

Abhinav Kumar

AI Engineer

Location: Milton Keynes, UK **Phone:** +447867238292 **GitHub:** [www.github.com/abhinav61](https://github.com/abhinav61)

Email: abhi292394@gmail.com **LinkedIn:** <https://www.linkedin.com/in/abhinavkumar2910/>

Professional Summary

Results-driven Engineer with over 5 years of experience designing, developing, and delivering high-quality, end-to-end software, ML models and AI solutions using deep learning in production environments. Adept at building scalable, maintainable, and performance-optimized code using industry-aligned languages, frameworks, and tools. Demonstrated expertise in machine learning, deep learning, MLOps, and cloud platforms (AWS, GCP, Azure), with successful deployments of intelligent automation and GenAI systems (GPT-3/4, LLAMA3, TTS, DALL-E 3, embedded models) across financial, e-commerce, and space tech domains. Skilled in vector search, prompt engineering, workflow orchestration, data analysis and statistical analysis. Proven ability to collaborate cross-functionally with product managers, designers, and engineers to align software solutions with business objectives. Continuously stays abreast of the latest advancements and trends in AI, continuously exploring and evaluating new technologies and approaches. Passionate about innovation, continuous learning, and building intelligent, real-world systems that create measurable business value.

Education

Master of Science – STEM: Management and Information Systems
Cranfield University - Cranfield, UK - 2:1

Sep 2022 – Aug 2023

- **Course:** Data Analyst with AI, Enterprise Management, Enterprise systems, Digital Engineering, Supply Chain, Operations Management, Software Engineering
- **Thesis:** Digitalisation and Automation of Space Engineering Task – Geometry Construction for Mechanical Analysis using PythonOCC and Artificial intelligence
- **Group Project:** Thermography based Inspections – OpenCV

Aspurian, Milton Keynes, UK

May 2024 – Present

Full Stack Developer – AI & Data

- Design, develop, and implement end-to-end AI pipelines, including data acquisition, pre-processing, feature engineering, model training, evaluation, and deployment.
- Contributed to the architecture, design patterns, reliability and scaling of existing ecommerce application
- Built end-to-end react applications deploying ARIMA based ML model, integrating **GPT-4, DALL-E 3, and TTS** for real-time audio, text, images and avatar-based inference for AI based Finance Reporting Tool
- Collaborated with client stakeholders and internal teams to understand business requirements and translate them into robust AI solutions.
- Processed unstructured PDF data, developed a **machine learning model**, and integrated it with an LLM using vector embeddings and ranking algorithms to enhance information retrieval, content generation and relevance.
- ETL data from multiple sources and applied **statistical methods for forecasting graph-based data** to support dynamic insights and automation in the intelligent reporting platform.
- Integrated **Stripe payment** API into the Reporting Tool.
- Developed REST APIs and automated workflows supporting model deployment and lifecycle management.
- Led R&D and Implemented RAG based intelligent chatbot integrating **Llama3 and nomic-embed-text models** for engineering support.
- Designed a scalable dashboard system integrating Upsales APIs, **saving 20% in operational costs**

Nitrexo, Cork, Ireland (Remote)

May 2023 – Sept 2023

Research Intern - Data & AI/ML

- Took ownership and developed end-to-end ML pipelines for automating **Space Engineering Task** based geometry construction in aerospace engineering.
- Read, interpret, and implement emerging ML and AI techniques in large datasets

- Contributed to the development of internal AI assets, frameworks, and best practices to enhance team efficiency and delivery excellence.
- Integrated PythonOCC with AI LLM to optimize mechanical analysis workflows.
- Deployed ML models to AWS with CI/CD practices, **reducing time-to-deployment by 60%.**
- Designed and executed A/B testing experiments to compare model performance and evaluate the impact of data-driven features on user engagement and system efficiency
- Collaborated with stakeholders to translate business challenges into AI-powered prototypes

Altimetrik (Client: PayPal), Bengaluru, India

Oct 2021 – Sept 2022

Senior Machine Learning Engineer/Data Scientist

- Built and maintained ML production model for **fraud detection application**, reducing fraud rates from **12% to 5%.**
- Translated analytics insights into scalable backend systems using Django and Python libraries
- Integrated and managed MLOps using Docker, Kubernetes, and Git workflows
- Created modular APIs and improved system throughput with multithreaded model serving.
- Enhanced team collaboration by mentoring junior developers with analyzing and writing program specifications, leading to a 40% reduction in project delays and a 25% improvement in project delivery timelines.
- Integrated third-party APIs, services, and libraries to enhance the functionality and capabilities of the platform
- Experience building large-scale infrastructure applications and writing maintainable code in multiple programming languages
- Tested, and maintained applications and scripts to support various aspects of the **digital banking platform**

