MUHAMMAD TAHA

Full-Stack AI Engineer

Islamabad, Pakistan | +92 317 5434059 | ch.tahaarif2005@gmail.com
taha-portfolio-website.vercel.app | github.com/mtahaarif | linkedin.com/in/muhammad-taha-21a163256/

PROFESSIONAL SUMMARY

Full-Stack AI Engineer and Computer Engineer specializing in end-to-end deployment, bridging advanced machine learning with scalable web architectures. Built edge-quantized vision and speech models running under 4 MB on Raspberry Pi 5, calibrated clinical prediction ensembles over 40GB+ of EHR data, and nine production Next.js platforms for US healthcare, home-care, and B2B clients. Adept at full-stack systems design across low-latency API pipelines, dual-database architectures, and biometric verification suites.

EDUCATION

National University of Sciences and Technology (NUST) Islamabad, Pakistan | 2022 – 2026
Bachelor of Computer Engineering
Relevant Coursework: Data Structures & Algorithms, Object Oriented Programming, Software Engineering, Machine Learning, Computer Vision, Digital Image Processing, Artificial Intelligence & Decision Support Systems

PROFESSIONAL EXPERIENCE

A&T Nexus Solutions LLC | *Full-Stack Engineer* Remote | Aug 2025 – Present

- **Client Delivery & Ownership:** Sole engineer across nine production Next.js platforms for US healthcare, home-care, IT-services, and B2B clients, owning requirements, architecture, deployment, and post-launch maintenance across 290+ commits. Two detailed under Projects Section.
- **Security & Accessibility Baseline:** Standardised hardened defaults across client sites, including strict Content-Security-Policy, X-Frame-Options, and MIME-sniff protection headers; honeypot fields and HTML-entity escaping on public forms; WCAG-compliant, reduced-motion-aware interfaces throughout.
- **Data & Integrations Layer:** Engineered JWT-secured admin APIs over Prisma/PostgreSQL and pooled MySQL backends with direct-to-Vercel-Blob upload brokering, plus third-party workflows spanning Resend, SMTP/Nodemailer, Google Sheets, WhatsApp, and Calendly.
- **Legacy Migration & Tooling:** Authored 15+ Node.js migration and asset-optimization scripts porting a WordPress practice site into Next.js (HTML extraction, image de-duplication, WebP conversion, orphaned-asset validation) over SSH-tunnelled MySQL; rebuilt three more legacy sites (GoDaddy, WordPress, Wix) into typed content modules.
- **B2B Digital Transformation Platform:** Built the firm's flagship React 19/TypeScript site with react-three-fiber 3D scenes and scroll-driven Framer Motion storytelling, backed by validated serverless contact and newsletter routes (Resend, FormSubmit).

TruID Technologies PVT. LTD. | *Computer Vision Engineer Intern* NSTP, Islamabad, Pakistan | Jul 2025 – Sep 2025

- **Biometric Liveness & Anti-Spoofing:** Engineered a cross-domain screen-replay detector via 2-stage transfer learning on a MobileNetV2 backbone across face, hand, and card presentations. Applied per-channel CLAHE and 3-pass unsharp masking over a stratified 17,834-image corpus, reaching **99.8% validation accuracy** after 45 epochs.
- **Identity Document Fraud Prevention:** Architected a Feature Pyramid Network with Convolutional Block Attention Modules on MobileNetV2, fusing a top-down pyramid into a 256-channel representation. Used LAB-space CLAHE (L-channel only, preserving color fidelity) plus an edge-sharpening kernel to surface splice artifacts, overcoming a 1:5 class imbalance for **93.8% test accuracy** at 0.97 precision on genuine documents.
- **Signature Forgery Verification:** Designed a two-stage deep metric learning verifier using custom hard-negative triplet mining and margin-based Triplet Loss, mapping signatures to L2-normalized 256-dimensional embeddings for a downstream classifier evaluated via EER and ROC-AUC, reaching **92.7% accuracy** across 5,000 pairs.
- **Model Optimization & Profiling:** Built inference pipelines across all 4 vision models on Tesla P100, profiling evaluation steps to a strict 31 ms or lower batch execution latency for scalable KYC deployment.

