Yash Kumar – AI Developer | AI/ML Engineer

Professional Summary

AI Developer with 2+ years of experience architecting and deploying production-ready multimodal AI systems. Specialized in Retrieval-Augmented Generation (RAG), Computer Vision, and Generative AI, delivering solutions that improved model accuracy by 10% and reduced inference latency by 15%. Skilled in designing and scaling enterprise-level AI pipelines using LangChain, LangGraph, FastAPI, Docker, and Azure Cloud.

Proven track record of transforming raw data into business impact, such as reducing report generation time by 99%, improving workflow efficiency by 25%, and driving automation with computer vision models. Experienced in working across cross-functional teams, simplifying AI outputs, and enabling actionable insights for decision-making.

Core Skills

- **Programming & Data: Python, SQL, Pandas, NumPy, Scikit-Learn**
- **AI/ML & DL: Machine Learning (Supervised/Unsupervised), Deep Learning, ANN, RNN, CNN, LSTM, GRU, Transformers**
- **Generative AI & LLMs: LLMs, RAG, AI Agents, LangChain, LangGraph, LangSmith, LoRA, QLoRA**
- **NLP & CV: NLP, Computer Vision, Image Processing, OCR, OpenCV**
- **Frameworks & Tools: TensorFlow, PyTorch, FastAPI, Docker, Git/GitHub, FAISS, Pinecone**
- **Cloud & Deployment: Azure Cognitive Services, AWS SageMaker, Linux, Ubuntu, Windows**
- **Other Expertise: EDA, Data Cleaning, Data Visualization, Statistical Analysis, Orchestration Frameworks, Automation (n8n, crewAI), Agile**
- **Soft Skills: Problem-solving, teamwork, stakeholder communication, continuous learning**

Experience

Cyber Infrastructure Pvt. Ltd. | AI/ML Engineer

Oct 2024 – Present | Indore

- **Built GenAI + RAG applications → improved query resolution by 20%, reduced response time by 25%.**
- **Designed end-to-end AI/ML pipelines using Python, FastAPI, Docker, Azure (scaling to 7K+ records).**
- **Deployed CV + NLP automation models (LangChain + LangGraph) → boosted workflow efficiency by 10%.**
- **Partnered with stakeholders to deliver business-driven AI insights.**

AlmaBetter | Data Science Fellow (Applied AI Projects)

Jun 2023 – Sept 2024 | Delhi

- **Developed and optimized ML/GenAI models in computer vision & predictive analytics.**
- **Built dashboards (Power BI, Tableau) for decision-making, improving stakeholder engagement.**

DIFACTO Robotics | Robotics Engineer

Jan 2020 – May 2023 | Bangalore

- **Designed CV-powered robotic systems → raised production accuracy by 20%.**
- **Built deep learning inspection modules with OpenCV → reduced manual checks by 30%.**
- **Led Kaizen teams → ensured global compliance and process improvement.**

Key Projects

1. Automated RAG-Driven Maritime Intelligence System (LangChain + Azure)

- **Problem: Manual SQL report processing slowed operations.**
- **Solution: Created a centralized AI-powered RAG portal integrating SQL records + maritime safety manuals.**
- **Built ingestion pipelines, embeddings, and semantic search.**
- **Designed RAG pipeline → dynamically fetched best practices and safety guidelines.**
- **Deployed as a production-grade Azure solution (semantic search + APIs).**
- **Impact: Report generation time reduced >99%, faster decision-making, stronger compliance.**

---

2. AI-Powered Digital Magazine Viewer (LLM + GenAI + CV)

- **Problem: PDF magazines unreadable on mobile; poor engagement.**
- **Solution: Built PDF-to-HTML AI pipeline: OCR + OpenCV + CLIP to extract structured text/images.**
- **Classified content (article, ad, poster), skipping irrelevant parts.**
- **Created Inshorts-like UX → covers + headlines + clean HTML article views.**
- **Deployed with FastAPI + Docker + MongoDB.**
- **Impact: Boosted readability, engagement, and user satisfaction.**

---

3. Smart Trolley – Retail Product Verification with YOLOv8

- **Problem: Barcode-only checkout allowed theft & mismatches.**
- **Solution: Built real-time video verification pipeline.**
    - **Detected new product placement with video + OpenCV.**
    - **Passed frames through YOLOv8 model → verified vs barcode.**
- **Created training pipeline for millions of SKUs with auto-finetuning.**
- **Delivered as POC with live demos.**
- **Impact: Reduced theft risk, validated feasibility for scalable retail automation.**

Education

Guru Gobind Singh Indraprastha University – B.Tech in Mechatronics Engineering (2015 – 2019)
CGPA: 8.2 / 10.0

The Union Academy Sr. Sec. School, Delhi

- **10th: 74%**
- **12th: 72%**

Personal Interests

- **Music: Playing instruments, exploring rhythm and creativity.**
- **Cooking: Experimenting with cuisines, blending precision with artistry.**
- **Continuous Learning: Exploring orchestration frameworks, multimodal AI, and creativity-driven problem-solving.**

1. What should we know about your life story?

I began my career in robotics engineering, where I built computer vision–powered robotic systems for global automotive clients. That foundation in robotics gave me a deep appreciation for precision and efficiency. Over time, I transitioned into AI and machine learning, where I now specialize in multimodal and generative AI systems. My journey has always been about solving real-world problems through technology—whether automating manufacturing or creating AI platforms for enterprises. Outside of work, I nurture my creativity through playing musical instruments and cooking, both of which balance my technical mindset with artistic expression.

