Siddhant Kumar Jha

| @ siddhantjha365@gmail.com | LinkedIn | GitHub | Portfolio | Bangalore,India

Education

James Cook University — Townsville, Australia/Singapore

B.Sc. in Statistics and Data Science; ***GPA: 5.62/7.00*** — *Mar 2020 – July 2023*

Minor in Computational Econometrics; ***GPA: 6.18/7.00*** — *Mar 2020 – July 2023*

Skills

Languages: Python, R, SQL

Technologies: LLama Index, Langchain, MongoDB, Scikit-Learn, PyTorch, Keras, TensorFlow

Fundamental Skills: Bayesian/Non-Bayesian Statistics, Schema Restructuring, Structural Querying, LLMops,RAGops, Vector Databases and Hierarchical Knowledge Graphs

Work Experience

ParamNetwork LLC — Bangalore, India

Data Engineer — *Aug 2023 – Present, Full-time*

- Worked on the Reconstruction of Schema structures within pre-existing State machines to be better optimized for the constituent parts of the knowledge graph within the system.
- Worked on High-dimensional PCI analysis of sub-schema lexicon to be optimized for hierarchical semantic abstraction
- Designed Text-SQL query pipeline with grok and local LLMs , to enable dashboard level querying of systemic supply chain data
- Created document reading engines for invoices and checklists to read tables ,addresses into a structured schema output

Publications

- What is a black box problem, a mathematical look(Substack)
- The promise of KANs as interpretable machines
- Literature Review of recent advancements in Hypergraph Learning as it relates to optimizer (Paper Draft)

Projects

LLM Friendly Schema Reconstruction | *ParamNetwork*

- Restructured pre-existing state machine schemas to improve server efficiency for clients including Bosch-Travis, Taneira, mhrl, EXIM bank, etc., by eliminating an additional properties sub-schema.
- Organized table and column names based on semantic proximity in n-dimensional PCI clusters to enhance query efficiency and dynamically adapt to user input in a hierarchical querying system.

Document-AI

- Implemented a document processing pipeline using two LLMs: LayoutLM to segment document parts (e.g., tables, personal details) into images, followed by OCR to convert these into text strings. A reader LLM, such as OpenAI's model, along with Kor and Pydantic, is then used to generate structured outputs conforming to predefined schema structures based on document type.

SQL-Text Dashboard with Local LLM

- Developed a SQL-text dashboard using LangChain, integrated with vector database ChromaDB and a database dictionary, to enable natural language queries and real-time data visualization.
- Utilized local LLMs, including Qwen-2 and LLaMA-3 13B, with text-to-RAG chaining to enhance query processing and dynamic response generation.