

ANBER AZIZ

AI Systems Engineer | Full-Stack Developer | ML Researcher
anberaziz6@gmail.com | (+92) 318 4636679 | Lahore, Pakistan
github.com/AnberAziz5 | linkedin.com/in/anber-aziz-70b028266

RESEARCH STATEMENT

I am a final-year Software Engineering student at Lahore College for Women University with a focused research interest in artificial intelligence, machine learning systems, neural network architectures, and intelligent software infrastructure. My independent work spans production-grade LLM security, unsupervised anomaly detection, edge-native inference pipelines, real-time RAG systems, and autonomous multi-agent orchestration. I am seeking a funded MS or PhD assistantship in AI, Machine Learning, Data Science, Intelligent Systems, or Computer Science at a US university, where I can build on this applied foundation through rigorous academic research. In parallel, I bring full-stack and AI engineering skills directly applicable to software product teams and freelance client engagements.

EDUCATION

Lahore College for Women University (LCWU)

Lahore, Pakistan

Bachelor of Science in Software Engineering

September 2022 Expected June 2026

- Relevant coursework: Data Structures and Algorithms, Artificial Intelligence, Machine Learning, Database Systems, Web Engineering, Software Design and Architecture, Computer Networks, and Operating Systems.
- Final year capstone: CyberGuard, an AI-powered cybersecurity threat intelligence platform with real-time risk analysis, role-based dashboards, incident lifecycle management, and automated threat prioritization.
- Active participant in university-level programming competitions, hackathons, and technology seminars.

RESEARCH INTERESTS

- Large Language Model reliability, hallucination mitigation, and guardrail engineering
- Autonomous multi-agent systems and agentic AI orchestration frameworks
- Retrieval-Augmented Generation (RAG) architectures and live knowledge grounding
- Unsupervised and semi-supervised anomaly detection for operational AI systems
- Edge-native and serverless machine learning inference pipelines
- Predictive analytics and ML-driven decision support in high-stakes domains (healthcare, logistics, infrastructure)
- Neural network applications in computer vision, natural language processing, and structured prediction
- Data engineering, feature pipelines, and scalable ML infrastructure design

RESEARCH PROJECTS AND TECHNICAL PORTFOLIO

CyberGuard AI Threat Intelligence System

2025 Present

Final Year Research Capstone | Python, Machine Learning, Role-Based Architecture

- Designed and developed a full-stack AI-powered cybersecurity intelligence platform as the senior capstone at LCWU, covering the complete lifecycle from threat ingestion to automated reporting.
- The platform delivers real-time threat monitoring, automated risk scoring using ML inference, incident lifecycle management, and role-based analytical dashboards tailored to different operator permission levels.
- A threat prioritization engine surfaces the highest-severity indicators automatically, reducing manual triage overhead for security analysts and enabling faster incident response.
- Integrates structured machine learning inference with a reporting module that generates actionable security briefs without requiring manual analyst input.

Omni-Node Operations Mesh Autonomous SRE Pipeline

2025

Distributed Microservices | Isolation Forest (scikit-learn) | Gemini 2.5 | MongoDB Atlas

- Built an event-driven microservice system that streams live server cluster telemetry (CPU, RAM, API latency) through a bidirectional WebSocket mesh across three independently deployed services.

- Applied a scikit-learn Isolation Forest model to perform unsupervised geometric outlier detection on incoming telemetry, replacing brittle static threshold alerting with statistically grounded anomaly scoring.
- When the ML engine returns a critical anomaly score, the system autonomously submits the telemetry context to Google Gemini 2.5 Flash, which diagnoses the root cause and generates exact Linux remediation commands within milliseconds.
- Every autonomous AI decision is stored in a MongoDB Atlas audit ledger with a strict schema, creating a permanent immutable trace for post-mortem analysis and compliance review.
- Deployed across a multi-cloud topology: Python ML engine and Node.js event broker on Hugging Face Spaces via Docker, and the React command center on Vercel.

