AI-Driven LLM Automation Agent for Data Retrieval

Anjali Garg - agarg37@uic.edu , Basil John Milton M - bmuth5@uic.edu , Karishma Kamble - kkambl2@uic.edu, Mohan Kumar Areti - maret3@uic.edu , Yun(Hailey)Wu - ywu247@uic.edu
Faculty mentor: Prof. Fatemeh Sarayloo (fsaraylo@uic.edu)

Digital copy of this poster

Introduction

Founded in 1980, CCC is a technology leader providing solutions for insurers, automotive manufacturers, repair facilities, and more, with extensive claims data and a strong network in the auto industry.

Project Overview:
CCC aims to develop an agent-based engine using Large Language Models (LLMs) for automated vehicle data extraction and valuation. This tool will support CCC agents by automating data retrieval and presenting data in user-friendly formats.

300+ Insurance Partners

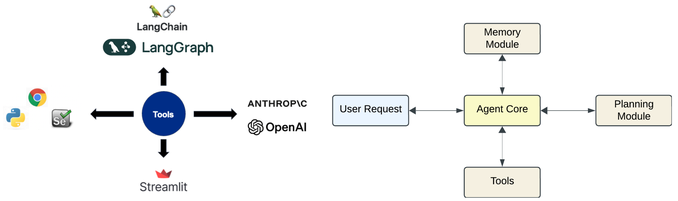
27,000+ Collision Repair Facilities

Data-Driven Insights

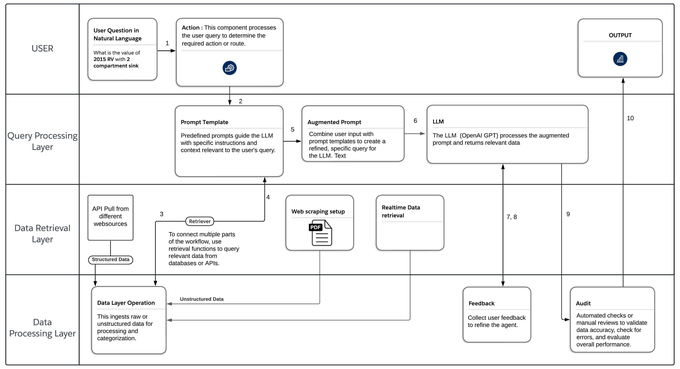
Business Problem

1. **Accurate Data Retrieval:** Automate the retrieval of up-to-date vehicle information from diverse online sources.
2. **Support for Vehicle Valuation:** Aid agents and adjusters by reducing manual effort in vehicle valuation processes.
3. **Adaptability:** The system must seamlessly work with both structured APIs and unstructured data (e.g., web content).

Methodology



How to Build Efficient AI-Based Automation

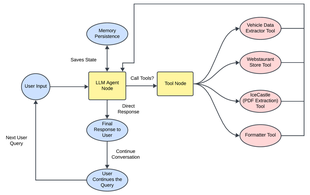


Data Collection:
Description: The multi-agent system gathers data from diverse online sources through web scraping. This approach supports flexibility and scalability in data acquisition.
Key Tools: The system employs customized web scraping methods to extract information from both structured and unstructured sources, ensuring comprehensive data collection across varied websites.

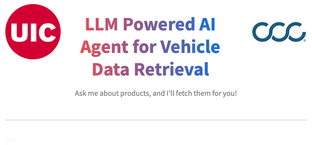
Data Processing:
Description: Once collected, the data undergoes processing to ensure it meets CCC's standards for relevance and accuracy. LLM integration refines and formats the data to align with CCC's specific use cases.
Key Tools: The LLM (OpenAI GPT & Anthropic) is utilized to interpret, clean, and organize the data, preparing it for efficient use in decision-making and analysis.

User Interface:
Description: The user interface is designed to present query results in formats such as tables or lists, tailored to user preferences. This streamlined UI simplifies access to organized data.
Key Tools: Streamlit powers the UI, offering an interactive and user-friendly platform where users can access and explore data outputs efficiently.

Result



User Interface



Business Output

1. **Data Retrieval Speed:** Reduced retrieval time from 10 minutes to 5 minutes.
2. **Workflow Efficiency:** Increased cases handled daily from 100 to 140.
3. **Cost Reduction:** Lowered data acquisition cost by 20%.
4. **Error Rate:** Reduced from 5% to 3.5%.
5. **Claim Processing Time:** Decreased by 25%.
6. **System Scalability:** Improved system scalability from 50% to 87.5%.

Functionality Developed

User-Friendly Interaction

Intelligent Query Processing: LLM Agent Node

Context Retention: Memory Persistence

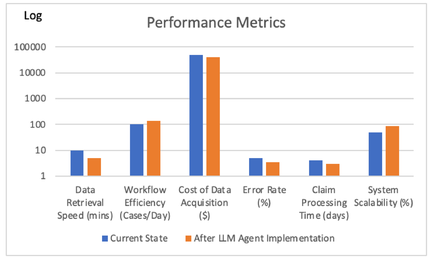
Dynamic Tool Integration: The Tool Node
- Vehicle Data Extractor Tool
- Webstaurant Store Tool
- IceCastle (PDF Extraction) Tool
- Formatter Tool for structured data output

Quick Response: For straightforward queries, the LLM responds immediately.

Detailed Answers: After gathering information from tools, the agent provides a complete answer.

Continued Interaction: Users can follow up with additional queries.

Conclusion



The AI-based automation agents effectively streamline CCC's workflow by providing accurate vehicle data from multiple sources, reducing manual effort, and improving efficiency across the board. This solution ensures better resource utilization, lowers operational costs, and enhances system scalability, paving the way for more effective and data-driven decision-making.

Integration Potential
- **Data Enrichment:** Real-time updates to vehicle databases.
- **Advanced Analytics:** Structured data for insights and trends.
- **AI Decision Support:** Foundation for predictive analytics.
- **Real-Time Pipeline:** Live data streaming across CCC systems.

Analytical Problem

1. **Data Aggregation:** Efficiently gather and consolidate data from multiple data sources.
2. **Flexible Multi-Agent System:** Build a dynamic, multi-agent framework to handle varying data sources and requirements.
3. **Accuracy and Efficiency:** Ensure high accuracy in data extraction and reduce processing time to support CCC's workflow.

Challenges and Solutions

1. **Data Integration:** Combined structured APIs and web scraping for unstructured data using a multi-agent system.
2. **Legal Compliance:** Prioritized API usage to align with terms and used web scraping only on publicly accessible data.
3. **Dynamic Content Handling:** Utilized Selenium for interactive sites and developed parsers for diverse data formats.
4. **Data Accuracy:** Integrated LLMs for data validation and formatting, ensuring consistency with CCC's standards.
5. **UI Complexity:** Designed a modular, adaptable Streamlit UI to support varied use cases and deliver a user-friendly experience.
6. **Scalability:** Optimized multi-agent workflows and cloud storage to efficiently handle large data volumes and requests.

Acknowledgement

Anjalli Garg
/anjaligarg1010

Basil John Milton M
/bjmm1296

Karishma Kamble
/karishma-kamble-2809

Mohan Kumar Areti
/mohan-kumar-areti-a27298178

Yun (Hailey) Wu
/hailey-wu-32408922b