Shriyansh Singh

+1-930-333-5141 | **shriyansh.singh24@gmail.com** | **LinkedIn** | **GitHub**

SUMMARY

ML Engineer specialized in search relevance, query understanding, and language models. Experienced in building personalized recommendation systems and domain-adapted NLP solutions.

PROFESSIONAL EXPERIENCE

Machine Learning Engineer *April 2024 - Dec 2024*

Hyphenova AI *Los Angeles, CA*

- **Designed** and **implemented** a personalized search ranking system using gradient boosting models that improved query relevance by 28% across enterprise customers
- **Developed** novel user interaction signals that enhanced search personalization, resulting in a 35% increase in user engagement metrics
- **Built** a query understanding pipeline using transformer models that effectively captured user intent across ambiguous search queries, improving zero-result searches by 45%
- **Integrated** LLM-powered question answering capabilities with traditional search, enabling direct answers for 62% of natural language queries

Software Engineer, NLP *July 2022 - July 2023*

Enterprise Business Technologies Pvt Ltd *Mumbai, India*

- **Led** the development of domain adaptation techniques for pre-trained language models, achieving 40% better performance on industry-specific tasks with minimal labeled data
- **Implemented** efficient retrieval mechanisms that combined dense and sparse representations to improve accuracy while maintaining sub-100ms latency requirements
- **Created** an evaluation framework that measured natural language generation quality across multiple dimensions, enabling data-driven model selection
- **Collaborated** with product teams to translate customer needs into technical requirements, ensuring ML solutions addressed real business problems

KEY PROJECTS

Enterprise Knowledge Graph Search | *Python, PyTorch, Elasticsearch, Neo4j* *Oct 2023 – Feb 2024*

- **Architected** a hybrid search system that combined knowledge graph traversal with neural retrieval to answer complex multi-hop questions
- **Developed** custom embedding models fine-tuned on enterprise data that outperformed general-purpose embeddings by 32% on domain-specific tasks
- **Implemented** an efficient caching and indexing strategy that reduced query latency by 65% while maintaining result quality

Adaptive Ranking Framework | *Python, TensorFlow, Pandas, LightGBM* *Mar 2023 – Sep 2023*

- **Designed** a multi-stage ranking system with specialized models for different query types, improving relevance across the query distribution
- **Built** an A/B testing infrastructure that enabled rapid experimentation and statistical validation of ranking improvements
- **Created** feature extraction pipelines that captured user, document, and contextual signals for personalized search experiences

TECHNICAL SKILLS

Machine Learning: Search Ranking, Query Understanding, Natural Language Processing, Language Models
ML Frameworks: PyTorch, TensorFlow, Hugging Face Transformers, Scikit-learn, LightGBM, XGBoost
Search & Retrieval: Vector Search, BM25, Hybrid Retrieval, Query Expansion, Personalization, Elasticsearch
Programming: Python, Java, SQL, C++, Go (basic)
Data Processing: Pandas, NumPy, Spark, Ray, Feature Engineering, ETL Pipelines
Tools & Infrastructure: Git, Docker, Kubernetes, CI/CD, MLflow, Weights & Biases, AWS, GCP
Research: Experimental Design, A/B Testing, Model Evaluation, Literature Review

EDUCATION

Indiana University Bloomington Aug 2021 – May 2023

Master of Science in Computer Science *Indiana, United States*

- Specialization: Machine Learning and Artificial Intelligence
- Relevant Coursework: Information Retrieval, Statistical Machine Learning, Natural Language Processing, Advanced Algorithms, Distributed Systems