Yash Vijay Kucheriya

ykucheri@asu.edu | +1 (623) 276-6592 | linkedin.com/in/yash-kucheriya

SUMMARY

Software Developer with 3 years of experience spanning backend engineering, AI engineering, and data science/analytics. Strong expertise in **Python** and **Java**, with hands on experience building scalable backend systems using **Spring Boot**, RESTful APIs, and cloud-native architectures on **AWS**. Experienced in developing AI and data-driven applications, productionizing ML models, and leveraging **AI-assisted development tools** to accelerate delivery. Adept at working across the full lifecycle from system design to deployment and optimization in a fast paced, collaborative environments.

TECHNICAL SKILLS

- **Languages:** Python, Java, TypeScript, Go, C++, SQL
- **Backend & APIs:** Spring Boot, Node.js (TypeScript), Flask, Laravel, REST APIs
- **Frontend:** React, Angular, HTML, CSS
- **Machine Learning & Data:** TensorFlow, Scikit-learn, Keras, XGBoost, Hugging Face, Pandas, OpenCV, ARIMA, LSTM
- **Cloud & Databases:** AWS (ECS, EC2, Lambda, DynamoDB, S3, SageMaker), PostgreSQL, MongoDB, Snowflake, Spark, MariaDB, GCP, Kafa
- **DevOps & Tooling:** Docker, CI/CD, Git, Jenkins, Terraform, CloudWatch, IBM Sterling
- **Developer Productivity:** AI-assisted tools (Cursor, GitHub Copilot, Claude Code)

EDUCATION

Arizona State University, Tempe, AZ   *Jan 2024 – Dec 2025 | GPA: 3.8/4.0*
MS in Computer Science. Coursework: Statistical ML, Algorithms, Cloud, NLP, Data Mining, Networks

Walchand Institute of Technology, Solapur, India   *Jun 2016 – May 2020 | GPA: 9.23/10*
B.E. in Computer Science and Engineering. Coursework: Data Structures, Operating Systems, Big Data, Web Development

WORK EXPERIENCE

ZenZiee (YesTech Corp.), NJ, USA — *Software Engineer*   Jun–Aug 2025

- Built **ZenZiee**, a gamified social application integrating **real-time APIs, analytics dashboards**, and **user sentiment tracking** for 100+ beta users.
- Deployed containerized microservices using **AWS ECS/ECR, ALB, and CloudWatch**, achieving 99.9% uptime and low-latency data synchronization.
- Designed **event-driven pipelines** with **DynamoDB Streams, AWS Lambda, and S3**, reducing processing latency by 60% and enabling parallel execution.
- Collaborated with ML and frontend teams to embed **emotion classification models**, increasing user engagement by 25%.
- Leveraged **AI-assisted development tools (Cursor, GitHub Copilot)** to streamline debugging, testing, and code reviews.
  Tech: AWS, Docker, Python, PostgreSQL, React, Swift, Xcode, iOS, Google Analytics, Athena

Buildoors Lab Ltd, London, UK — *Data Scientist*   Jan–Dec 2023

- Developed a blockchain-based fraud detection system, reducing risk exposure by 35% through **real-time anomaly detection**.
- Automated ETL pipelines and built compliance dashboards using **Pandas, SQL, Tableau, and Power BI**, improving data throughput and reporting reliability.
- Deployed ML models on **AWS SageMaker** with automated retraining and monitoring workflows, including evaluation metrics to track performance over time.
  Tech: Python, TensorFlow, SQL, Pandas, Power BI, Tableau, AWS SageMaker, S3, Excel

Infosys Ltd, Pune, India — *Software Engineer*   Nov 2020–Jul 2022

- Designed and developed **Java Spring Boot microservices** and **REST APIs** for enterprise billing systems, reducing manual effort by 50% and improving accuracy by 40%.
- Improved backend performance by refactoring **MSSQL queries**, increasing throughput by 35% across high volume workflows.
- Managed 1M+ **EDI transactions per month** using **IBM Sterling**, ensuring reliability under strict SLAs.
- Implemented **CI/CD pipelines, monitoring, and deployment automation**, reducing release cycles by 45% and mentoring junior engineers across cross-functional teams on backend best practices.
  Tech: Java, Spring Boot, MSSQL, IBM Sterling B2B, REST/gRPC, Jenkins, Linux, .NET, C#, Business Intelligence

PROJECTS

Logical Language Modeling for Reasoning   Jan–May 2025

- Fine-tuned **LLaMA 3.2** on **OpenWebMath** with reweighted loss; boosted **LogiQA** accuracy 28%.
- Applied **attention diagnostics** and AI-based prompt optimization for logical consistency and interpretability.
- Built scalable inference scripts for distributed GPUs and cloud checkpoints. *Tech: PyTorch, Hugging Face, LLaMA 3.2, CUDA GPUs, OpenWebMath*

Cloud-Based Face Recognition   Aug–Dec 2024

- Built **Flask REST API** for CNN inference with **OpenCV preprocessing**, deployed via **AWS EC2 + Docker**.
- Added **S3-based data versioning** and implemented concurrency-safe inference handling. *Tech: TensorFlow, OpenCV, Flask, AWS EC2, S3, Docker, Python*

Sales Forecasting (Time Series)   Jan–May 2025

- Combined **ARIMA, LSTM, and XGBoost** for ensemble forecasting; achieved 93% accuracy and 30% lower MAE.
- Automated feature engineering using **Pandas** and **NumPy** to produce rolling predictions in cloud jobs. *Tech: Python, Pandas, TensorFlow, ARIMA, XGBoost, Matplotlib, Data models*

Fake News Detection   Jan–May 2024

- Built hybrid **TF-IDF + BiLSTM** classifier reaching 95% accuracy; deployed for real-time screening.
- Integrated REST-based moderation service and **Dockerized deployment** for scalability. *Tech: TensorFlow, Keras, Flask, REST APIs, NLP, Docker, Semantic analysis*