

Shriyansh Singh

+1-930-333-5141 | shriyansh.singh24@gmail.com | [LinkedIn](#)

SUMMARY

AI Solutions Architect with expertise crafting **GenAI prototypes** and **consulting** clients on ML implementations while translating **business objectives** into scalable technical solutions.

PROFESSIONAL EXPERIENCE

AI Solutions Consultant

April 2024 - Dec 2024

Hyphenova AI

Los Angeles, CA

- **Advised** enterprise clients on **GenAI strategy**, translating business requirements into technical architecture designs that accelerated AI adoption by 65% across 12 organizations
- **Constructed** rapid prototype applications using **LangChain**, **Hugging Face**, and **Google Cloud Vertex AI** that provided proof-of-concept validations for customer-specific use cases
- **Delivered** 15+ technical workshops and presentations on **model fine-tuning approaches**, prompting strategies, and cost optimization techniques to technical and executive audiences
- **Collaborated** with product teams to integrate customer feedback into development roadmaps, influencing feature prioritization based on real-world implementation challenges

Technical Solutions Specialist

May 2022 - Oct 2022

Enterprise Business Technologies Pvt Ltd

Mumbai, India

- **Demonstrated** AI/ML capabilities through **proof-of-concept solutions** developed using **TensorFlow** and **PyTorch** that showcased practical business value to potential clients
- **Facilitated** 8 technical discovery sessions to understand client requirements, identifying optimal architecture patterns that aligned with business objectives and technical constraints
- **Developed** reusable solution accelerators and technical assets using **Python** and **Terraform** that reduced implementation time by 40% for common use cases
- **Communicated** complex technical concepts to diverse audiences through customized presentations that bridged the gap between business stakeholders and technical teams

PROJECTS

Generative AI Customer Solution Framework | *Google Vertex AI, LangChain, Hugging Face, Streamlit, Docker* Dec 2024

- **Architected** modular framework for rapidly building vertical-specific GenAI applications that shortened proof-of-concept development time from weeks to days for customer engagements
- **Implemented** parameter-efficient fine-tuning approaches using **LoRA** and **PEFT** techniques that optimized model adaptation for domain-specific tasks while reducing GPU requirements by 70%
- **Documented** comprehensive technical guides and reference architectures that enabled sales teams to effectively communicate solution capabilities and implementation approaches to customers

Enterprise RAG Solution Accelerator | *Google Cloud, Vector Databases, LangChain, FastAPI, Terraform* May - Nov'24

- **Designed** scalable Retrieval-Augmented Generation architecture that enabled enterprises to build AI applications on top of their proprietary knowledge bases, improving response accuracy by 87%
- **Pioneered** infrastructure-as-code approach using **Terraform** modules that enabled consistent deployment across development, staging, and production environments with governance controls
- **Presented** the solution to 25+ client organizations through interactive demonstrations that showcased the business value proposition while addressing technical implementation considerations

SKILLS & CERTIFICATIONS

Programming: Python, Java, SQL, Shell Scripting, YAML, JSON

AI/ML Technologies: TensorFlow, PyTorch, Hugging Face, LangChain, scikit-learn, Vertex AI, OpenAI

Cloud & Infrastructure: Google Cloud Platform, AWS, Kubernetes, Docker, Terraform, CI/CD

Data Processing: Vector Databases, Airflow, Spark, Beam, Pandas, NumPy, Embeddings

Solution Design: Technical Architecture, Prototyping, MLOps, LLMOps, RAG, Fine-tuning

Communication: Technical Presentations, Business Value Discovery, Workshops, Documentation

EDUCATION

Indiana University Bloomington

Aug 2023 – May 2025

Master of Science in Data Science

Indiana, United States

- Relevant Coursework: Generative AI Applications, Machine Learning Systems, Natural Language Processing, Cloud Computing, Business Analytics
- GPA: 3.8/4.0