Predictive Supply Chain Nervous System

2025

[XGBoost Classification](#) | [Folium Geospatial Mapping](#) | [Gemini LLM Triage](#) | [Streamlit](#)

- Developed an enterprise-grade logistics prediction platform that assigns a real-time delay probability score to each active shipment using an XGBoost classifier trained on synthetic historical supply chain telemetry.
- Feature engineering covers container weight, route distance, historical port congestion indexes, and weather pattern variables, giving the model multi-dimensional context for each risk assessment.
- High-risk shipments are plotted on an interactive Folium geographic map, clustering anomalous routes for immediate visual identification by operations managers.
- An integrated Gemini LLM triage agent automatically generates specific, actionable rerouting and buffering directives for flagged shipments, replacing the reactive crisis management process.
- Includes a full synthetic telemetry generation pipeline in Python to produce training datasets that simulate complex real-world supply chain variables.

OpenScholar Live Academic Research Engine

2025

[Dynamic RAG](#) | [arXiv API](#) | [Qdrant Cloud Vector DB](#) | [FastAPI](#) | [React](#) | [Vite](#)

- Built a serverless academic research engine that solves the information-lag problem inherent in static RAG systems by querying the live arXiv API in real time rather than relying on pre-indexed corpora.
- Downloaded papers are parsed, chunked, embedded on the fly using sentence transformers, and indexed into Qdrant Cloud with active payload indexing keyed to individual paper IDs for isolated per-session workspaces.
- The synthesis layer uses Groq as the primary high-throughput inference engine and automatically falls back to Google Gemini under rate-limit conditions, ensuring uninterrupted operation at zero cost.
- Frontend is a dual-pane React application built with Vite and Tailwind CSS v4, exposing an exploratory search tracking panel alongside a deeply synthesized literature review output.
- Backend FastAPI service containerized and deployed to Hugging Face Spaces; frontend deployed to Vercel with automated production rollouts via GitHub integration.

AI-Powered CI/CD Pull Request Reviewer

2025

[FastAPI](#) | [GitHub Webhooks](#) | [Google Gemini 2.5 Flash](#) | [Docker](#) | [Hugging Face](#)

- Designed a stateless webhook microservice that intercepts GitHub pull request events (opened and synchronize actions) in real time, fetches the full file diff, and submits it to Google Gemini 2.5 Flash for autonomous code review.
- The LLM is prompted to prioritize security vulnerabilities (hardcoded credentials, injection vectors, improper error handling), logic errors, and best practice violations before surfacing stylistic feedback.
- Review output is published as formatted Markdown directly to the GitHub pull request timeline, giving reviewers structured, actionable feedback without a manual first-pass inspection cycle.
- Fully containerized with Docker and deployed to Hugging Face Spaces as a persistent webhook listener operating continuously with zero downtime.

Enterprise AI Resume Checker with LLM Guardrail Gateway

2024

[Python](#) | [Flask](#) | [Groq Llama 3.3 70B](#) | [Pydantic Schema Validation](#) | [Vercel Serverless](#)

- Built a hardened full-stack microservice that semantically maps candidate qualifications against job description requirements using the Groq LPU inference engine running Llama 3.3 70B.
- Designed a three-layer LLM Guardrail Gateway: a regex-based prompt injection scanner that blocks adversarial instruction-override attacks with a 403 response; a PII redactor that strips all email addresses and phone numbers from resume text before transmission to protect candidate privacy under GDPR principles; and a Pydantic BaseModel schema validator that enforces deterministic JSON output structure regardless of model variation or incomplete responses.
- Deployed as a Vercel Serverless Function with automatic scaling and minimal cold start latency, serving both static frontend assets and the Python worker from a single vercel.json configuration.

Autonomous Multi-Agent Researcher

2024

[Multi-Agent Orchestration](#) | [Google Gemini](#) | [DuckDuckGo Search](#) | [Streamlit](#)

- Architected a sequential state-machine pipeline with three specialized LLM agents: a Planner that decomposes the user query into targeted search vectors; a Researcher that executes live web queries against the DuckDuckGo search index; and a Synthesizer that compiles the aggregated results into a fully cited Markdown intelligence report with an executive summary, thematic sections, and inline source references.
- Deliberately separates planning, retrieval, and synthesis across distinct agents to prevent context degradation and confabulation that arises when a single model handles all three tasks simultaneously.
- Implements rate-limit safeguards and graceful degradation logic so the pipeline continues producing partial results even when external web queries fail or return incomplete data.
- Hosted on Hugging Face Spaces as a Streamlit application with persistent containerized deployment at zero infrastructure cost.