Clarivate Analytics, Bengaluru, India

Mar 2021 – Oct 2021

Software Engineer – NLP & Backend Systems

- Applied NLP techniques to recognize **peer review content** across large scholarly complex data in the application
- Developed scalable REST APIs and implemented SQL-based analytics queries.
- Deployed backend models and implemented notification systems integrated with MS Teams.
- Participated in model versioning, testing, and CI/CD using AWS and Git-based workflows.

Anora Semiconductor Labs, Bengaluru, India

May 2020 – Jan 2021

Software Engineer – Data Science

- Led, Designed and deployed a **Web-based analytics dashboard** leveraging ML algorithms for trend detection.
- Led the development of LMS using Django, Python, and custom classification models.
- Integrated ML components for real-time recommendation and improved system responsiveness.
- Collaborated across small teams to streamline product delivery in a startup environment.
- Took an active part in discussions to improve UI/UX through decision making thus being an active member of the start-up company
- Lead technical design discussions for medium to large-scale projects

DXC Technology - Bengaluru, India

Oct 2018 – Apr 2020

ITO Svc Delivery Consultant

- Worked in the development of **Marathon Web-based application** to track petroleum tanks using Django framework
- Designed and developed a chatbot tracking app
- Participated in low to moderately complex initiatives and projects associated with the technology domain, including installation, upgrades, and deployment efforts
- Identify opportunities for service quality and availability improvements within the technology domain environment
- Implemented the REST web services in the integration of a chatbot with the Marathon mailbox
- Utilized ORM frameworks such as SQLAlchemy to interact with relational database

Trainee Software Engineer

- Designed and developed a **Web-based Financial Analytics Tool** for data analysis and Visualization
- Reviewed, validated, and automated testing to improve code quality.
- Experienced in deploying web application in server and setting up required Python environments
- Used Python Pandas and Bokeh for data visualization and analysis
- Improved HR recruiting system from 55% to 80%

CORE SKILLS

- **Languages:** Python, SQL, Bash, JavaScript, HTML, CSS, C#, React, TypeScript
- **Data Tools:** Pandas, NumPy, Seaborn, Matplotlib, PySpark, Databricks, Data warehouse Data Lakes,
- **Frameworks/Tools:** Scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers, Fast API, Flask, DRF, .Net, Airflow, Django, LangChain, LangGraph
- **ML & AI & Deep Learning:** LLMs (OpenAI GPT-4, GPT3, Llama3, Nomic-Embedded-Text, DeepSeek), NLP, OpenCV, Machine Learning Pipelines, Model Monitoring, Embeddings, Linear Regression, Knowledge Graph, Random Forest, CNN, RNN, LSTM, XGBoost, Keras
- **Cloud & MLOps & DevOps tools:** AWS (EC2, S3, Lambda, CloudWatch, CloudFormation), GCP, Azure, Terraform, CI/CD, Git, Docker, Kubernetes, Load Balancing, Vercel AI, pg_vector
- **Databases/Vector Databases:** PostgreSQL, MySQL, SQLite, Chroma DB, Faiss, Pinecone
- **Practices:** Version Control, Test-driven Development (TDD), Agile/Scrum, API Development, OAuth
- **Others** – Hadoop, Code Reviews, ETL, SCRUM, Code Quality, LoRA, QLoRA, fine-tune, RAG, Apache Spark, Linux, Windows, Jupyter notebooks, AutoML, Data Structure, Object- Oriented Programming/Design, System development, Redis, Kafka, data manipulation, shell scripting, System Design, Web scraping, ORM, Data Modelling, Multithreading, System integration, GitHub Copilot, ChatGPT, Authentication & Authorization, chunking,
- **Quantitative Analytics:** Statistical Modeling, Time Series Analysis, Recommendation Analysis, Forecasting, Regression Analysis, Risk Modeling, Optimisation Hypothesis Testing, Experimental Design, A/B Testing