Akshit Gupta

f20231136@hyderabad.bits-pilani.ac.in | +91 99585 71609 | linkedin.com/in/akshit-gupta | github.com/akshit9162

Education

BITS Pilani, Hyderabad Campus CGPA: 8.12
B.E., Electrical and Electronics Engineering; Minor in Computing and Intelligence *2027*

Delhi Public School, Sushant Lok, Gurgaon 95.2% | 96.6%
Class XII | Class X (CBSE) *2023 | 2021*

Technical Skills

Languages: Python, C++, SQL, Java, C
Machine Learning & NLP: PyTorch, deep learning, Transformers, NLP, representation learning, LoRA, YOLO, RL (A2C)
LLM & Information Retrieval: RAG, hybrid (dense + BM25) retrieval, reranking, vector search, LlamaIndex, LangChain
Evaluation & Statistics: ROC-AUC, precision/F1, RAGAS, Kendall τ / Spearman ρ, ablations, drift under temporal splits
Data & Graphs: PostgreSQL, Apache AGE (graph database), knowledge graphs, NumPy, Pandas, SciPy, SQLite
Systems & MLOps: FastAPI, REST APIs, Docker, pytest, Redis, Kafka, ONNX Runtime, Terraform, AWS, CI/CD
Coursework: Machine Learning, Reinforcement Learning, Artificial Intelligence, Computer Vision, DSA, DBMS, OS, OOP

Experience

Upsure.AI May 2026 – Jul 2026
AI/ML Engineering Intern

- Architected an event-driven **FNOL claims pipeline** (**Kafka**, Kogito rules on an **Apache AGE graph**, FastAPI state machine) routing cases to **ML models**, re-engaging users via SSE and queueing edge cases for time-bounded **human review**.
- Built a vehicle-capture pipeline on **Vision Transformers** (SigLIP + DINOv2 + ONNX orientation head), cutting latency **67% (82s to 27s)**, plus a **YOLO11s + DINOv2**-LogReg damage-detection ensemble (**ROC-AUC 0.893, FPR 0.009**).
- Deployed a CPU-optimized **FastAPI** service for damage detection and **multilingual OCR** (**ONNX Runtime** + RapidOCR), raising **precision from 0.56 to 0.83** and cutting OCR latency from **∼14s to ∼4s**.
- Built an **autonomous LLM QA agent** (Claude/GPT, **structured outputs**) that writes Playwright, Postman and k6 test suites from a codebase and **self-heals**: failing tests go back to the LLM, which patches them or **classifies** them as real bugs.
- Integrated Gitleaks, Trivy and SonarQube into **CI** across multiple **GitLab** repositories, cutting manual testing time by **80%**; added a **code-fix agent** that **retrieves** relevant source files and tests LLM patches in isolated git worktrees.

Coffeee.io May 2025 – Jul 2025
Software Development Intern

- Built an **AI-powered resume parser** in **Python** with **spaCy** and **NLTK NER** to extract data from **PDF/DOCX** files.
- Added **regex** and **fuzzy matching**, batch processing, JSON/CSV export, **SQLite** storage and a **Streamlit** UI.

Projects

RL Video Summarization: Paper Reproduction | akshit9162.github.io/prlvs-reproduction Mar – Sep 2026

- Implemented a 2024 **reinforcement learning** paper from scratch in **PyTorch** (two **A2C actor-critic** policies on an **LSTM**), documenting **10 details** it leaves out and resolving **3 equations** unusable as printed.
- Built an **evaluation harness** (**Kendall** τ vs 20 annotators, **bootstrap CIs**, **Wilcoxon tests**) showing the result does not reproduce: $\tau \approx 0$ vs the published **0.080**, matching an independent reproduction.
- Traced the failure to the paper's reward (maximising it directly scores $\tau = 0.006$ vs **0.73** for a human-informed selector); a **5-fold cross-validated** importance model reaches τ **= 0.24** on unseen videos.

Abstain or Route: Multilingual Intent Routing | akshit9162.github.io/abstain-or-route Sep 2026

- Wrote **LoRA from scratch** (**pytest**-tested) for an **NLP Transformer** routing **60 intents** in **10 languages**: at **0.14%** of parameters it ties full fine-tuning on English data and is within **3 points** on multilingual data.
- English-only training drops Swahili to **24%** vs **87%** in English; **multilingual** training on the same data lifts it to **59%**.
- Measured **calibration** per language; sending the least-confident **20%** to humans lifts Hindi from **75% to 85%**.

Grounded RAG Platform | akshit9162.github.io/Grounded-RAG-Platform Aug 2026

- Built a **citation-grounded RAG** service (**FastAPI**, Claude) with **hybrid BM25 + dense retrieval** and **cross-encoder reranking**, so every answer cites its sources; **Dockerized** on AWS.
- Hit **1.0 context recall** and **0.947 answer relevancy** on **RAGAS**, with a **hallucination guardrail**.

Temporal EHR Intelligence Platform | akshit9162.github.io/Temporal-EHR-Intelligence-Platform Jan – Mar 2026

- Built a **Streamlit** ML app over **7 relational** health-record tables with Decision Tree, SVM and MLP **classifiers**.
- Measured **data drift** on a **pre/post-2020 temporal split** and found and removed a **label-leaking feature**.

Achievements

- **Codeforces Expert** (max 1606) • LeetCode 1627, 400+ problems solved • **NTSE** Scholar, State Rank 155, Haryana.