VIVEKANAND KUMAR

AI / ML Engineer | Generative AI | RAG Pipelines | Computer Vision
Noida, Uttar Pradesh, India | +91-8651240027 | vivekanandkumar865124@gmail.com
linkedin.com/in/vivekanand2003 | github.com/vivekanandkumar001 | huggingface.co/vivekananda0010 | progeni.live

PROFESSIONAL SUMMARY

AI/ML Engineer (B.Tech CSE – AI, 2026) who builds and ships RAG pipelines, LLM-powered applications and real-time computer vision systems. I own projects end to end – data, model, backend, frontend, Docker and deployment on Hugging Face Spaces. Currently an AI-Ready Full-Stack Intern at Mobiloitte Technologies. Seeking fresher GenAI / LLM engineering roles.

TECHNICAL SKILLS

GenAI & ML	LLMs (LLaMA 3 / 3.3-70B), RAG, Prompt Engineering, LangChain, ChromaDB, Ollama, NLP, PyTorch, TensorFlow, Scikit-learn
Computer Vision	YOLOv5, OpenCV, ONNX
Languages	Python, SQL, JavaScript, TypeScript
Backend & Web	Flask, FastAPI, REST APIs, Node.js, Express.js, React, Next.js, MongoDB, MySQL, WebAssembly
Tools & Cloud	Docker, Hugging Face Spaces, Git, GitHub, GitHub Actions

EXPERIENCE

Intern Software Engineer – AI-Ready Full-Stack
Aug 2026 – Present

Mobiloitte Technologies India Pvt. Ltd., New Delhi

- Selected for the Train-to-Hire track (AI-Ready Workforce Program) to build live AI-enabled full-stack applications.
- Work with senior engineers on full-stack development, AI integrations, API testing, peer code reviews and debugging.

PROJECTS

Progeni Atlas AI – AI Career Platform with Video Interviewer
atlas.progeni.live

LLaMA 3.3-70B, RAG, Flask, FastAPI, SQLAlchemy, Next.js 14, Tailwind CSS, Web Speech API, Docker, GitHub Actions

- Built an end-to-end AI career platform (resume analysis, job search, mock interviews) with a Python backend (Flask/FastAPI + SQLAlchemy) and a Next.js 14 + Tailwind frontend, deployed live on Hugging Face Spaces.
- Developed **Swami AI**, a voice + vision mock interviewer using the Web Speech API and webcam; it conducts real-time technical interviews with an LLM and produces structured feedback after the session.
- Engineered **dual-mode ATS resume scoring**: a general audit (no JD) and a JD-targeted analysis, each with its own weighted scoring pillars, so scores reflect what the user actually needs.
- Implemented RAG-based vector search for resume-to-job matching and integrated job-board APIs with a scraper; search guarantees a minimum of 5 matches, sorted by real posting recency.
- Overhauled free-tier reliability so the app stays usable under API rate limits and quotas; added multi-domain support and in-app feedback collection.
- Containerized with Docker (gunicorn) and automated CI/CD through GitHub Actions to Hugging Face Spaces on every push.

Progeni-Lens – Real-Time Object & Risk Detection
Live Demo (Hugging Face)

YOLOv5, OpenCV, ONNX Runtime, Python, Hugging Face Spaces

- Built a multi-mode computer vision system covering **exam monitoring, traffic analysis, security and industrial risk detection**, working on both uploaded images and live webcam streams.
- Exported YOLOv5 to ONNX and tuned the inference pipeline: **25% lower latency** and a sustained **30+ FPS** on standard hardware (no GPU required).
- Deployed as a public Hugging Face Space so anyone can try real-time detection from the browser.

ProgeniBench – 30 Private Browser Utilities
bench.progeni.live

Vanilla JavaScript, HTML5, CSS3, Client-Side Web APIs

- Designed and shipped a privacy-first portal of **30 client-side tools** (CSV converter, subtitle repair, photo cropper, contact de-duplication and more); everything runs in the browser with **zero server uploads**.
- Built with no backend or framework dependency; each tool has its own SEO-optimized page with Open Graph metadata. Audited and fixed the codebase using AI coding agents with a verification-first workflow.

WhatsSwift – Private WhatsApp Utilities
wakit.progeni.live

JavaScript, FFmpeg (WebAssembly), Client-Side Web APIs

- Built a free, 100% in-browser toolkit: **30-second status video splitter**, direct chat via wa.me, QR code generator and audio compressor.
- Runs FFmpeg compiled to WebAssembly directly in the browser, so videos and audio are processed on the user's device with zero uploads; SEO-ready page for every tool.

EDUCATION

B.Tech in Computer Science & Engineering (AI) | CGPA: 7.06/10
2022 – 2026

Greater Noida Institute of Technology (GNIOT), Greater Noida

CERTIFICATIONS & TRAINING

- Student Development Program on Data Science (AI/ML) – E&ICT Academy, IIT Kanpur (MeitY)
- AI/ML Training Program – Greater Noida Institute of Technology | Python Coder & Dataset Creator – Kaggle