UNITED WHOLESALE MORTGAGE

Design Proposal

Prepared for: O’Reilly Media Inc, Architectural Katas Q4 2025

Prepared by: UShore Challengers, Technology Innovation Group

October 22, 2025

EXECUTIVE SUMMARY

Objective

The company MotorCorp is observing a customer churn due to two main challenges. The assignment is to address the same in an efficient way through a Generative Artificial Intelligence way.

Goals

The goal is to enhance the customer satisfaction levels of a vehicle rental and there-by boost the count of customers.

Solution

The procedure was initiated to understand the reasons of the challenges, formulate the practical solution proposals, derive the business use-case out of those that helps to build an universal multi-modal Generative AI architecture. This helps to build a roadmap in order to execute the proposed solutions.

Project Outline

1. Introduction

The MotorCorp is a vehicle rental company who own the electric vehicle categorized as bikes, scooters and cars. Each of the vehicle has a device that captures the customer and geo-location data and delivers to the central repository. As per the historical observations it is identified that the firm is facing tough challenges with regard to allocating a priority plan to deliver a service person for the battery replacement as well as the other dimension of vehicle not being returned at the right locations. For the same, a Generative AI solution is proposed to better serve the customers and gain the market reputation with a societal responsibility. Thus, the end to end generative AI solution also highlights the need of responsible AI and the system monitoring frameworks respectively.

