VIKNESH M

AI Engineer | LLM Training | Finetuning | RAG | AI Agents

Coimbatore, India | vikneshoftheleaf@gmail.com | 7708605707 | Portfolio | GitHub | Hugging Face | LinkedIn

SUMMARY

AI engineer who trains language and diffusion models from scratch and ships LLM apps (RAG, agents) end to end. M.Sc. Computer Science, Durham University Visiting Scholar, 3x hackathon winner.

SKILLS

AI / ML: PyTorch, Hugging Face (Transformers, Diffusers, Hub), Transformer and Diffusion Model Training, RAG, LLM APIs (Gemini), AI Agents, Prompt Engineering

Engineering: Python, TypeScript, JavaScript, SQL, FastAPI, Node.js, Next.js, REST APIs, Gradio

Cloud & Tools: Docker, Git, Linux, GitHub Actions, Cloudflare, Vercel, Azure, Google Colab, PostgreSQL, MongoDB

AI PROJECTS

Atom-1.0 and Atom-1.0-Instruct | PyTorch, Hugging Face, Colab | Model *2026*

- Designed and trained from scratch a 29M-parameter decoder-only language model (8 layers, 512 hidden size, RoPE, RMSNorm, SwiGLU, tied embeddings, 8K vocabulary, 512-token context) on a Colab GPU, with step-interval checkpoints synced to the Hugging Face Hub.
- Instruction-tuned it to write short stories from required words and features, and published both models on Hugging Face with a demo Space.

PtahX32 | PyTorch, Diffusers, CLIP, Hugging Face | Model *2026*

- Trained a text-to-pixel-art diffusion model from scratch on a single Colab T4: conditional UNet with cross-attention, frozen CLIP text encoder, DDPM training, DDIM sampling with classifier-free guidance, EMA weights, and fp16 mixed precision.
- Built a resumable pipeline that checkpoints to the Hub every 500 steps with sample grids and loss curves; reached 0.0022 validation loss and released the weights and a full model card under MIT.

Silverline Board Documents RAG Chatbot | Gemini Flash, RAG, Next.js, Vercel | GitHub *Oct 2026*

- Built a production-oriented RAG chatbot over board-meeting PDFs, DOCX minutes, and XLSX financial tables that answers only from the source documents, deployed on Vercel with a chat UI and an embeddable website widget.

Gemini Browser Agent | Gemini Live API, Chrome Extension (MV3) | GitHub *2026*

- Built a voice-controlled agent where Gemini Live picks from 43 browser tools, with approval gates for risky actions and scheduled tasks.

Interviewest | Gemini Live API, Web Audio, Cloudflare Pages | Live Demo *2026*

- Built a real-time voice interview simulator with duplex audio streaming, 4 interviewer personas, and in-browser resume parsing, with all user data kept private in the browser.

Lead-forge | Gemini, SerpApi, Cloudflare Workers, FastAPI | Live Demo *SerpApi India Hackathon*

- Built a pipeline from a Google Maps search to a qualified lead, a Gemini-generated website, and a ready-to-review outreach message.

EXPERIENCE

Developer | Zorixo Ventures, Tiruppur *June 2026 – Sep 2026*

- Contribute across the full SDLC from requirements to production deployment; design data models and optimized queries across relational and NoSQL databases.

Software Engineer Intern | LexDash, Chennai *Dec 2025 – Jan 2026*

- Built RESTful APIs backed by MongoDB and full-stack features with React.js, Next.js, and Tailwind CSS.
- Worked with the product team in an agile environment to ship features and resolve bugs.

Freelance Developer | Self-employed *5+ clients*

- Delivered web apps and AI-enabled tools from requirements to cloud deployment and support.

EDUCATION & CERTIFICATIONS

M.Sc., Computer Science | Bharathiar University, Coimbatore *2024 – 2026*

Visiting Scholar, Talent Programme (AI & Data Science) | Durham University, UK *Mar – Jun 2024*

Certifications: GenAI Powered Data Analytics (TCS); Technology Software Development (Citi). **Achievements:** 3x hackathon winner.