

Shayan Ahmed

AI Engineer | End-to-End AI Systems - Models, Applied Systems and the Data Behind Them

Email: shayankhattak250@gmail.com

LinkedIn: linkedin.com/in/shayankhattak250

Portfolio: shayankhattak.com

Location: Pakistan - open to relocation; remote to any timezone

Availability: Part-time and project-based

Professional Summary

AI Engineer who builds AI systems end to end and owns every stage of the pipeline: adapting and training models, evaluating honestly whether they hold up, putting them to work on real tasks with the guardrails they need to be safe in front of users, and building the data pipelines the models depend on in the first place. Currently architecting a neuro-symbolic beneficial-ownership determination engine for KYB and AML onboarding at a regulated cryptocurrency exchange, where grounded extraction feeds a deterministic reasoning engine rather than a model being trusted to decide. Experience also covers generative image models with LoRA adaptation and identity-preservation evaluation, AI security and guardrails, and training-data pipelines at scale for on-device models. Experienced with LLM applications, RAG, graph and rules-based reasoning, model evaluation, computer vision, data engineering and QA.

Core Skills

Model Training and Evaluation: LoRA and QLoRA adaptation, dataset preparation, conditioning, similarity and alignment scoring, labelled evaluation sets, blind comparison, adversarial and failure-case testing

Applied AI: Operational discovery, workflow analysis, human-in-the-loop design, prompt engineering, workflow orchestration, webhooks, REST API integration, production readiness

AI and LLM Engineering: LLM applications, Retrieval-Augmented Generation (RAG), vector search, hybrid retrieval, structured data extraction, OCR, document understanding

AI Security: Prompt injection defence, indirect prompt injection, RAG poisoning, tool abuse, jailbreak resistance, authentication and authorisation scoping, secrets management

Generative AI and Computer Vision: FLUX.1, SDXL, LoRA, QLoRA, IP-Adapter, PuLID, ControlNet, ArcFace, InsightFace, MediaPipe, CLIP evaluation, identity preservation, image dataset curation

Data for AI: Web scraping at scale, data collection, dataset curation, data cleaning, PII and secret filtering, deduplication, annotation, metadata design, knowledge graph extraction, training data pipelines

Reasoning and Compliance Systems: Neuro-symbolic architecture, LLM grounding and citation verification, graph algorithms, linear algebra (matrix inversion, spectral radius), interval arithmetic, bi-temporal data modelling, entity resolution, regulatory rules engines, AML, KYB, FATF, OFAC, EU AMLR, MiCA, threat modelling, property-based testing

Programming and Tools: Python, NumPy, Pydantic, TypeScript, Node.js, SQL, Playwright, Scrapy, Crawl4AI, Firecrawl, Docker, Redis, Git, S3-compatible storage, Label Studio, CVAT

Quality Engineering: Functional testing, end-to-end testing, regression testing, automated test coverage, defect reproduction, release validation

Professional Experience

Operations Specialist (Contract)

Origins Market Place & Technology, Dubai, UAE (remote)

Cross-functional scope in a startup: AI data pipelines, model training and evaluation, applied AI systems

Products: ZBX, GhostDrive, and an AI companion platform

July 2024 - Present

ORION-KYB - beneficial-ownership determination engine for KYB/AML onboarding, built for ZBX. Architect and lead engineer (2026 - in development)

- Architected a seven-stage KYB determination pipeline - ingestion, grounded extraction, verification, entity resolution, graph reasoning, screening, review - replacing a prior design in which an LLM generated per-document parsing scripts and executed them in microVMs.

- Audited that prior design and identified a structural defect rather than a tuning problem: an exact-sum validation rule requiring shareholdings to total 100% measurably incentivised the model to fabricate a shareholder. Replaced it with interval consistency checking.

- Implemented integrated-ownership propagation over cyclic shareholding networks using the Brioschi input-

output formulation, resolving circular structures that defeat chain multiplication; recovers a 91.81% effective holding from a 4.24% direct stake on the reference case. Singular matrices from closed cross-holdings are flagged as a laundering typology and routed to review rather than failed.

- Implemented interval-arithmetic propagation for banded register data, since the UK PSC register publishes 25-50% ranges rather than percentages. Proved entrywise monotonicity of the inverse via its Neumann series and enforced it in CI across 20,000 randomised networks.

- Designed a bi-temporal, append-only ownership graph separating valid time from transaction time, so the system can answer the question regulators ask after the fact: why was this change not acted on sooner?

- Implemented a versioned, effective-dated rules engine holding thresholds and jurisdiction lists as data, serving FinCEN CDD, EU AMLR 2024/1624, UK PSC banded control types, AMLR Article 58 trust roles, the OFAC 50% Rule and the BIS Affiliates Rule from a single computation.

- Produced the threat model covering untrusted code execution, prompt injection carried in adversary-supplied PDFs - detected by diffing the text layer against the rendered page - and adversarial evasion including threshold structuring at 24.9%, chain dilution and closed loops. The pipeline has no tools that act, so an injection can corrupt one extraction but cannot cause a side effect.

- Built the evaluation harness from open registers, using Companies House PSC daily snapshots and GLEIF Level 2, measuring precision and recall on determinations rather than schema-validation pass rate, with OCR field accuracy reported separately as the declared pipeline ceiling.

