

PRANJAL SRIVASTAVA

Corpus Christi, TX | +1 346 375 2373 | pranj6004@gmail.com | [linkedin.com/in/pranj-srivastava07](https://www.linkedin.com/in/pranj-srivastava07) | pranjmax.github.io/pranj-portfolio/

SUMMARY

AI engineer who ships privacy-first, production-ready LLM experiences end to end. I build on-device RAG apps (WebLLM/WebGPU), client-side hallucination guardrails, and voice-AI workflows that prioritize trust, latency, and zero data leave-the-browser guarantees. Strengths include retrieval design (chunking, embeddings, Top-k tuning), prompt engineering and evaluation, and clean JS/TS frontends with CI/CD and test discipline. Comfortable owning architecture, data flows, and developer UX, with practical DevOps hygiene for repeatable demos, reliable releases, and recruiter-ready documentation.

CORE SKILLS

Languages & Frameworks: Python, Java, TypeScript/JavaScript, Node.js/Express, React, SQL
AI & Retrieval: RAG, embeddings (Transformers.js/MiniLM), cosine similarity, Top-k tuning, chunking strategies, guardrails
Model & Prompting: Prompt libraries, eval sets, safety/refusal patterns, grounding via citations, answer-diffing and red-teaming
Data & Storage: pdf.js ingestion, IndexedDB/localForage vectors, JSON feeds, basic analytics and event logs
Cloud & Infra: Static hosting (GitHub Pages), Vercel serverless APIs, ngrok, CORS, Docker, health checks
DevOps & Quality: Git/GitHub Actions CI, Swagger/OpenAPI, tests and regression checklists, Agile/Scrum
Security & Governance: RBAC, validation, audit logging, change control

EXPERIENCE

Tinker Tech Logix - Software Developer

Apr 2024 - Sep 2025

- Built and enhanced modular web apps; shipped features across admissions, fees, exams, and results modules.
- Designed and tuned SQL schemas/queries with caching for high-load screens to improve page responsiveness.
- Implemented RBAC with server-side validation and audit logging for traceable change control.
- Wrote unit/integration tests and added GitHub Actions CI; documented APIs with Swagger/OpenAPI.
- Containerized services with Docker; added health checks and improved error handling for sturdier releases.
- Collaborated via PRs and code reviews; supported staged rollouts and rollback plans.

ECS - Intern, Cloud Server & Data Management

May 2019 - Aug 2019

- Assisted cloud migration planning, stability hardening, and performance tuning of data platforms.
- Supported backup and recovery procedures and contributed to scale-out planning and data handling policies.

SELECT PROJECTS

Private Doc Chat - On-Device RAG (Browser-Only) | Live on GitHub Pages

- Pure client-side PDF Q&A;: local embeddings (Transformers.js MiniLM) + WebLLM (WebGPU) + cosine Top-k retrieval; zero backend, zero paid APIs.
- Sliding-window chunking with overlap plus similarity thresholding; strict quote-only answers with inline citations to reduce hallucinations.
- IndexedDB vector persistence with Clear Data, Service Worker offline shell, accessible keyboard-friendly UI; static CI/CD to GitHub Pages.

Hallucination Guard - Client-Side LLM Evidence Checker | Demo + Code

- Verifies each claim in an LLM answer against user-provided PDFs/text entirely in-browser via embeddings + retrieval pipeline (React + TS).
- Trust-first UX: Review -> Sources -> Report with clickable evidence and grounded fix drafts; designed for recruiter-ready clarity.
- Privacy-first and zero-ops: no uploads, no API keys; automated CI/CD for reproducible demos.

Vanessa - Voice AI Acquisitions Agent (Vapi + Node/Express) | Repo + Demo Packet

- Browser-dialer voice AI that detects seller intent quickly and logs structured leads to a live dashboard.
- Webhook pipeline (/vapi/events -> /api/leads) with deterministic qualification rules, live badges, and health/JSON feeds.
- Clear safeguards: polite branches (No/CallLater/DNC), runtime cap, and reliability fixes (CORS/ngrok); includes ops runbook.

EDUCATION

M.S., Computer Science - Texas A&M; University - Corpus Christi

2022 - 2023

B.Tech., Computer Science - SRM Institute of Science and Technology

2017 - 2021

CERTIFICATIONS

- micro1 - Certified Software Engineer (AI Interview), Sep 2025
- Complete ServiceNow Developer Course (Udemy)
- Programming in Python for Everyone (Coursera)
- ChatGPT Prompt Engineering for Developers
- Building Real-Time Video AI Applications (Nvidia)