

Pratham Satani

pratham.satani@outlook.com | github.com/prathamsatani | [linkedin.com/in/prathamsatani](https://www.linkedin.com/in/prathamsatani) | +1 (617) 858-7620

EXPERIENCE

- Intern, Machine Learning Engineer | Softmax AI

Jan 2025 – Jun 2025

 - Designed complete **end-to-end ML workflows** from data ingestion through deployment, reducing adaptation time for new AI solutions by 40%.
 - Deployed **production-ready pipelines** on **Google Cloud Platform**, ensuring high availability, fault tolerance, and enterprise-level scalability.
 - Engineered automation systems powered by advanced **Computer Vision models**, streamlining employee verification and improving operational efficiency.
 - Collaborated with senior engineers to optimize AI infrastructure and workflows, enhancing overall reliability and reducing system latency.
- Intern, Software Engineer – Machine Learning | CasePoint LLC

Jan 2024 – Dec 2024

 - Developed machine learning solutions across **Computer Vision, NLP, and LLMs**, enabling process automation and improved business decision-making.
 - Deployed AI-driven systems on **GCP and AWS**, ensuring production-grade scalability, reliability, and cost-efficiency.
 - Optimized ML model workflows using **multimodal embeddings** and **clustering techniques**, increasing recommendation accuracy and personalization.
 - Partnered with **cross-functional engineering and product teams**, aligning solutions with stakeholder requirements and improving adoption.

PROJECTS

- AI-powered Attendance System | Softmax AI

 - Developed a facial recognition pipeline using **Python, OpenCV, and TensorFlow** to achieve >95% accuracy in real-time employee verification.
 - Implemented **geo-fencing with GPS integration**, preventing unauthorized remote check-ins and ensuring compliance with organizational policies.
 - Built a **cross-platform mobile app** using role-based design for employees, managers, and admins, improving accessibility and reporting efficiency.
 - Architected and deployed the backend on **FastAPI + MySQL with Google Cloud Platform**, ensuring secure logging, audit trails, and enterprise scalability.
- Image-to-Image Product Recommendation | CasePoint LLC

 - Designed a recommendation engine using **Python and Vertex AI** to generate personalized results from multimodal embeddings, improving user engagement.
 - Applied **fashionSigLIP embeddings with cosine similarity and clustering**, enhancing relevance of top-10 recommendations through multi-stage filtering.
 - Deployed **Vertex AI Vector Search** for low-latency retrieval across millions of embeddings, while integrating **AWS DynamoDB and S3** for data storage.
 - Containerized the pipeline with **FastAPI and Cloud Run**, ensuring scalable APIs and seamless integration with the e-commerce infrastructure.
- RAGBot: Retrieval-Augmented Chatbot | CasePoint LLC

 - Built a conversational pipeline using **LangChain, Python, and FastAPI**, enabling context-aware responses with Gemini 2.0 Flash.
 - Implemented **HNSW graphs with ANN search in ChromaDB**, reducing retrieval latency and improving accuracy of retrieved documents.
 - Developed ingestion pipelines for structured data with **APIs/Databases** and unstructured data with **PDF/HTML parsers**, expanding coverage of sources.
 - Created a **Next.js frontend with TypeScript and embeddable iframe design**, ensuring cross-platform compatibility and smooth user experience.

SKILLS

Technical Skills	Python (proficient), UNIX/Linux
Frameworks	TensorFlow, PyTorch, Keras, Scikit Learn, Django, FastAPI
Databases	MySQL, PostgreSQL, ChromaDB, DynamoDB, SQLite
Cloud Platforms	Google Cloud Platform, Amazon Web Services
Certifications	Affective Computing (IIT Ropar) – 87%, Social Networks (IIT Ropar) – 66%

EDUCATION

- Master of Science in Artificial Intelligence | Northeastern University | Boston, MA, USA

Sep 2025 – Dec 2027

 - Related Coursework: Data Processing for AI, Algorithms
- Bachelor of Technology, Information Technology | Gujarat Technological University | India

Sep 2021 – Jun 2025

 - Related Coursework: Python Programming, Machine Learning, Artificial Intelligence, Linear Algebra and Calculus