- Specified detect-then-branch ingestion, routing digital-text PDFs directly and scanned documents through OCR with page-level retrieval, after benchmarking showed direct PDF-to-vision-model extraction underperforms staged pipelines on long scanned financial documents.

ZBX - regulated cryptocurrency exchange and payments platform. Operations and applied AI (September 2025 - Present)

- Ran day-to-day exchange operations including deposit and withdrawal exception handling, stalled SEPA transfers, missing crypto memos and tags, virtual IBAN reconciliation and transaction investigation.

- Led KYB workflow discovery for corporate onboarding: mapped bottlenecks, analysed documentation requirements, and defined where automation could operate without touching the compliance decision.

- Designed the compliance dossier the determination engine produces for human review. The system recommends; it does not approve.

- Designed and implemented the guardrails on customer-facing AI, defending against prompt injection, unauthorised access and actions, and exposure of internal secrets or sensitive information. Tested these controls against misuse and adversarial scenarios to verify the system remained within its authorised scope.

- Led operational discovery on the support side: analysed real customer queries, identified recurring problems, mapped resolution paths, and scored workflows on volume, standardisation and risk to determine automation suitability.

- Collected and structured the operational knowledge base the system retrieves from, and wrote, tested and iterated its prompts against production failure cases.

- Contributed to the Phase 1 support agent now live in production, working with a Senior AI Engineer who owned the implementation; supported production readiness, deployment and ongoing improvement. Stack: LLM, RAG, vector and keyword retrieval, internal platform APIs, blockchain RPC.

AI companion platform - identity-preserving image generation. Generative AI and Computer Vision (2024 - 2025)

- Performed the human verification automated tooling could not settle: separating temporary attributes such as glasses, hats and accessories from permanent facial identity, rejecting blur, distortion and over-smoothed skin, and validating crops preserved facial structure across 1:1, 4:5 and 9:16 formats.

- Collected, curated, labelled and structured identity-anchor and scenario datasets with facial landmarks, head pose, embeddings and aesthetic scores as queryable metadata. Tools: InsightFace, ArcFace, MediaPipe, Dlib, OpenPose, MiDaS, Label Studio, CVAT.

- Contributed to the captioning convention separating permanent identity from conditional attributes, and to the dataset layout for anchors, scenario presets and control references.

- Contributed to generative model training and conditioning: LoRA adaptation over FLUX.1 and SDXL with pose and depth conditioning. Tools: Kohya_ss, Hugging Face Diffusers, ControlNet, IP-Adapter, PuLID.

- Contributed to model evaluation using ArcFace cosine similarity for identity retention, CLIP alignment for prompt fidelity, and blind A/B testing against a visual defect checklist.

GhostDrive - on-device agent operating system. AI Training Data (July 2024 - 2025)

- Collected, scraped and curated the source training corpus across four verticals: operational playbooks,

autonomous agent interaction traces, protocol and cryptographic specifications, and dense technical documentation. Selection optimised for knowledge density because target models run on-device. Tools: Python, Crawl4AI, Firecrawl, Playwright, Scrapy, HTTPX, asyncio.

- Organised and prepared data for the team processing pipeline and structured it for repeat ingestion rather than one-off delivery.

- Contributed data collection, curation and preparation feeding the team five-stage quality gate: boilerplate removal, PII and secret filtering, exact and near-duplicate detection, density scoring, and entity and relation extraction. Tools: Trafilatura, Presidio, SHA-256, MinHash LSH, GLiNER, spaCy, Airflow.

- Researched distribution and versioning approaches for a peer-to-peer knowledge network using content-addressed datasets. Tools: DVC, IPFS, Filecoin, MinIO and S3.

QA Engineer (Part-time, concurrent engagement)

CLiqsy - CRM platform, remote

July 2026 - Present

- Functional, end-to-end and regression testing across CLiqsy CRM workflows, including defect reproduction, edge-case coverage and release validation alongside the development team.

- Built automated regression coverage in Playwright.

Education

Bachelor of Science in Computer Science

COMSATS University Islamabad, Wah Campus

Completed 2023

Technical Foundations

Completed the full AI engineering curriculum. The areas below are studied coursework knowledge, listed separately from the Core Skills above because they have not yet been applied in production work.

Mathematics for machine learning: Calculus and gradients, probability and statistics, optimisation (linear algebra is listed under Core Skills, having been applied in production work)

Classical machine learning: Supervised and unsupervised learning, decision trees, gradient boosting, bias and variance, evaluation metrics

Deep learning: PyTorch, backpropagation, training loops, CNN and RNN architectures

Transformers and LLM internals: Tokenisation, embeddings, self-attention, encoder and decoder architectures, KV cache, quantisation

MLOps and production AI: Docker, CI/CD, model serving, autoscaling, cost and latency optimisation

GPU and distributed inference: Quantisation, batching, memory bandwidth, tensor and pipeline parallelism

Additional Information

Availability: Part-time and project-based only; currently engaged and not available full-time. Remote, working to any timezone by agreement; on-site possible for short periods where a project requires it.

One product name is withheld under confidentiality; details and references are available on request. Contributions are described precisely: work marked "contributed" was delivered as part of a team, not independently. No performance metrics or business outcomes are stated because none are public.