Bridging Natural Language and Reservoir Simulation: A Configurable Architecture for Democratizing Subsurface Modeling

CLARISSA: Conversational Language Agent for Reservoir Integrated Simulation System Analysis

[Authors removed for blind review]

Abstract

Reservoir simulation remains one of the most technically demanding disciplines in the energy sector. Despite decades of software evolution, the fundamental barrier persists: translating engineering intent into executable simulation input requires specialized expertise that takes years to develop. This paper presents the architecture of CLARISSA (Conversational Language Agent for Reservoir Integrated Simulation System Analysis), a system designed to bridge the gap between natural language interaction and domain-specific simulation syntax.

We propose a multi-stage pipeline that decomposes translation into discrete, validated steps: speech recognition, intent classification, entity extraction, asset validation, syntax generation, and physics-based verification. The architecture addresses deployment heterogeneity through configuration-driven design, enabling air-gapped installations alongside cloud-native deployments. We present a domain mesh data architecture with purpose-built storage for different data types, and a hybrid synchronous-asynchronous communication pattern for long-running simulation workflows.

CLARISSA separates concerns across three architectural pillars: LLM-based reasoning for planning and explanation, reinforcement learning for skill acquisition based on simulator-derived numerical outcomes, and neuro-symbolic governance for constraint enforcement and approval workflows. This separation enables independent auditability while maintaining system coherence.

This paper contributes reference architecture patterns applicable beyond reservoir simulation to any domain requiring natural language interfaces to complex technical systems. The combination of multi-stage translation pipelines, configurable infrastructure, and event-driven communication provides a template for democratizing access to specialized technical capabilities.

Keywords: *reservoir simulation; natural language processing; microservices architecture; domain-specific languages; configurable systems; event-driven architecture; AI agents*

Target Audiences

Software Engineering: Microservices patterns, Data Mesh architecture, Configuration-Driven Design, Event-Driven Communication

AI/ML in Energy: Multi-stage NLP Pipeline, Model Abstraction Layer, Reinforcement Learning from Simulator Feedback

Petroleum Engineering: ECLIPSE Integration, Reservoir Simulation Workflows, Physics-Based Validation, Field Asset Management

NLP Community: Domain-Specific Translation, Multi-Stage Validation Pipelines, Confidence-Based Clarification Strategies