Akshit Bhatia

www.linkedin.com/in/being-akshit-bhatia • +49 1516 8555138

Bochum, Germany • *bhatia2akshit@gmail.com*

AI/ML Engineer and Python Backend Developer with 3+ years of experience architecting real-time systems, with integrated AI solutions.. Expert in creating production ready web applications, and deployment on Azure and AWS clouds. Proven ability to modernize complex legacy codebases, implement robust CI/CD workflows, and deliver production-grade systems under strict performance constraints.

Skills & Technologies

- **Languages:** Python, Java, JavaScript, TypeScript, SQL
- **Tools & Frameworks:** PyTorch, OpenCV, Hugging Face Transformers, LangChain, LangGraph, SmolAgent, CrewAI, PydanticAI, FastAPI, Flask, Django, Streamlit, React, HTML, CSS, Docker, Kubernetes, Jenkins, Terraform, Airflow, Redis, PostgreSQL, Weaviate, Git, CI/CD, Azure, AWS, Confluence, JIRA, Git.
- **Expertise:** AI/ML System Design, Real-Time AI Applications, Agentic AI & LLMs, Retrieval-Augmented Generation (RAG), Computer Vision & Image Generation, Model Fine-Tuning (LoRA), ML Pipelines & MLOps, Cloud-Native Deployment, Multi-Tenant Architectures, API Design, DevOps & Automation, Production Monitoring & Reliability

Work Experience

AI/ML Engineer   **Kumo Clouds Solutions**   **June 2025 - December 2025**
Cologne, Germany

- **Designed and deployed a multi-tenant, real-time interview bot using Python, LLMs, Django, PostgreSQL, and Redis**, enabling context-aware conversational AI with low-latency responses.
- **Implemented a retrieval-augmented generation (RAG) pipeline using embedding-based search and Weaviate,** extracting structured knowledge from 1,000+ pages of technical documentation to improve information accessibility.
- **Built end-to-end CI/CD pipelines using Jenkins, Terraform, Docker, and Azure,** enabling automated builds, infrastructure provisioning, and zero-downtime deployments.

Python Developer   **Auto1**   **April 2024 - October 2024**
Berlin, Germany (Remote)

- **Built an AI-driven news intelligence agent using Python, LangChain, FastAPI, and LLMs,** continuously monitoring industry and competitor news and generating sales-oriented summaries via contextual retrieval.
- **Integrated web search tools as agent actions and stored hundreds of articles in a vector database,** enabling semantic search and grounded LLM responses.
- **Automated Google Ads campaign management using Python APIs,** reducing manual operational effort.
- **Implemented CI/CD workflows using Jenkins and Airflow,** ensuring reliable deployments, scheduling, and production monitoring with Sentry.

ML Engineer - Internship   **PropertyExpert**   **July 2024 - October 2024**
Langenfeld, Germany (Remote)

- **Developed an AI image generation pipeline using Vision Transformers, GPT-2 captioning,** and diffusion models, generating context-consistent synthetic images for damaged infrastructure scenarios.
- **Trained a ResNet-50-based fake image detection model using PyTorch,** classifying real vs. synthetic artifacts to support content-agnostic quality assurance pipelines.

Python Developer - Internship   **Blue Avenir**   **July 2022 - February 2023**
Dusseldorf, Germany

- **Improved predictive analytics pipelines using Python, Pandas, XGBoost, and feature optimization,** increasing model accuracy by 6% on structured business datasets.
- **Built a semantic document search pipeline using LLMs and vector embeddings,** enabling efficient information retrieval from unstructured documents.

Research Assistant - WHK   [Computational Social Science](#)   **April 2021 - May 2022**
Paderborn, Germany

- **Trained and fine-tuned transformer-based language models (BART, GPT-2, RoBERTa) using Hugging Face Transformers** to generate contextually relevant and persuasive counter-arguments.
- **Developed Flask-based inference APIs in Python,** enabling controlled experimentation and evaluation of generated text outputs.

Software Developer   [Sopra Steria](#)   **October 2015 - October 2019**
Delhi, India

- **Developed and deployed production-grade enterprise applications using Java, EJB, Hibernate, and SQL,** ensuring scalability, reliability, and maintainability in large systems.
- **Built a face recognition system using OpenCV and scikit-learn for secure access control,** and implemented clustered ticket grouping logic to streamline assignment workflows.
- **Collaborated in agile, distributed teams, contributing to CI pipelines, system upgrades, and production releases** using Jenkins and version control tools.

Education

Masters: Informatiks, University of Paderborn, Germany   **October 2019 - August 2024**
Thesis: "Train Neural Networks to detect Images generated with AI": Developed and deployed two end-to-end computer vision pipelines in Python. The work covered dataset preparation, model training and evaluation, hyperparameter tuning, and performance analysis, with a strong focus on generalization, robustness, and real-world applicability of image authenticity detection.

Bachelor of Technology: Computer Science, Indraprastha University, Delhi   **August 2011 - May 2015**
Thesis: "InfoRank: an algorithm to rank Reviews": designed and implemented an entropy-based algorithm to rank large-scale user reviews by informational value, applying statistical methods and feature engineering to text data. The project involved data preprocessing, relevance scoring, and evaluation against baseline ranking approaches.

Projects & Extra

REACT Based Agentic AI System   **December 2025**
Designed and implemented an Agentic AI application using LLMs and tool orchestration, enabling safe tool execution, state tracking, and decision-making across web search, mathematical reasoning, and governance writing, with a lightweight API consumable by automation frameworks such as Trigger.dev.

LLM Jokes Better   **October 2024**
Fine-tuned Mistral 7B using LoRA and prompt engineering to generate contextually relevant, high-quality humor, demonstrating parameter-efficient LLM fine-tuning, controlled text generation, and applied generative AI system design.

Certifications

IBM Certificate: [Develop Generative AI Applications: Get Started](#)   **December 2025**
IBM Certificate: [Build RAG Applications: Get Started](#)   **December 2025**
AWS: [Generative AI with Large Language Models](#)   **October 2024**

Languages

- English: Professional Proficiency
- German: Professional Proficiency