

Jaspreet Singh Nahal

Delhi, India

jaspreetnahal100@gmail.com

+91-9999604054

[LinkedIn](#) – [GitHub](#) – [Portfolio Website](#)

SUMMARY

Computer Science Engineering student specializing in AI and Machine Learning, with hands-on experience building LLM-based systems, full-stack applications, and scalable backend services. Developed RAG-powered chatbots, algorithmic trading platforms, and automation tools using Python, React.js, Flask, and modern APIs. Experienced in hackathon-driven development, clean system design, and performance optimization. Solved 300+ LeetCode problems and actively seeking Software Engineering and AI/ML Engineering roles.

TECHNICAL SKILLS

Programming Languages: Python, Java, C++, JavaScript

Frameworks & Libraries: React.js, Node.js, Express.js, Flask, FastAPI

Machine Learning & NLP: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, LangChain, Retrieval-Augmented Generation (RAG), LLM Fine-Tuning

Databases & Vector Stores: PostgreSQL, MySQL, MongoDB, SQL, FAISS

Cloud, DevOps & Tools: Docker, CI/CD (GitHub Actions), Streamlit, Git, LeetCode

EXPERIENCE

Machine Learning Intern – Legal Chatbot with RAG

August 2025 – October 2025

Expedien eSolutions

Noida, India

- Worked with a team to build a domain-specific RAG-based chatbot using LangChain, Hugging Face Transformers, and FAISS.
- Preprocessing legal documents, chunking with LangChain, and generating embeddings with sentence-transformer models.
- Implemented a modular retrieval-to-generation pipeline using RetrievalQA for accurate legal question answering.
- Developed a demo UI in Streamlit and evaluated chatbot on accuracy, relevance, and latency.

Algorithmic Trading Developer

November 2024 – December 2024

Project Experience/Freelance

Noida, India

- Enforced an automated trading system implementing the SMA (Simple Moving Average) crossover strategy
- Designed an intuitive dashboard to monitor trade signals and visual market trends, enabling better user decision-making
- Enhanced backend performance using Flask and PostgreSQL, resulting in a 30% reduction in trade latency
- Scaled the system to handle increased trading volume with optimized data handling and async processes

PROJECTS

Context-Grounded GPT–RAG System | GPT, PyTorch, Transformers, FAISS, NLP, RAG

February 2026

- Built a transformer-based GPT language model from scratch in PyTorch, implementing causal self-attention, MLP blocks, custom weight initialization, and GPT-2 tokenizer integration; pretrained the model on 1B tokens from FineWeb-Edu using shard-based tokenization, memory-efficient storage, and gradient-clipped training.
- Implemented a Retrieval-Augmented Generation (RAG) pipeline using Sentence Transformers and FAISS for context-grounded, low-hallucination question answering, achieving sub-200ms retrieval latency, and designed a confidence-based gating mechanism using embedding distance thresholds.
- Fine-tuned the model for instruction following using selective loss masking to train only on assistant response tokens, improving factual accuracy and chat reliability.
- Developed an interactive chat system with hybrid GPT/RAG inference, top-k sampling, repetition penalties, heuristic-based repetition prevention, and context quality filtering.

Resume Skill Extractor | Python, REST-API Development, NLP, PDF Parsing, React

June 2025

- Engineered an AI-powered skill extraction tool leveraging NLP pipelines (spaCy, regex, TF-IDF) for entity recognition from resumes.
- Built REST APIs with FastAPI, extracting resume skills with 85%+ accuracy and processing each file in under 2 seconds.
- Created a named entity mapping model to identify tech vs non-tech keywords from unstructured text.
- Deployed the backend model and UI on a containerized dev environment using GitHub Actions.

SMA Crossover Trading System | Python, Flask, PostgreSQL, Data Visualization

December 2024

- Developed an automated trading pipeline using SMA crossover and moving average convergence strategies.
- Used Python and scikit-learn to analyze historical data patterns and validate signal consistency.
- Engineered a backtesting environment to simulate real-world trades and improve model reliability.
- Connected the strategy to live data via REST APIs and optimized latency-sensitive Flask endpoints.

AI Document Reader and Invoice Matching System | Python, Flask, AI Models

October 2024

- Built an AI-powered document parser using OCR and NLP to classify and match invoices based on text similarity.
- Applied cosine similarity and keyword extraction for accurate field-to-field invoice matching.
- Achieved 85% match precision across 300+ invoice pairs using sentence embeddings and pattern matching.
- Developed a Flask-based REST API to automate document ingestion and classification workflows.

AI Chatbot | React.js, Python, Flask, RESTful API

September 2024

- Built an NLP chatbot with 92% intent recognition accuracy and 1s response time, handling 1,000+ user sessions.
- Integrated Named Entity Recognition (NER) and intent classification for personalized response generation.
- Achieved 92% intent recognition accuracy using custom-trained models and real-time Flask API deployment.
- Used MongoDB for dynamic session storage and conversation history handling.

EDUCATION

Dr. A.P.J. Abdul Kalam Technical University

Uttar Pradesh

Bachelor of Technology in Computer Science Engineering (Artificial Intelligence)

2023 - 2027

CERTIFICATIONS ACHIEVEMENTS

Google: Prompt Design in Vertex AI, Responsible AI: Applying AI Principles with Google Cloud
IBM: Accelerating Deep Learning with GPUs
GUVI: GUVI Hackathon - GUVI Geek Networks, IITM Research Park
BCG X (via Forage): GenAI Job Simulation – AI-powered Financial Chatbot Development

HACKATHONS

AI & Automation Unpacked Hackathon – IBM

July 2025

- Built an AI-driven proof-of-concept solution using IBM Granite to demonstrate innovative business applications of AI
- Gained hands-on experience with foundational AI concepts and automation techniques to promote economic growth and decent work

NationBuilding Hackathon – Unstop

April 2025

- Architected an AI-based Restaurant Manager chatbot to automate bookings, staff queries, and feedback loops using Flask and MongoDB

Google Solutions Challenge – Google

March 2025

- Collaborated on a real-world solution addressing local issues using Google technologies. Focused on scalable architecture and user-centric design

GUVI Hackathon – IITM Research Park

February 2025

- Implemented an NLP-based Resume Skill Extractor to identify key candidate strengths from resumes for recruiters