2. Current State

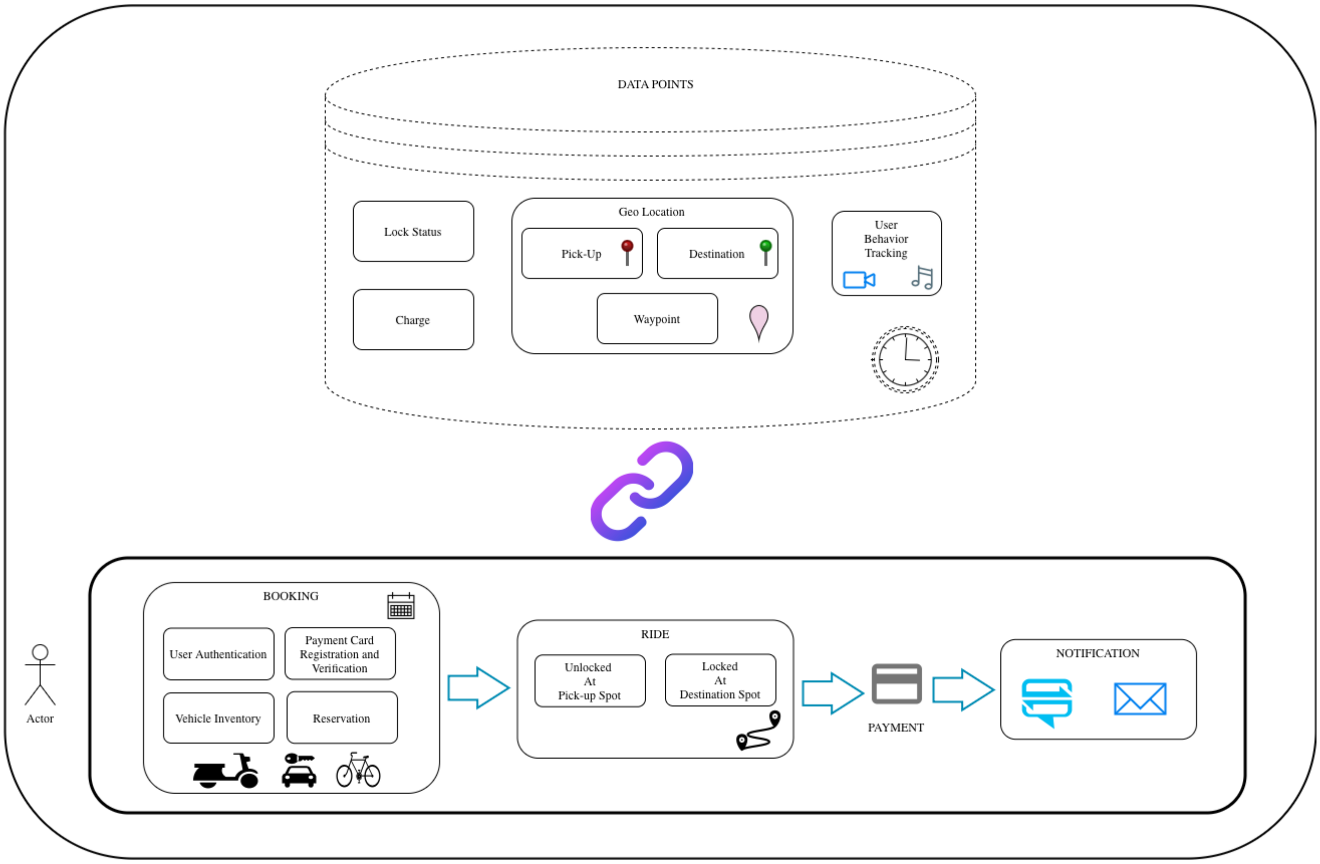

Figure 1: Current System

The current rental system of the firm is not open-source and the above is only an assumption. Especially about the data points that the vehicle device transmits to the database. It is for this purpose that the usage of GPS, video/ audio capture was populated in the above figure to mention the required set of data records to effectively address the business use-case. The system is very plain that utilizes the booking platform to initiate the reservation with on demand payment registration. After which the ride begins with live tracking and upon the return of a vehicle the fare will be debited from the customer account and the same will be notified to the customer either via SMS message or electronic mail as per the choice.

3. Analysis and Solution Proposals

The project scope is to explore the reasons that cause the customer churn and thus formulate the solution proposals. The figure 2 titled 'challenge and solution proposal' categories the situations that impact the escalation of a negative. For the same, the solutions were drafted to address the drawbacks of the current state that leads for a poor business operations and thus the financial loss. The block concerning the solution proposals has the derived GenAI use cases to construct the future state architecture diagram. Although the use-cases are multiple the design proposal is universal for all of those provided the usage of large language models with retrieval augmented generation mechanism holds true.

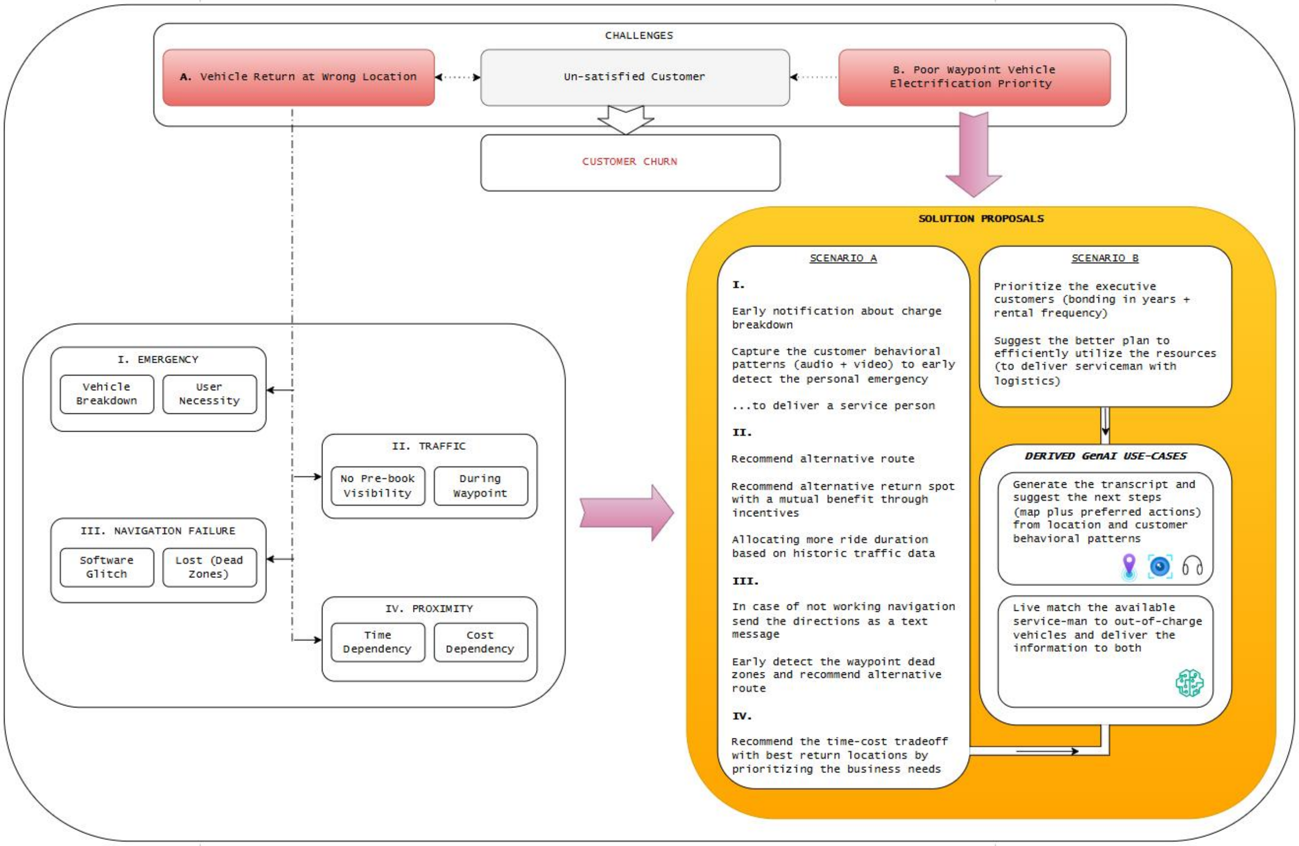
CHALLENGES
A. Vehicle Return at Wrong Location
Un-satisfied Customer
B. Poor Waypoint Vehicle Electrification Priority
CUSTOMER CHURN
I. EMERGENCY
Vehicle Breakdown
User Necessity
II. TRAFFIC
No Pre-book Visibility
During Waypoint
III. NAVIGATION FAILURE
Software Glitch
Lost (Dead Zones)
IV. PROXIMITY
Time Dependency
Cost Dependency
SOLUTION PROPOSALS
SCENARIO A
I.
Early notification about charge breakdown
Capture the customer behavioral patterns (audio + video) to early detect the personal emergency
...to deliver a service person
II.
Recommend alternative route
Recommend alternative return spot with a mutual benefit through incentives
Allocating more ride duration based on historic traffic data
III.
In case of not working navigation send the directions as a text message
Early detect the waypoint dead zones and recommend alternative route
IV.
Recommend the time-cost tradeoff with best return locations by prioritizing the business needs
SCENARIO B
Prioritize the executive customers (bonding in years + rental frequency)
Suggest the better plan to efficiently utilize the resources (to deliver serviceman with logistics)
DERIVED GenAI USE-CASES
Generate the transcript and suggest the next steps (map plus preferred actions) from location and customer behavioral patterns
Live match the available service-man to out-of-charge vehicles and deliver the information to both