TECHNICAL SKILLS

Languages & Core Tools: Python, TypeScript, JavaScript, C++, SQL, MATLAB, Bash, Linux, Git
Deep Learning, LLMs & MLOps: PyTorch, TensorFlow, Keras, Hugging Face, LLMs, RAG, LightGBM, XGBoost, Optuna, SHAP
Vision, Audio & Edge AI: OpenCV, Librosa, Whisper, ResNet, MobileNet, 3D Gaussian Splatting, TFLite (INT8/NF4), Raspberry Pi 5
Frontend Web Architecture: Next.js 16 (App Router), React 19, TypeScript, Tailwind CSS, Framer Motion, Vite, WCAG Accessibility
Backend, APIs & Cloud: FastAPI, Flask, Node.js, REST APIs, Server-Sent Events, Prisma, PostgreSQL, MySQL, AWS EC2, Vercel

PROJECTS

Machine Learning and AI Systems

MedTraceAI: ED to ICU Clinical Deterioration Model | *Polars, XGBoost, Transformers, Optuna, SHAP*

- **Architecture:** Architected a dual-branch deterioration model forecasting ICU transfer 6 hours ahead, combining 192 engineered tabular features (NEWS2, shock index, MAP, BUN:creatinine ratio, threshold-crossing counts) with a custom Clinical Event Transformer over variable-length event sequences, fused by an isotonic-calibrated stacking ensemble.
- **Data Engineering:** Built an out-of-core Polars LazyFrame pipeline with predicate pushdown over 40GB+ of MIMIC-IV, streaming 37.1M chart rows into an OMOP-standardised event-triplet schema keyed on LOINC concept IDs, making the model hospital-agnostic with no site-specific item codes.

- **Optimization:** Ran a 300-trial Optuna XGBoost search with 5-fold out-of-fold predictions, training the Transformer branch with Focal Loss, label smoothing, and OneCycleLR scheduling; corrected an unstable meta-learner by retraining on population-aligned out-of-fold predictions.
- **Results & Explainability:** Achieved **0.892 AUROC, 0.667 PR-AUC, and 0.051 Brier score** on a held-out cohort of 19,248 ED visits, a +0.11 to +0.17 AUROC gain over published NEWS2 baselines. Integrated TreeExplainer SHAP for per-prediction top-5 feature audit trails.

SERENITY: Multimodal Mental Health AI System | *TensorFlow, PyTorch, TFLite, Raspberry Pi 5, FastAPI, React*

- **Edge AI & Perception Models:** Trained a ResNet-18 facial emotion classifier (93.0%) and a CNN-BiLSTM-Attention speech emotion network (80.6% across 7 classes) on a class-balanced 53,487-utterance corpus aggregated from 9 public datasets. Applied INT8/NF4 quantization to deploy a **3.86 MB TFLite** model on Raspberry Pi 5 at 27.5% peak CPU.
- **Hybrid Systems Architecture:** Architected a privacy-first edge-cloud pipeline executing Whisper STT and all biometric perception locally, streaming lightweight metadata to an EC2-hosted quantized LLM over Server-Sent Events via a pooled async client with sticky failover and circuit breaking, at 8–12s end-to-end latency.
- **Deterministic Clinical Safety Layer:** Built a retrieval-augmented reasoning engine over sentence-chunked therapeutic sources, gated by a rule-based router that computes risk scores and cognitive-distortion flags independently of the LLM and locks generation into one of four therapeutic frameworks via a backend-enforced state machine, removing the model from safety-critical decisions.
- **React Clinical Dashboard:** Delivered a six-page React 18 clinical interface with longitudinal PHQ-9/GAD-7/PCL-5 tracking (severity banding, 7-day worsening flags, inter-assessment velocity), automated SBAR handoff generation with deterministic fallback, and a C-SSRS triage and geolocated SOS escalation toolkit.