SynthTox Engine Edge-Native Clinical Drug Interaction Triage

2024

[Cloudflare Workers](#) | [Groq LPU](#) | [Cloudflare Vectorize](#) | [D1 SQLite](#) | [Llama 3.1 and Llama 4](#)

- Designed a zero-hallucination clinical decision support pipeline deployed entirely on Cloudflare's global edge network, achieving sub-second end-to-end triage latency without a centralized backend or third-party cloud database round trips.
- A multimodal Llama-4-Scout model performs OCR on mobile device photographs to extract drug entity names. An edge-level output sanitizer applies regular expressions and word-count heuristics to strip conversational drift artifacts before they contaminate the retrieval pipeline.
- Drug entities are embedded using Cloudflare Workers AI (bge-base-en-v1.5, 768 dimensions) and matched against a Cloudflare Vectorize index of FDA drug interaction data. Matched vector IDs are resolved to clinical context text through an intra-network Cloudflare D1 SQLite database, keeping all retrieval traffic within Cloudflare's edge fabric.
- The final inference call runs at temperature 0.0 with a 100-token hard budget against Llama-3.1-8B, forcing the model to output exactly one categorical triage token: SAFE, CAUTION, SEVERE, or UNKNOWN, with no generative preamble permitted.

Feature Control Panel Full-Stack Feature Flag Management Platform

2024

[MERN Stack](#) | [MongoDB](#) | [Express.js](#) | [React](#) | [Node.js](#)

- Built an internal DevOps administration portal for managing runtime feature toggles without redeploying code, supporting a complete CRUD lifecycle (POST, GET, PATCH, DELETE) across a three-tier MERN architecture.
- Implements automatic token normalization that converts messy user input into system-safe kebab-case identifiers in real time, and maintains live telemetry counters for total flags, active flags, and offline flags.
- Character-by-character search bar and environment category filters (Development, Testing, Production) allow operations teams to isolate specific flags instantly during incident response scenarios.

HypeWear Storefront Zero-Dependency Vanilla JavaScript SPA

2024

[Vanilla JavaScript](#) | [CSS3](#) | [Zero External Dependencies](#) | [O\(n\) Filtration Engine](#)

- Engineered a full-featured e-commerce single-page application with no framework, library, or build toolchain, demonstrating complete mastery of native browser APIs, DOM orchestration, and JavaScript engine behavior.
- The multi-dimensional product filtration engine performs a single-pass O(n) reduction over three simultaneous attributes: search substring matching, price range boundaries, and category isolation. Hard boundary evaluation prevents substring collision (for example, a query for one gender category does not bleed into another).
- Cart state management uses a localized reactive data model that dynamically computes aggregate pricing, itemized quantities, and free-shipping threshold conditional rendering without any external state library or virtual DOM.
- CLS (Cumulative Layout Shift) is mitigated through strict CSS3 variable tokenization and object-fit container constraints, ensuring high-resolution product photography never triggers a DOM reflow at any viewport width.

Task Manager MERN Stack Application with JWT Authentication

2024

[MongoDB](#) | [Express.js](#) | [React](#) | [Node.js](#) | [JWT](#) | [bcrypt](#)

- Built a complete full-stack task management application with user registration, login, and persistent session management via JWT tokens, covering the full authentication and authorization flow from bcrypt password hashing to protected route middleware.
- REST API exposes seven structured endpoints across auth and task resources. All private routes are gated by a custom JWT verification middleware. Frontend supports task filtering by All, Pending, and Completed status.

Personal Portfolio Website Full-Stack MERN

2024

[React](#) | [Express.js](#) | [MongoDB](#) | [TypeScript](#) | [Responsive Design](#)

- Developed a full-stack personal portfolio with a React and TypeScript frontend, an Express.js API backend, and a MongoDB database seeded with academic course records, certifications, Al-Khidmat volunteer history, and project metadata.
- Implemented responsive design, scroll animations, SEO metadata optimization, and a type-checked codebase with zero TypeScript errors verified via tsc at build time.

