



QUERY MANAGEMENT SYSTEMS FOR LARGE LANGUAGE MODELS (LLMs)

IITM Technology Available for Licensing

Problem Statement

- **Accurate tactile pulse replication** is essential for cardiovascular training, but current systems fail to mimic the heart's continuous, dynamic waveforms.
- **Traditional mechanical or pneumatic simulators** often ignore ventricular shape and volume, **producing unrealistic pulse patterns**.
- **Modern methods** improve realism **but cannot fully replicate** natural blood flow; hydro-electromagnetic actuation offers high-fidelity, human-like pulses.
- There is a **need** for a system that **uses hydro-electromagnetic actuation** to deliver precise and realistic pulse simulation for effective training.

Research Lab

PROF. PRABHU RAJAGOPAL
Dept. of Mechanical Engineering

TRL (Technology Readiness Level)

TRL 3 Experimental proof of concept

Technology Category/Market

Category- Artificial Intelligence & Machine learning/ Information & Communication Technologies (ICT)

Industry: Data management Industry, Cloud computing & service providers, Enterprise software/AI/SaaS, Data governance industry, Cloud Hosts, consulting & system integrators

Applications: Internal enterprise Knowledge management & search, Customer support & Help desk, Business Analytics, Manufacturing/IoT, e-commerce, Media, Telecommunication, Cloud & SaaS providers

Market report: The global market value of AI Data management in 2025 is 38.27 billion USD. By 2035, estimates show the market reaching up to 234.95 billion USD with a CAGR of 22.34%

Intellectual Property

- IITM IDF Ref 3238
- **IN 202541024879 Patent Application**

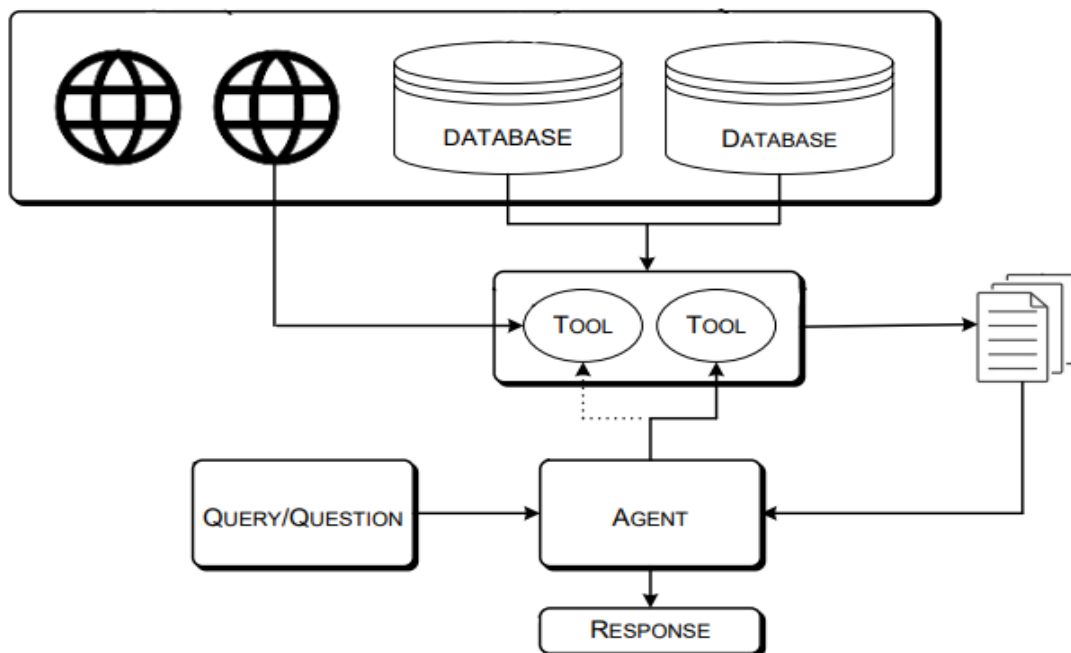


FIGURE: Query Management System for LLM

CONTACT US

Dr. Dara Ajay, Head TTO
Technology Transfer Office,
IPM Cell- IC&SR, IIT Madras

Email: headtto-icsr@icsrps.iitm.ac.in

ttooffice@icsrps.iitm.ac.in

Phone: +91-44-2257 9756/ 9719



Technology

The system receives user queries through an input engine that captures them accurately.

Queries are processed in stateless, meta-state, or multi-processing modes depending on complexity and context.

In stateless processing, the system generates similar queries, simulates documents, retrieves data chunks with two retrievers, and ranks them for LLM response.

Meta-state processing breaks complex queries into subqueries, selects tools for each, and synthesizes outputs into a response.

Multi-processing mode uses a state machine to determine which tools to call, their order, and coordinates outputs for the response.

The output engine evaluates retrieved information, selects the best mode and sequence, and guides the LLM to generate an accurate response

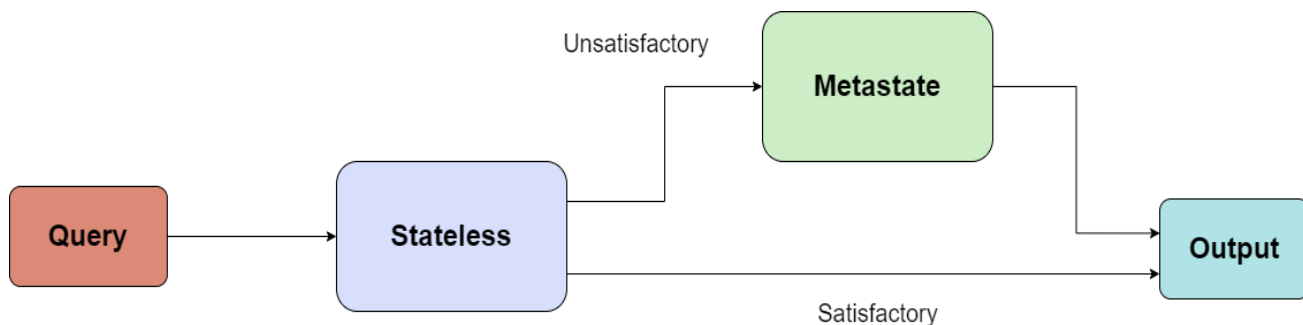


FIGURE: Pipeline of Query Management System for LLM

Key Features / Value Proposition

- The system enhances response accuracy, relevance, and adaptability by **using multiple query processing modes and real-time tool integration**.
- Its **multi-mode architecture adjusts processing strategies dynamically** based on the complexity and needs of each query.
- **Stateless mode** delivers fast, resource-efficient responses for simple or discrete queries **without requiring contextual continuity**.
- **Meta-state mode** effectively **handles complex, multi-layered queries** by decomposing them and synthesizing outputs from multiple tools.
- **Configured iteration limits** ensure a balance between response **accuracy, computational efficiency, and processing speed**.

CONTACT US

Dr. Dara Ajay, Head TTO

Technology Transfer Office,
IPM Cell- IC&SR, IIT Madras

Email: headtto-icsr@icsrpis.iitm.ac.in

ttooffice@icsrpis.iitm.ac.in

Phone: +91-44-2257 9756/ 9719