3D Environment Reconstruction | *OpenCV, PyTorch, gsplat, Open3D*

- **Classical CV Pipeline:** Architected a six-stage reconstruction pipeline over 49 calibrated views of the DTU Robot Image Dataset, using SIFT extraction, FLANN-RANSAC epipolar filtering, and two-view DLT triangulation to recover a 30,500-point sparse cloud.
- **MVS Densification:** Engineered plane-sweep Multi-View Stereo evaluating 96 depth hypotheses per pixel with illumination-invariant NCC photo-consistency scoring, then cleaned 166,303 raw points to 130,029 via observability masking, statistical outlier removal, and a 15 mm radius filter, rejecting 21.8% noise.
- **Neural Rendering:** Trained a 3D Gaussian Splatting model in PyTorch and gsplat over 15,000 iterations, injecting the MVS cloud as geometric initialization to bypass local minima, and ran adaptive clone/split/prune densification to **25 dB PSNR** at 0.04 combined L1+SSIM loss.

Full-Stack Software Engineering

Client work delivered at A&T Nexus Solutions LLC

Haines City Dental: Practice Platform & Headless CMS | *Next.js 14, Prisma, PostgreSQL, MySQL*

- **Architecture & Custom CMS:** Engineered a Next.js 14 site and bespoke admin CMS with HTML5 drag-and-drop reordering, protected by edge middleware, httpOnly JWT cookies, and inactivity auto-logout tracked across seven event types.
- **Dual-Database Backend:** Integrated Prisma/PostgreSQL for provider records alongside a lazily-initialised MySQL pool for legacy CMS content, with direct-to-Vercel-Blob upload brokering, cutting content update time **90%** from 30+ minutes to under 2.
- **Performance & Reliability:** Achieved sub-100kB initial loads via SSR and automated cache revalidation; built SMTP appointment workflows with mailto fallback and exponential-backoff client fetching.

Open Arms & Heart Services: Home-Care Agency Platform | *Next.js 16, React 19, TypeScript, Resend*

- **Architecture:** Architected nine static pages on typed content modules, with pinned horizontal scroll, sticky stacking cards, and an interactive care finder mapping 13 needs to a care level.
- **Contact & SEO:** Built a Resend contact route with honeypot, validation, HTML escaping, and a no-JavaScript 303 fallback; added MedicalBusiness and FAQPage JSON-LD, llms.txt, and a strict CSP.
- **Results:** Cleared every critical finding of an SEO audit opened at 37/100; mobile Lighthouse on Home rose from 57 to 87–91, blocking time fell from 865 to 57 ms.

SiteOptimizer: Search-Visibility Audit Platform | *Next.js 16, PostgreSQL, Playwright, Zod* *Personal project*

- **Audit Engine:** Built a breadth-first crawler with one parse pass per page feeding 332 declarative checks across 15 categories; fingerprinting of 35 platforms gates inapplicable checks and prints every skip reason.
- **Scoring Model:** Designed a severity-weighted score with exponential decay and PageRank-plus-traffic page importance; **0 of 2,000** fuzz trials lowered the score after a fix, versus 17% for Sitechecker's formula.
- **AI Visibility:** Added LLM content grading with Zod-validated output and per-URL answer-engine readiness analysis, including AI crawler access and llms.txt generation.

LICENSES & CERTIFICATIONS

Deep Learning Specialization — *DeepLearning.AI | Dec 2025*
Machine Learning Specialization — *DeepLearning.AI | Jul 2025*
AI for Everyone — *DeepLearning.AI | Jun 2024*
Introduction to Front-End Development — *Meta | Oct 2023*
CS50: Introduction to Programming with Python — *Harvard's CS50 | Sep 2023*