PROFESSIONAL EXPERIENCE

Full-Stack and AI Developer Independent Projects

January 2026 Present

[Open-Source](#) | [Freelance Available](#) | [Remote](#)

- Actively designing, building, and shipping AI-integrated web applications and backend services, with a focus on production-grade LLM pipelines, RAG systems, ML-powered APIs, and full-stack MERN applications.
- Maintaining multiple public repositories covering autonomous agent orchestration, ML anomaly detection, edge-native inference, and developer tooling. All projects are documented with system architecture diagrams, deployment guides, and API specifications.
- Available for freelance engagements in web development, AI and ML integration, RAG pipeline construction, full-stack application development, and cloud deployment on Vercel, Hugging Face, and Cloudflare.

Web Development Intern NovaSole Pakistan

June 2024 September 2024

[Kasur, Pakistan](#) | [E-Commerce Platform Development](#)

- Developed and maintained frontend and backend components of a live production e-commerce platform, improving user experience, checkout flows, and platform reliability.
- Optimized website performance through image compression and browser-level caching strategies, achieving measurable reductions in page load time.
- Applied on-page SEO practices including structured metadata, semantic HTML markup, and canonical URL management, contributing to measurable growth in organic search traffic.
- Conducted systematic cross-browser testing and debugging to ensure consistent functionality across all major browser environments.

IT Infrastructure Intern MITE (Mubashar's Institute)

May 2024 June 2024

[Kasur, Pakistan](#) | [Network Administration and Technical Support](#)

- Supported the installation and ongoing maintenance of campus-wide network infrastructure serving more than 230 students and staff.
- Maintained CCTV surveillance systems and provided first-line technical support, minimizing downtime and keeping critical academic operations running.

TECHNICAL SKILLS

AI and Machine Learning: LLM integration (Gemini, Groq, Llama, GPT APIs), RAG pipeline design, multi-agent orchestration, XGBoost, scikit-learn (Isolation Forest, classification), neural network concepts (CNNs, RNNs, Transformers), prompt engineering, Pydantic schema validation, vector embeddings, semantic search

Data Science and Analysis: Python data stack (Pandas, NumPy, Matplotlib, Plotly, Folium), feature engineering, synthetic dataset generation, data cleaning and transformation, exploratory data analysis, Streamlit dashboard development

Programming Languages: Python (primary), JavaScript and TypeScript (ES6+), C++, SQL, HTML5, CSS3

Backend and API Engineering: Node.js, Express.js, FastAPI, Flask, REST API design, JWT authentication and authorization, WebSocket real-time communication (Socket.io), webhook integration, microservice architecture

Frontend Development: React.js, Vite, Tailwind CSS, Vanilla JavaScript SPA architecture, responsive web design, SEO optimization, progressive web app concepts

Databases and Storage: MongoDB (Atlas), MySQL, Mongoose ORM, Qdrant Cloud (vector database), Cloudflare D1 (SQLite), PostgreSQL basics

Infrastructure and DevOps: Docker (containerization and multi-service orchestration), Hugging Face Spaces, Vercel, Cloudflare Workers (edge serverless), Git, GitHub, Postman, Google Cloud basics, Linux command line

CERTIFICATIONS AND PROFESSIONAL DEVELOPMENT

- Google IT Support Professional Certificate Google via Coursera
- Data Analysis Using Python IBM via Coursera
- Data Visualization Using Python IBM via Coursera
- MERN Stack Web Development Online Professional Coursework

- Advanced JavaScript and Modern Frontend Engineering Online Professional Coursework

VOLUNTEER AND COMMUNITY ENGAGEMENT

- Al-Khidmat Foundation Pakistan: volunteer contributor supporting community welfare and social development programs in the Kasur region, with a focus on education access and community outreach.
- Active participant in university-level technology events, programming competitions, and AI/CS seminars organized at LCWU.

LANGUAGES

English: C1 Proficient (CEFR) **Urdu:** Native speaker

REFERENCES

Available upon request from faculty supervisors at LCWU and professional supervisors from internship placements at NovaSole Pakistan.