Figure 2: Challenges and Solution Proposal

The collected datapoints in section 2 associates the data sources defined in section 4.3. The solution proposal block utilizes the A, B, I, II, III, IV wherever applicable and this is from the other blocks respectively.

4. Future State

4.1 Why AI and not ML?

It is very cost effective to utilize the vastly available open-source ML models to leverage the projected use cases. However there occurs a trade-off to compromise with the broader perspective of not being able to utilize the benefits that the large language models offer in terms of content relevance memory and latency with high throughput to conclude the predictions. Also it involves huge work efforts to create data labels and train the models as per the necessities. Thus, a readily available LLM models will better guide the scenario to collect a through a generative response and in real-time.

4.2 Unstructured Data Scoping

4.2.1 Regulatory Acts

The importance to abide by the GDPR and other EU acts is very widely applicable to address the social and environmental factors. Thus, the agenda of this project scope highlights the necessity of governance as in section 4.3. This provide the responsive consideration to collect the solution offerings from a large language model.

4.2.2 About Data Sources

The usage of retrieval augment generation mechanism assumes the data to reside in semantic search adaptable database. The prospect cover the dimension of image, video and text data to understand the customer live behavioral patterns and attach it to the historical context to gather the GenAI responses.

4.3 GenAI Solution Architecture

Modules: Front-end, DevOps, Back-end, Data Sources, Infrastructure, LLM observability, Responsible AI

Front-end: The web application or an edge device serve to populate the live prompt feed analytics coming directly from an actor, who in this case is the auto-data (example behavioral patterns, service person timetable, etc..) from a customer. The collective responses highlighted as a track record in the same platform is after an observation by a customer to take proactive decisions/steps. This help to reduce the customer churn through early detection of an issue that might lead to dis-satisfaction.

DevOps: The pipeline to bind the front-end and the backend environments. The docker file contains the sequence of steps to collect a LLM response from a live (without customer intervention) prompt.

Back-end: The orchestrator layered with a model context protocol serves as the core to interact with the several agents. And the interaction with them is via an A2A protocol (Agent to Agent). The agents are namely the configuration file, model call, semantic vector store database, monitoring tool, responsible AI dashboard.

