

PANKAJ JARIAL

+91-8626977490 ◊ Chandigarh, India

pankajjarial.job@gmail.com ◊ [linkedin.com/pankiae](https://www.linkedin.com/pankiae) ◊ github.com/pankiae

OBJECTIVE

Generative AI Engineer with **2.5+** years of experience building **LLM-driven and multimodal** AI systems using **Python, FastAPI, and PyTorch**. Skilled in **RAG, LangChain, vector databases, and AI agent architectures** for scalable, production-grade applications.

SKILLS

Programming & CS Fundamentals	• Python • Object-Oriented Programming (OOP) • Dynamic Programming • Familiar with Design Patterns & System Design • Networking Basics
Backend Engineering	• FastAPI • Django REST Framework • Flask • PostgreSQL • MongoDB • Redis • Celery
Generative AI	• Git • AWS • Docker • Supervisor • API integration LLMs • RAG • Multi-Model AI Agents • Tool-Use, Memory & Delegation • NLP • Vector Databases • prompt engineering • embedding models • MLOps • LangChain • Agent-to-Agent Communication

APPLIED AI RESEARCH PROJECTS

Image Seg AI. Developed ImageSegAI (Python, PyTorch, FastAPI) leveraging Meta's SAM2 for zero-shot object segmentation, enabling precise masking and background removal. Integrated OpenAI GPT-image-1 for AI-powered inpainting.

Context & Control (C2) Core. Monitors token usage in chat/research apps; when thresholds (e.g. 5 K tokens) are exceeded, automatically fragments the conversation. A custom 'delegation' agent analyzes new queries, retrieves only the most relevant embedded fragments, and passes them (plus the fresh prompt) to the LLM, significantly reducing token overhead.

EXPERIENCE

AI/ML Engineer Associate - level 3

smartData Enterprises Inc.

Sept 2025 - Present
Mohali, Punjab

- Working on the traditional ML and AI Regulation project. Build the tool in the Computer Vision for detecting the objects in the image on the basis of the prompt (text input).

Gen AI Python Engineer

Tricky WebSolutions Pvt. Ltd.

Aug 2024 - Sept 2025
Mohali, Punjab

- Deployed AI-agent workflows from scratch, using Python function-calling patterns and eval() framework rather than turnkey agent frameworks, to keep dependencies light, simplify debugging, and make future customization easier.

Research Intern

Tricky WebSolutions Pvt. Ltd.

May 2023 - July 2024
Mohali, Punjab

- Got Exposure in the Gen AI tools, RAG, vector database. Contributed in the legacy project while maintaining the packages and make it suitable for the modular architecture.

PROJECTS

Curico. Multi-Model Agentic work flow, which works in loop for generating the steps for activity, questions & answers and Image. Whisper + TTS voice tutor guides children, when they gets stuck. Celery+Redis pipeline off-loads image generation, embeddings, and moderation; retry logic smooths traffic spikes. ([Try it here](#))

Invisioned. Integrated multiple LLMs such as OpenAI, Stability AI, Claude, and Perplexity AI, along with web URL citation results and computer vision models like GroundingDINO and SAM2. Delivered a cost-effective solution with a total expense of only \$10/month for all the models\$ currently used by over 6,000 users across both the Play Store and the App Store. ([Try it here](#))

EDUCATION

Bachelor of Computer Applications, H.P. University

Relevant Coursework: Python, Responsive Web Design, Networking Fundamentals.

2019 - 2022