ROHITH GARAPATI

Hyderabad, India
rohith22gksr@gmail.com | +91 9182194370
Portfolio | LinkedIn | GitHub

SUMMARY

A Computer Science Engineering fresher with 2+ years of hands-on experience in prompt engineering and AI-driven development. Skilled in using modern tools and methods to solve problems, develop programs, and build AI solutions. Experienced in rapid prototyping, code generation, and model development through advanced prompt engineering techniques. Competent and dedicated to developing best-in-class software, AI systems, and AI agents. Open to opportunities in AI, software engineering, and product development.

EDUCATION

GITAM University, Hyderabad   *2021 – 2025*
B.Tech, Computer Science & Engineering
Abhyasa International Residential School (ICSE)   *2016 – 2019*

SKILLS

- **AI/ML:** Deep Learning, NLP, Computer Vision, Transformers, Attention Mechanisms, Generative AI, Large Language Models (LLMs), Model Optimization (FP4, Quantization, QAT, PTQ), AI Agents, Vision Transformers.
- **Transformer Architecture:** Multi-Head Attention, Self-Attention, Cross-Attention, Positional Encoding, Layer Normalization, Feed-Forward Networks, Residual Connections, Autoregressive Models.
- **LLM Development:** Model Design, Training Strategies, Fine-tuning, RLHF, Context Window Optimization, Prompt Engineering, Inference Optimization, Model Compression.
- **Prompt Engineering:** 2+ years experience with advanced prompting techniques, chain-of-thought reasoning, few-shot learning, prompt optimization, and AI model interaction design.
- **AI-Aided Development:** AI Coding Assistants (Cursor, VS Code), Automation, AI Agent Workflows, APIs (Openrouter, Gemini).
- **Software Engineering:** Full-Stack Development, API/System Design, UX, Agile, Debugging, Deployment, Iterative Dev.
- **Frameworks/Tools:** PyTorch, TensorFlow, Hugging Face, FastAPI, Flask, Streamlit, Git, Docker, K8s, AWS, GCP, NVIDIA NIMS.
- **Core Concepts:** Data Structures & Algorithms, Product Thinking, Hardware Optimization, UI/UX Principles.

EXPERIENCE

AI-Aided Software Developer & Problem Solver   *2022 – Present*
Project-Based

- Designed and implemented diverse software solutions and AI models using AI-assisted development tools and advanced prompt engineering.
- Leveraged 2+ years of prompt engineering expertise and agent workflows to deliver efficient end-to-end systems.
- Developed proficiency in optimizing AI model interactions through sophisticated prompting strategies and workflow automation.

AI/ML & Product Design Enthusiast   *Self-Directed Learning*

- Explored image generation, large models, and prompt design through hands-on experimentation.
- Gained strong product and UX skills via iterative building and real-world testing.

PROJECTS

More at: github.com/INFINITYone22

AI/ML & Deep Learning

- **INFINITYone22_Multimodal:** Designed large-scale multimodal transformer architecture with custom attention mechanisms and hardware-optimized inference.
- **A-Unified-Transformer:** Autoregressive transformer model architecture for unified text, image, and audio generation with novel cross-modal attention.
- **ReasoningTransformer:** Custom transformer architecture with enhanced reasoning capabilities and chain-of-thought attention patterns.
- **TriQuant:** Novel quantization framework for transformer models with ternary-precision (2-bit) optimization and attention mechanism preservation.

AI Tools & Web Apps

- **Image-vector-database:** Web app to store/search images using AI and embeddings.
- **SemantiCore:** Local document search app that understands meaning, not just keywords.
- **Resume-analyzer:** AI tool to evaluate resumes with NLP.
- **ContextFlow:** Research on handling long context in LLMs.

System Design & Infra

- **Hierarchical_MultiModal_AI_CNC:** Modular system to automate CNC manufacturing with AI.
- **Inferium:** AI inference speed simulator across NVIDIA GPUs.
- **System-Design-for-Large-Multimodal-Language-Model:** Code & notes for scalable model systems.

Dev Tools & Programming Languages

- **AION-lang:** A human-readable language for AI/ML workflows.
- **python-projects:** 60+ Python programs across domains.

Web Dev & Visuals

- **Portfolio Website**: Modern, responsive portfolio website showcasing AI/ML projects with cyberpunk-inspired design, smooth animations, and optimized performance.
- **AI-Concepts-Visualization-Web-HTML:** Web project to visualize AI concepts.
- **Web-Development-designs:** Modern UI/UX web design projects.