FOUNDATION FACTORY

PUTTING GENERATIVE AI INTO BUSINESS

SELLMA

Streamlined Event Lake for Multi-Agent Action Architecture

a whitepaper from:
Prof. Dr. Peter Gentsch, B.I.G. GmbH
Dr.-Ing. Christian Gilcher, Value Cloud GmbH

Table of Contents

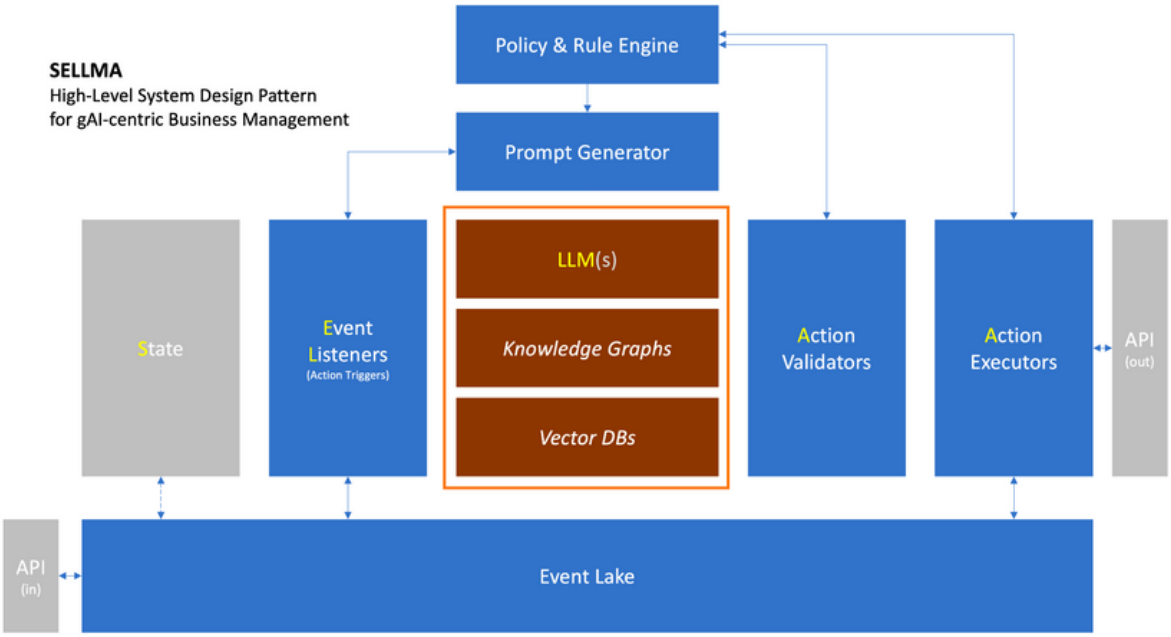
Abstract and Introduction

Abstract

The Streamlined Event Lake for Multi-Agent Action Architecture (SELLMA) is a novel software architecture pattern designed to enable organizations to leverage the power of large language models (LLMs) in a systematic, context-aware, and policy-driven manner. This whitepaper presents the various components and mechanisms that make up the SELLMA architecture, illustrating how they interconnect to provide a comprehensive and flexible solution for event-driven, AI-powered decision-making processes.

Introduction

As organizations increasingly adopt AI technologies to improve decision-making and business processes, there is a growing need for efficient and effective ways to integrate these technologies into real-life business processes. Large language models (LLMs) offer significant potential in this regard, thanks to their ability to process and generate human-like text based on a wide range of input data. The SELLMA architecture pattern addresses this need by providing a structured approach to incorporating LLMs into software systems for next-generation business operations.



SELLMA Architecture Components

The SELLMA architecture pattern consists of the following components:

Event Lake
The Event Lake is a central repository that holds all events generated within a particular organization. These events can include a wide variety of data types, such as user interactions, system logs, and external data feeds. The Event Lake provides a unified, scalable, and easily accessible data storage solution that can be queried to retrieve relevant information for further processing.

