Kapish Yadav

github.com/Git-Kapish | linkedin.com/in/yadavkapish22 | yadavkapish22@gmail.com | +91-8851946409

Summary

Final-year ECE undergraduate at DTU (Class of 2027) with strong CS fundamentals and hands-on experience building and rigorously testing scalable distributed systems in C++17, Python, and TypeScript using agentic AI coding workflows (Claude Code, Codex, Copilot) with disciplined human review. Also builds LLM-integrated applications (LangChain, RAG, agents) and works across cloud-agnostic infrastructure (AWS, GCP).

Skills

Languages: Python, TypeScript, JavaScript, C++ (C++17), React

Systems & Backend: Distributed systems design, REST API design, FastAPI, Node.js

AI-Native Development: Agentic coding workflows (Claude Code, Codex), spec-driven builds, reviewing & validating AI-generated code

Testing & Reliability: Test strategy design, automated test suites (CTest, unit/functional), code coverage metrics, CI/CD (GitHub Actions), telemetry-driven reliability, incident/defect resolution, performance profiling & optimization, Docker, Git

Cloud: AWS, GCP

Databases & Messaging: PostgreSQL, MongoDB (relational & non-relational), Redis (protocol-level)

Machine Learning & AI: LLM API integration, agent architecture, prompt engineering, RAG, LangChain/LangGraph, vector databases, PyTorch, scikit-learn, XGBoost, algorithm design & performance optimization on large-scale datasets

Projects

MiniCache — Distributed In-Memory Cache with Capacity-Aware Sharding July 2026
C++17 · RESP2 Protocol · LRU/LFU Eviction · Consistent Hashing **GitHub**

- Architected a distributed, Redis-like in-memory cache from first principles in C++17– RESP2 wire protocol (compatible with redis-cli), sharded concurrent storage, and consistent hashing (FNV-1a, 100 virtual nodes/shard) to minimize key relocation on cluster resize.
- Directed the build via agentic coding workflows (Claude Code) against self-authored PRD/ARCHITECTURE/AGENTS.md specs, reviewing and validating all AI-generated code before merge.
- Built a 7-suite CTest automation framework (protocol, sharding, eviction, concurrency, persistence, replication) and engineered pluggable LRU/LFU eviction, crash-recoverable persistence (AOF, snapshotting), and leader-follower replication for fault tolerance.
- Benchmarked throughput via `redis-benchmark` across 4 shards: 9,711 SET ops/sec (p99: 18ms) and 11,517 GET ops/sec (p99: 7ms).

TabNet Studio — From-Scratch PyTorch Implementation & Research Workbench July 2026
Python · Pytorch · FastAPI · React/Typescript · Docker · XGBoost **GitHub | Live Demo**

- Implemented every component of "TabNet: Attentive Interpretable Tabular Learning" (Arik & Pfister, AAAI 2021) directly from the paper's equations in PyTorch Sparsemax attention, Ghost Batch Normalization, GLU feature transformers, sequential attentive decision steps, sparsity regularization, and feature attribution rather than wrapping an existing library.
- Built afull research workbench (FastAPI backend, React/TypeScript frontend) for configuring hyperparameters, training models with live loss/accuracy curves, visualizing step-by-step Sparsemax attention heatmaps across the forward pass, and running batch inference on uploaded datasets.
- Benchmarked the from-scratch model against Logistic Regression, Random Forest, and XGBoost baselines on accu racy, F1, training time, and inference latency, backed by a CI/CD pipeline (GitHub Actions) and Docker Compose orchestration for reproducible testing and deployment.

Work Experience

Software Engineer Intern | Globe Capacitors Pvt. Ltd. Jun 2025 – Jul 2025

- Designed and shipped backend automation connecting the company's ERP platform, relational databases, and Slack for automated status tracking, reporting, and notifications on AWS, cutting manual administrative effort by 40% across 3 operations teams.
- Diagnosed and resolved 30+ production defects by profiling and refactoring legacy backend code on GCP-hosted infrastructure, improving application reliability and cutting average load times by 25% for 100+ internal users.
- Automated previously manual approval routing and governance workflows by designing and shipping Python-based backend APIs, increasing workflow approval completion rates by 15%.

Education

Delhi Technological University 2023 – 2027
B.Tech in Electronics & Communication Engineering

Certifications

Software Engineering Job Simulation | JPMorgan Chase & Co., Forage Dec 2025
View Certificate

- Built an end-to-end backend service spanning Kafka-based message ingestion, Spring Data JPA transaction logic, SQL persistence, and a live external REST API via `RestTemplate`.
- Diagnosed and resolved cross-system failures across message ingestion, database operations, and external APIs using Maven test suites and debugger-driven inspection.