Infrastructure: The several of the infrastructure as a code principles can usually be automated and the figure highlights the necessary to leverage the published modules and connect them for parallelism.

LLM Observability: The dashboard projects the metrics in the dimensions of prompt, model, generic **Kubernetes** API usage. This is part of an automated backend pipeline as the live data is required for periodic update.

Responsible AI: The automated governance of the entire system via backend pipeline is through the ISO standardization. Each of the sub-module inside the framework has their own fundamentals that needs to be connected to have a broader vision of responsiveness which adds a societal value and there-by gains the market capitalization.

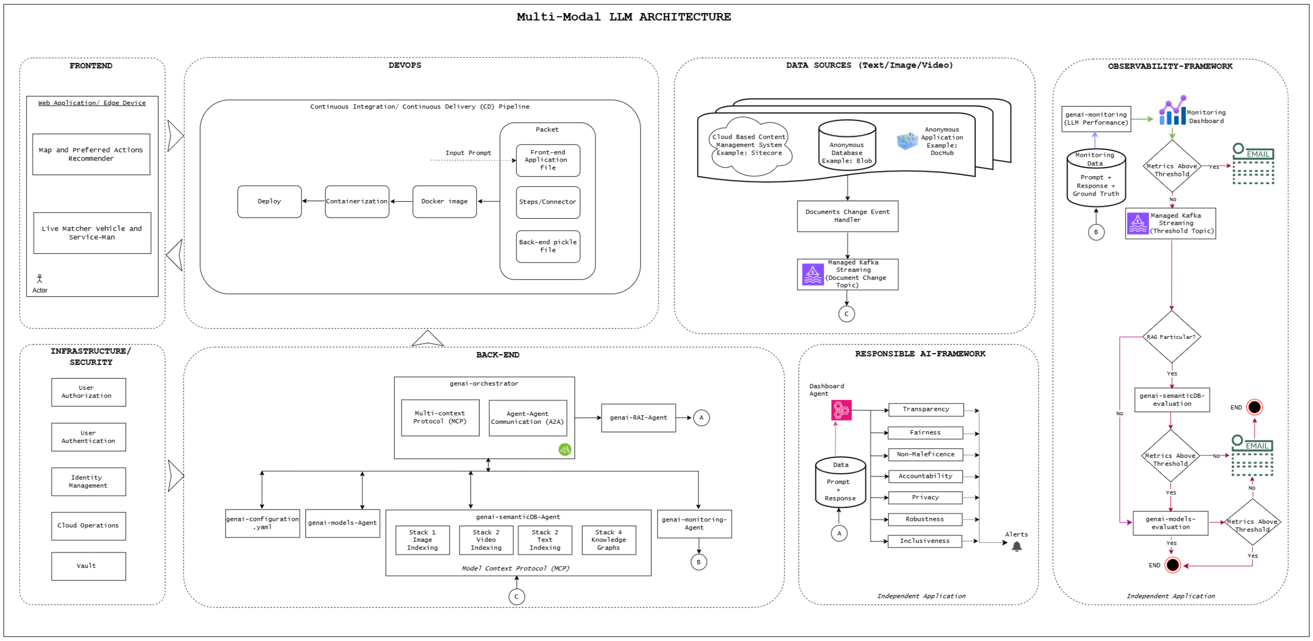

Figure 3: Multi-modal GenAI Architecture

5. Conclusion

The project outlined initiated the necessity to perform the exploratory data analysis for deriving the GenAI business use-case out of the proposed solutions under the mentioned scenarios. Ultimately the built architecture/design diagram conceptualize the resolution factors of the identified issues.

References

[1] Kishor Datta Gupta, Mohd Ariful Haque, Hasmot Ali, Marufa Kamal, Syed Bahauddin Alam and Mohammad Ashiqur Rahman. CONTINUOUS MONITORING OF LARGE-SCALE GENERATIVE AI VIA DETERMINISTIC KNOWLEDGE GRAPH STRUCTURES, 2025.

[2] Andrés Herrera-Poyatos, Javier Del Ser, Marcos López de Prado, Fei-Yue Wang,Enrique Herrera-Viedma, and Francisco Herrera. RESPONSIBLE ARTIFICIAL INTELLIGENCE SYSTEMS:A ROADMAP TO SOCIETY'S TRUST THROUGH TRUSTWORTHY AI, AUDITABILITY, ACCOUNTABILITY, AND GOVERNANCE, 2025.

[3] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. Retrieval-Augmented Generation for Large Language Models: A Survey, 2024.