Ansh Jetli

linkedin.com/in/anshjetli | a-jetli.github.io | ansh.jetli@gmail.com

Education

Rutgers University - New Brunswick — New Brunswick, NJ
Bachelor of Science in Computer Science · ***GPA: 3.4*** — *Expected Dec 2026*

Raritan Valley Community College — Branchburg, NJ
Associate of Science in Computer Science — *Dec 2024*

Relevant Coursework: *Data Structures, Computer Architecture & Assembly, Design and Analysis of Computer Algorithms, Principles of Programming Languages, Software Methodology, Systems Programming, Intro to AI*

Projects

RAG Pipeline over AI/ML Papers | *Python, LangGraph, FastAPI, ChromaDB, OpenAI, Pydantic* — May 2026

- Built a retrieval-augmented generation system over approx ∼500 arXiv papers
- Engineered hybrid retrieval fusing OpenAI semantic embeddings (ChromaDB) and BM25 keyword search via reciprocal rank fusion, then filtered candidates with a local cross-encoder reranker (FlashRank) for second-stage relevance.
- Designed a LangGraph stateful agent (planner, retriever, grader, reformulator, synthesizer) with Pydantic constrained structured outputs, supporting subquery decomposition and grader driven retrieval retries.
- Profiled end-to-end latency and cut worst-case query time 55% (80s → 35s) by replacing hosted reranker API with a local ONNX cross-encoder (67% of CPU request time) and tightening the agent's retry loop.
- Deployed as a FastAPI API on Hugging Face Spaces via Docker; pre-serialized the BM25 index so the server loads it at startup instead of rebuilding in memory.

LLM-Driven Text RPG Engine | *Python, OpenAI API, Pydantic, Pygame* — May 2026

- Built a text-based RPG where an LLM acts as a real-time dungeon master while a deterministic Python engine retains authoritative control over all game state, preventing model hallucination of player data.
- Enforced a typed JSON contract using Pydantic schemas with classified retry logic on parse failures, feeding malformed responses back to the model with correction prompts.
- Designed a three-tier memory system: authoritative engine state, a sliding hot-context window auto-compressed by a cheaper summarization model, and on-demand cold storage in per-region and per-NPC markdown files.
- Wrote a regression eval harness covering schema compliance, phantom state-change prevention, and prose-style rules, versioned by system-prompt hash to compare prompt iterations.

TailoRes | *Next.js, TypeScript, Supabase, Gemini AI, React-PDF, Tailwind CSS* — Oct 2025

- Collaborated in a three person hackathon team to design, develop, and deploy a full-stack app in 24 hours.
- Built an AI powered resume tailoring software that matches user experience to job descriptions using Gemini AI.
- Implemented automated skill extraction and a responsive interface with real-time editing and ATS ready PDF generation with React-PDF
- Engineered a validation layer that matches AI-generated output against original resume data using substring containment checks and regex filtering to detect fabricated metrics and hallucinated content
- Integrated Supabase for authentication and database-backed resume storage.

MicroCaml Interpreter & Compiler | *OCaml* — April 2025

- Built a recursive descent parser for an ambiguous grammar, resolving operator precedence and associativity conflicts through left factoring to support single token lookahead parsing.
- Implemented a Hindley-Milner type inference engine using constraint generation and unification to catch type errors before evaluation.

Technical Skills

Languages: Python, Java, SQL, OCaml
Frameworks & Libraries: Gemini API, OpenAI API, FastAPI, ChromaDB, LangGraph, Pandas, NumPy
Skills/Concepts: Git, GitHub, Object Oriented Programming, Data Structures, Algorithms, Functional Programming, Software Design, Computer Architecture, RAG