 6 in 10 submissions returned (strict bar)
- Identity verified and background check passed (Mercor)
- Clinical Medicine skill (Outlier medical-expert projects); 5 languages, IELTS 8.5; works on Italian-locale projects
- AI research: Project Genesis (artificial life) and NeoResist-MD (computational oncology); builder of Tandem and refusal-gate (public code on GitHub)

EXPERIENCE

Expert QA Reviewer | Mercor - Project Odin, Phase 3 (computer-use agent data) Oct 2026 - Present

- Contracted directly into the reviewer tier (identity verified, background check passed). Audit other experts' computer-use task submissions (screen recordings, narration, files, deliverables) before client delivery.
- Promoted to the Trusted reviewer tier and selected for Odin's top-performer incentive, which doubled my effective pay rate through performance bonuses.
- 80+ review units in the first three weeks, nearly all of them in a single week, consistently under the 45-minute handling-time target.
- Hold a strict bar: about 6 in 10 submissions returned with written, checklist-referenced feedback to writers.
- Score each task on a multi-item tolerance checklist where one failing item returns it: window-occlusion limits (>25% for >6 s), what/where/why narration on 70%+ of actions, workflow completeness, PII, metadata and audio quality.

Expert AI Output Grader - Slide Decks | Mercor - design-grading project (Feather platform) Jul 2026 - 2026

- Expert hourly contract (identity verified). Graded AI-generated presentations against written specs across six task types (aesthetic ranking, design instructions, style matching, template following, deck completion, content grading) on a forced-spread 1-7 scale, under 45-minute timed tasks with strict quality rules.
- Wrote per-guideline rationales: every design instruction marked met / not met with slide-level visual evidence, plus an overall score justification and stacked ranking.
- Built a 22,500-character personal grading manual (checklist built from the prompt before viewing, failures reported as ratios) and recalibrated after reviewer feedback.

Expert Annotator - Robotics Video Data (paid) | micro1 - Project Titan Proton (Proton Robotics) Sep 2026 - Oct 2026

- Annotated robotics task videos with interval-level tags (idleness and unproductive motion over 3 seconds, task progress) to produce training data for embodied / robotics AI.
- Scored 100% on both certification quizzes (Q1, Q2) of the project course; paid contributor via expert.micro1.ai.

AI Evaluator - Medical Expert, Generalist and Italian Multilingual | Outlier AI (Scale AI) - 8 projects Dec 2025 - Nov 2026

- White Coat (Clinical Medicine skill): wrote natural STEM / clinical prompts, then rated and ranked model responses (RLHF preference ranking).
- Almanac (Clinical Medicine skill): wrote research-based prompts with verifiable answers designed to break frontier models; checked responses for factual errors, unsafe advice and hallucinated claims.
- Galaxy Speech: stress-tested LLMs over long multi-turn conversations to capture and document failure patterns.
- Aether (Generalist + Italian/Multilingual team) and Multimango Multilingual: instruction-following and cross-domain evaluation, Italian-locale tasks, audio and video ASR test-set transcription; 45 consecutive days above 50% tracked activity (Hubstaff).
- Sixth Sense, Mime Orchard, Cribbage Tail: prompt and response evaluation under split attempter/reviewer workflows and quality-score-gated access.

Voice-Agent and Function-Calling Evaluator | Labelbox / Alignerr - AVA Front Desk, Human eval v2 Oct 2026

- Ran scripted voice calls against a clinic front-desk AI agent and labeled the ideal timestamp of each tool call (patient lookup, two-identifier identity verification, appointment search and booking).
- Red-teamed the agent with a safety-refusal probe and a system-prompt / JSON-leakage prompt-injection probe; corrected transcripts and classified outcomes.

Audio Transcription and Labeling | Ermis Transcription Aug 2026

- Completed training modules 1-4 and production transcription: spoken-form rules, backchannels, non-lexical and clip-level tags, overlapping speech, speaker-nativity metadata.

PROJECTS

Tandem - private local AI across your devices | Creator and sole engineer Nov 2026

- Runs GGUF models on owned hardware behind an authenticated OpenAI-compatible API (/v1/models, streaming /v1/chat/completions) via llama.cpp, with automatic GPU selection and CPU fallback. ~18,000 lines of TypeScript, Kotlin and Swift.
- Pool mode splits one model by layers across PCs, Macs, Android and iPhone so models larger than any single device can run; native Android (Kotlin) and iOS (Swift) workers.
- Security: six-digit one-use pairing, X25519 + HKDF-SHA256, mutual HMAC proofs, AES-256-GCM transport with replay protection; devices never open a port.
- Measured: 24 ms p50 time to first token and 50.4 tok/s decode on integrated AMD Vulkan; routing explains why every device won or lost. 47 Vitest tests, a 47-check end-to-end production suite, 11 Android JVM tests, streaming benchmark harness.

refusal-gate - no source, no claim (public, GitHub) | Creator Nov 2026

- Python gate that makes any LLM refuse a health question unless real PubMed papers that actually study it back the answer: fetches abstracts (NCBI E-utilities), keeps only those an LLM judge rates as relevant, refuses below a threshold, and rejects drafts that cite PMIDs outside the evidence.
- LLM-agnostic, CLI with JSON output and exit codes, hardened XML parsing (defusedxml).

