Kunj P. Shah

kunjcr2@gmail.com, (628)-529-6990
Problem-solver passionate about building scalable AI systems to automate complex real-world tasks
LinkedIn, Github, Portfolio, San Francisco, CA

SKILLS

AI/ML	Large Language Models, Transformers, Retrieval-Augmented Generation (RAG), LoRA, PyTorch, TensorFlow, LangChain, LangFlow, n8n, OpenCV, Machine Learning, Deep Learning, Natural Language Processing (NLP), Information Retrieval, Recommendation Algorithms
Web & Database	Node.js, Express.js, React.js, Flask, Tailwindcss
Database and Tools	MongoDB, MySQL, Git, Docker, VertexAI, Microsoft Azure

EDUCATION

San Francisco State University — San Francisco, California
B.S. in Computer Science

- **GPA**: 3.97/4.00, *Dean's List* — Expected Graduation May 2027

EXPERIENCE

Dreamable Inc. — **San Francisco, California**
AI Agent Intern — *June 2025 – Present*

- Contributed to **fine-tuning a 7B-parameter open-source LLM** for internal document Q&A tasks; handled dataset curation, low-rank adaptation (LoRA), and model evaluation, on company-specific **NLP** prompts for document retrieval.
- Developed an **AI-powered outreach assistant** using *n8n*, LangChain, and OpenAI tools to automate messaging workflows; currently used by **14+ interns** to scale weekly outreach with minimal manual effort.
- Built a **lead generation pipeline** that verifies and ranks potential clients by email validity and interest score using custom agents, improving lead quality and boosting response rate by **~2.3×**.

Dyna Grow Design Solution — **Ahmedabad, India**
Web Developer Intern — *May 2024 – Jan 2025*

- Designed and launched a **responsive marketing website** using Node.js, Express.js, and EJS, tailored for an architecture firm's client showcase and service catalog.
- Improved **website performance**, leading to a **2× increase in qualified client inquiries** within the first 2 months of deployment.

PROJECTS

Custom LLM - KsM (github)

- Built a custom **215M-parameter** GPT-style language model with **18** transformer blocks and a **512**-token context window, trained on 5 large-scale literary datasets (~500K+ tokens) using a self-implemented tokenizer, attention mechanism, and training loop, achieving over **85% accuracy** on internal benchmarks and **MAE loss of 1.8**.

theHelper - AI Research Assistant (github)

- Engineered a PDF analysis tool using PyPDF2, BERT/BART transformers, and FAISS for semantic search, packaged in a Streamlit app for real-time summarization and Q&A — reduced manual review time by **70%** across 50+ academic and business documents; **actively used by peers and family** for coursework and client work.

Max – Advanced Voice-Activated Assistant (github)

- Created a fully voice-controlled AI assistant integrating LangChain, OpenAI, Hugging Face, and SpeechRecognition for tasks like web search, email, scheduling, maps, and YouTube — achieving 90% voice command accuracy in real-world usage.

Additional projects available at: GITHUB.
Eligible to work in the U.S. – no visa sponsorship required