2. What's your #1 superpower?

My ability to simplify and operationalize complexity. I can take highly technical AI concepts—like retrieval-augmented generation pipelines or multimodal computer vision models—and translate them into production-ready systems that directly solve business challenges. This combination of deep technical knowledge and practical problem-solving is what sets me apart.

3. What are the top 3 areas you'd like to grow in?

- **Scaling enterprise AI systems:** Designing solutions that can process massive volumes of real-time data with reliability and efficiency.
- **Mastering orchestration frameworks and AI agents:** Building autonomous, intelligent systems that can coordinate multiple tasks and tools seamlessly.
- **Leadership and mentoring:** Guiding younger engineers, fostering collaborative teams, and eventually leading large-scale AI innovation projects.

4. What misconception do your coworkers have about you?

People sometimes assume I'm "only technical." In reality, while I enjoy coding, model optimization, and system design, I also place equal importance on creativity, collaboration, and business strategy. I bridge the gap between AI development and business impact, ensuring my work is not just innovative but also practical and value-driven.

5. How do you push your limits?

I push myself by constantly stepping outside my comfort zone. Whether it's adopting new frameworks like LangGraph, optimizing inference pipelines, or designing unique

AI+business solutions, I embrace challenges as opportunities to learn. In my personal life, I take the same approach—when learning a new instrument or experimenting with a new recipe, I practice, fail, and refine until I succeed.

6. What motivates you the most?

What excites me is the opportunity to create systems that transform workflows and improve people's lives. When I see a maritime team cut report generation time from hours to seconds, or readers enjoy magazines in a new digital format, that impact motivates me to keep innovating. In a personal sense, music and food bring joy to people around me, and I see AI as another way to deliver value and happiness—just on a much larger scale.

7. How do you deal with setbacks?

For me, setbacks are experiments that didn't work out yet. I analyze what went wrong—was it data quality, model choice, or deployment pipeline—and then refine the approach. It's very similar to cooking: if a dish doesn't taste right the first time, I tweak the recipe until I get it right. That mindset keeps me resilient and focused on progress.

8. Who inspires you professionally?

I'm inspired by innovators who merge technical depth with creativity—AI researchers pushing the limits of multimodal learning, and entrepreneurs who translate that research into products that change industries. Their ability to blend vision with execution is something I constantly strive for.

9. How do you balance work and life?

I maintain balance by dedicating time to music and cooking. Playing an instrument sharpens my focus and brings calm, while cooking allows me to experiment and create outside of code. These activities recharge me so I can return to work energized and creative.

10. What's one project you're most proud of?

The RAG-driven maritime intelligence platform is my proudest project. By combining SQL data, semantic search, and maritime safety manuals, I built a system that reduced report generation time by 99%. It transformed how maritime teams accessed knowledge and made decisions, proving the tangible value of AI in real-world operations.

11. How do you approach teamwork?

I see teamwork like leading a band or running a kitchen. Every member has a distinct role, but success depends on coordination, timing, and trust. I encourage open communication, respect diverse skills, and focus on creating harmony within the team so that everyone can contribute at their best.

12. What's a unique strength you bring to teams?
I excel at bridging the gap between deep technical expertise and business clarity. I can explain complex AI outputs in simple, actionable terms, helping both engineers and business stakeholders align on goals. My mix of technical rigor, creative thinking, and communication skills allows me to add value beyond just code.

13. What do you enjoy learning about outside of work?
I enjoy exploring music theory, experimenting with new cuisines, and finding parallels between art and technology. Both music and cooking require creativity, precision, and iteration—qualities that also drive innovation in AI. This continuous learning outside of work feeds my curiosity and enhances my problem-solving mindset.

14. If you weren't in AI, what would you be doing?
I'd likely be a chef or musician. Both fields allow me to create, innovate, and share experiences with others. Whether composing a melody or designing a dish, the process of building something that resonates with people gives me the same fulfillment that AI development does.

15. Where do you see yourself in 5 years?
In five years, I see myself leading AI innovation teams, mentoring younger engineers, and driving projects that combine multimodal intelligence, automation, and creativity. My goal is to build human-centric AI products that don't just work technically but also feel intuitive, impactful, and accessible to everyone.