NeoResist-MD / NeoVax - computational oncology (public, GitHub) | Founder and lead researcher 2026 - Present

- Context-aware neoantigen prioritisation programme run end to end: research question, data pipeline, evaluation protocol, internal audit and manuscript.
- Claim-tiered evidence system (discovery / held-out / support-only) with audit trails and case-level provenance; ran an editor-grade audit of my own manuscript and withdrew results that did not survive it.
- NeoVax Builder: mutation-to-peptide (8-11-mer) generation with MHCflurry presentation scoring, strict input validation, transparent ranking and run-level reproducibility metadata.

Autoresearch - refusal-gated discovery engine | Architect 2026 - Present

- Automated search over interpretable ranking formulas with a machine-enforced promotion gate: 27,118 evaluations across 9 cancer contexts, 0 claims promoted. Built to refuse weak signals rather than report them.

Project Genesis - artificial-life research | Principal investigator 2026 - Present

- Fitness-function-free simulation of single-cell agents under existential stakes; powered ablations and null-baseline controls, 455 tests. Framework paper in preparation.

MedViva - AI medical oral-exam simulator | Product designer and builder 2026

- Rubric-scored viva simulator: generates structured clinical questions, interrupts long answers like a real examiner, returns scored feedback; practice and exam modes, PDF ingestion.

VivaMed Access - AccessED Hacks 2026 entry (public, GitHub) | Builder Oct 2026

- Accessible, multilingual, low-data health explainers with source list and AI-assistance disclosure.

CGM glucose prediction models (public, GitHub) | Analyst Dec 2025

- R scripts and metrics on anonymised continuous-glucose-monitoring data; two example models trained in Orange with results and confusion matrices.

AI implementation in radiology - review (public, GitHub) | Author Dec 2025

- Structured review of the current status of AI implementation in clinical radiology.

Biomedical signal processing and lab work | UNIVPM coursework 2024 - 2026

- MATLAB ECG denoising: 6th-order Butterworth high-pass and low-pass filtering (filtfilt vs filter, zero-phase vs causal) of 200 Hz ECG with spectral comparison; artifact filtering, feature extraction; Fourier/Laplace modelling. Microbiology antibiotic-resistance profiling with structured lab documentation.

EDUCATION

MD Candidate, Medicine and Surgery (LM-41) | Universita Politecnica delle Marche (UNIVPM), Ancona, Italy 2023 - 2029 (expected)

- Year 4 of 6. IMAT admission rank 104th worldwide, 7th at the institution. ERDIS fully funded merit scholarship. GPA 26/30.
- Anatomy, physiology, pathology, microbiology and immunology, genetics, clinical semeiotics, biomedical physics, signal processing, statistics.

CERTIFICATIONS

- Mercor - Introduction to Agents (AI agent evaluation), Oct 2026. Verify: no. MRC-2026-H2D5H9QV
- UNIVPM - Health and Safety at Work, High-Risk Healthcare (12 h, D.Lgs. 81/08), Jul 2026. Certificate S4EFcGI9ia
- Google - Foundations of User Experience (UX) Design (Coursera), May 2023. Verify: coursera.org/verify/AV5GNKGK3N6E
- IELTS Academic - Band 8.5

SKILLS

- **AI evaluation and QA:** RLHF / preference ranking, rubric design and calibration, reviewer QA checklists, hallucination and factuality checking, medical safety review, red-teaming and prompt injection, function-calling and voice-agent evaluation, computer-use / agent trajectory review, multi-turn failure analysis, PII review, adversarial prompt writing, robotics video annotation (interval / temporal tagging), design and slide grading
- **Human-data operations:** guideline interpretation, written feedback to contributors, gold sets and audit sampling, inter-annotator agreement, quality thresholds and defect scoring, handling-time targets, evidence provenance
- **Medical:** clinical medicine (trainee), pathology, immunology, oncology research, HL7, SNOMED CT, ICD-10-CM
- **Technical:** Python, TypeScript / Node.js, React, Kotlin, Swift, R, MATLAB, llama.cpp, OpenAI-compatible APIs, LLM-as-judge pipelines, PubMed / NCBI E-utilities, Vitest, Git
- **Tools and platforms:** Mercor Studio, Feather, micro1, Outlier, Labelbox / Alignerr, Hubstaff, Persona ID verification, Google Workspace

LANGUAGES

English (IELTS 8.5) | Italian (B2, Italian-locale AI projects) | Arabic (fluent) | Hindi (fluent) | Urdu (fluent)