

AMIT GUPTA

Senior AI Engineer | AI/ML Development | Python Expertise | Generative AI Specialist

+919060211542

@ amitinger25@gmail.com

<https://www.linkedin.com/in/amit-kumar-gupta-5005361a1/>

Pune, India



SUMMARY

AI Engineering professional with overall 7+ years of experience and with an M.Sc. degree in Mechatronics, expert in Python programming, large language models (LLMs), Gen AI and Agentic AI. Key achievements include deploying over 6 AI-driven solutions across various enterprise settings, achieving a 40% increase in operational efficiency, and integrating advanced AI stacks for Fortune 500 companies. Seeking a Senior AI Engineer position, where my skills in AI/ML development and data science will support your mission of enhancing investment processes to be more efficient, scalable, and insight-driven.

EXPERIENCE

ML/AI Specialist

UBS Asset Management (Via. Accenture Solutions Pvt. Ltd.)

11/2024 - Present Pune, India

- Increased customer engagement and reduced query time 80% by building a RAG-based Azure OpenAI-powered chatbot using Langchain and Azure Functions.
- Built a LLM response evaluator using metrics such as Ranking, Relevancy and confusion matrix.
- Built multi-agent chatbot using Langgraph, Azure, CosmosDB and CRM. We have covered 14 usecases some of them are Sales recommendation, Feedback manager and topic recognition.
- Worked on Orchestrator, Agents, tools framework for multi-agentic AI.
- Deployment using GitLab CI/CD pipeline and Docker images.

Senior AI Engineer

Turabit Solutions Private Limited

05/2024 - 10/2024 Remote

- Improved chatbot efficiency by 15% by enhancing detection of greeting words.
- Finetuned open source LLM like Mistral and Llama 7B model using PEFT, QLoRA technique using PyTorch and Huggingface.
- Built a RAG-based tools enabled AI agent for smooth onboarding of new joiners using Langgraph, Langchain, Pydantic.

Product Engineer

V-Labs Pvt. Ltd

02/2023 - 04/2024 Bangalore, India

- Implemented AI solution replacing legacy system, boosting performance from 35% to 91%.
- Applied deep learning and machine learning models, including RNN, LSTM, Transformers, BERT, and XGBoost, for advanced analytics.
- Experimented with Generative AI for text extraction using Llama 2, enhancing processing capabilities.
- Built a sentiment analysis pipeline for company's 10-K annual reports, improving financial insights using balance sheets.
- Conducted comprehensive statistical analysis for model validation, using descriptive and inferential methods.
- Led risk analysis and loan assessment projects to determine financial health of companies.
- Managed CI/CD for code version control using Azure DevOps, ensuring streamlined development processes.

Research Engineer

Ciena and ENSUN(Full-time roles during M.Sc.)

03/2022 - 01/2023 India

- Forecasted telecom network outages & developed recommendation systems (Ciena)
- Designed anomaly detection using PCA, CI-based methods, & clustering (KMeans, DBSCAN)
- Built & deployed multi-class text classification system with ETL pipeline (ENSUN)

Programmer Analyst

Cognizant Technology Solutions

11/2017 - 08/2019 India

- Boosted client sales by 15% by designing PoC recommendation system using order history & time-series analysis
- Conducted feature engineering using PCA, reducing features by 25% while maintaining accuracy
- Performed hypothesis testing (t-test, z-test) to validate assumptions
- Built foundational expertise in backend development and agile collaboration

KEY ACHIEVEMENTS



Chatbot Accuracy Improvement

Enhanced chatbot accuracy by 85% using RAG, boosting efficiency significantly in ML operations.



AI Solution Optimization

Enhanced AI solution accuracy from 35% to 91%, revolutionizing data handling processes.



Predictive Model Development

Achieved 94% accuracy in predictive modeling, elevating chatbot precision by 30%.



Feature Reduction Success

Reduced feature size by 25% via PCA in engineering tasks, enhancing data processing efficiency.

LANGUAGES

English

Proficient



Hindi

Native



German

Beginner



SKILLS

Agile	AWS	Azure	CosmosDB	Django
Docker	Fast API	Flask	Git	K-Means
MongoDB	Matplotlib	MLOps	NLP	
NumPy	Pandas	Plotly	PostgreSQL	
PyCharm	PySpark	Python	PyTorch	
Scikit-learn	Seaborn	SpaCy	SQL	
MCP	TensorFlow	GCP		

FIND ME ONLINE



LINKEDIN

<https://www.linkedin.com/in/amit-kumar-gupta-5005361a1/>

INTERNSHIPS AND PROJECTS

SaGer: Small vendor CRM.

08/2025 - Present Pune, India

Building a CRM for small street vendors to manage their sales, Sales prediction and community marketing platform. Deployed on GCP using automated CI/CD pipeline from GitHub.

- mysanvi.in

EDUCATION

M.Sc - Mechatronics

Universität Siegen

10/2019 - 01/2023 Siegen, Germany

- Thesis: Non-Rigid Puzzles without template
- Project: Explainable Pneumonia Detection using CNN (Image-based AI)

B.Tech

KIIT University

07/2013 - 05/2017 Bhubaneshwar, India

REFERENCES

Hasan Abu-Rasheed - Research Assistant, University of Siegen.

<http://linkedin.com/in/hasan-abu-rasheed>

Benjamin Hill - CEO, ENSUN GmbH

<https://drive.google.com/file/d/1RygY3umEKqHts3xfH4UL9LPJ5QqT9CwO/view?usp=sharing>

CERTIFICATIONS

Microsoft Azure AI-102: AI Engineer Associate

Microsoft

Generative AI - DeepLearning.AI

DeepLearning.AI | Accenture

MLOps

Udemy

AWS ML

Coursera | AWS

INTERNSHIPS AND PROJECTS

Automated-Portfolio maker

05/2025 - Present Pune, India

Building a portfolio maker using Django, Python, Langgraph, Node.js and React. This enables uploading of documents such as CV, profile pic, certificates, references.

- Under 10 seconds portfolio is generated. AI agent fetches top 10 relevant latest jobs posted and also answers the query about profile to the hirer.

Master Thesis: Non-Rigid puzzles

09/2022 - 01/2023 Siegen, Germany

Focus on finding the correspondences of the non-rigid deformations of the shapes without any template. Descriptors such as HKS, SHOT, Geodesic distance, Gaussian curvature, eigenvalues have been used.

- The MSE was jumped from 2.61 to 2.31.

INNCEPTION(Internship)

11/2021 - 03/2022 Remote

Gig work prediction for US seekers

- Used BERT based embedding to find the most relevant job for the people query.
- Used GPT to guide them how to proceed with the Gig.

Research internship(Studentenarbeit)-Universität Siegen

10/2020 - 10/2021 Siegen, Germany

Project: Pneumonia classification from x-ray images

- Got data from a Kaggle project with 4600 imbalanced data.
- Processed data by rescaling, balancing, PCA
- Used XGBoost, clustering and ResNet50 convolutional model.

iNeuron.ai(Internship)

05/2021 - 08/2021 Remote

Diabetes classification

- Data: Got data from a medical research team with 11 attributes.
- Data processing: Cleaning, Embedding, PCA, Clustering
- Models used: XGBoost for binary classification