State Representation Component
The State Representation Component (SRC) is responsible for generating state representations of the system based on the events stored in the Event Lake. The SRC enables the system to maintain a consistent view of the current state of the organization and its environment, facilitating the identification of relevant actions and decision-making processes.

Event Listeners
Event Listeners are specialized components designed to perform real-time analytics on the event streams generated within the system. These listeners are configurable and can be tailored to detect specific event constellations or patterns of interest. Upon detecting such a pattern, an Event Listener can trigger actions, which are then processed by the LLMs.

Large Language Models
The core of the SELLMA architecture consists of one or multiple LLMs, each with its specific domain knowledge and capabilities. When an action is triggered, the relevant context is extracted from the Event Lake and provided as input to one or more LLMs. LLMs are being complemented by local knowledge which can be handled through Knowledge-Graph type structure and/or Vector Databases. The combination and interaction of these structures combine a new type process intelligence that allows the system to run through business processes not in a hard-coded, pre-defined way bur rather find its own way, within the boundaries that are being provided by the Policy Engine.

Policy Engine
The Policy Engine is a critical component that defines and enforces the rules governing all allowed actions that the Event Listeners can trigger and the subsequent behavior of the LLMs. Such rules are defined on a meta-level. By providing a configurable and extensible mechanism for specifying these rules, the Policy Engine ensures that the SELLMA architecture remains adaptable and compliant with organizational policies and requirements.

Action Handlers
Action Handlers are responsible for interpreting the outputs generated by the LLMs and translating them into specific actions to be executed within the system. These handlers serve as the bridge between the AI-generated insights and the actual implementation of the recommended actions in the real world. They can be best though of as RPA-style software robots.

Action Validators
Action Validators play a crucial role in maintaining the quality and compliance of the actions executed by the system. By cross-referencing the outputs of the LLMs and the rules defined in the Policy Engine, these validators ensure that the generated actions meet the required quality standards and adhere to the established policies.

6

Implementation Considerations and Best Practices

Scalability and Performance
The SELLMA architecture is designed to accommodate varying workloads and data volumes, with components capable of scaling horizontally or vertically as needed. To ensure optimal performance and efficiency, it is crucial to implement caching and indexing strategies for the Event Lake, as well as load balancing and parallel processing for the LLMs.

Security and Privacy
Security and privacy concerns are being addressed throughout the implementation of the SELLMA architecture. This includes securing access to the Event Lake, protecting sensitive data, and implementing proper authentication and authorization mechanisms for the various components. Additionally, it is essential to consider data privacy regulations and ensure that the system complies with applicable laws and standards. This includes the option for running local LLMs that are based on fully auditable training data.

Monitoring and Logging
Monitoring and logging are essential for maintaining the health and reliability of the SELLMA architecture. Implementing a comprehensive monitoring solution will enable the detection of potential issues, performance bottlenecks, and security threats. Additionally, logging provides valuable information for debugging, auditing, and analyzing the behavior of the system.

Interoperability and Integration
The SELLMA architecture has been designed to seamlessly integrate with existing systems and software throughout a customers organization. This includes supporting standard data formats and communication protocols, as well as providing well-documented APIs for interacting with the various components of the architecture. The Event Lake can be seen as a node within a Corporate Event Mesh structure.

Conclusion

The Streamlined Event Lake for Multi-Agent Action Architecture (SELLMA) provides a comprehensive and flexible solution for organizations seeking to harness the power of large language models in a systematic and policy-driven manner, in order to incorporate a new level of productivity for their existing workforce. By incorporating real-time analytics, context-aware decision-making, and AI-powered action execution, the SELLMA architecture enables organizations to improve their business operations, while unlocking the full potential of AI-based insights and system-embedded intelligence.

This paper is currently in DRAFT modus and work-in-progress. A final version is envisaged to be published in September 2023.