Bridging PLEXOS and AI: A Multi-Agent, LLM-Based Framework for Transparent and Compliant Energy Planning

MSc Research Internship

Masters in Artificial Intelligence

Suvajit Lodh

Student ID: x23356987

School of Computing

National College of Ireland

Supervisor: Rejwanul Haque

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Suvajit Lodh		
Student ID:	x23356987		
Programme:	Masters in Artificial Intelligence	**Year:**	2024-2025
Module:	MSc Research Internship		
Supervisor:	Rejwanul Haque		
Submission Due Date:			
Project Title:	Bridging PLEXOS and AI: A Multi-Agent, LLM-Based Framework for Transparent and Compliant Energy Planning		
Word Count:	6590		
Page Count	21		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Suvajit Lodh
Date:	14 September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on the computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Bridging PLEXOS and AI: A Multi-Agent, LLM-Based Framework for Transparent and Compliant Energy Planning

Suvajit Lodh

x23356987

Abstract

PLEXOS is a simulation tool which has been traditionally used in energy infrastructure planning. Although strong, these tools are computationally intensive, opaque, and require expert interpretation, and therefore are limited in their responsiveness to quickly-changing policy and market conditions. As EU AI Act would require high-risk AI systems to be auditable and explainable, new methods that would be simultaneously efficient, explainable, and AI-compliant are of high demand. The idea to combine Large Language Models (LLMs) and energy system simulators to create an AI guidance overlaid over an energy system simulation is developed and explored in the thesis. The framework will automate the scenario, parameter estimation, simulation execution, KPI extraction, and so complete the scenario analysis faster, to increase transparency, and support superior decision-making. The architecture is based on Autogen multi-agent where coding is done specially in simulation management, data analysis, dynamic code execution, and error correction. A governance layer provides a bridge to regulatory compliance by means of audit trails and natural language justifications. Parallel computing is also provided through the use of high-performance facilities to execute a series of analytical workflows. Evaluation indicated that the system was able to perform analytical cycles in a few seconds consistently. Outputs were traceable and transparent and with a rational explanation of the same was improving user trust and readiness to comply. The system is much more efficient, transparent and aligned to regulations when compared to the traditional methods of planning. Keeping within an integrated architecture, it introduces a new standard of explainability and governance meta-mechanisms within AI-driven energy planning. The accuracy of the framework is limited by the accuracy of the underlying simulation models and reasoning of optimal paths that may still yield occasional output errors. Real-time adaption to the market and generalization to a broader domain are open problems of further research work.

1 Introduction

The modern energy system of the world is undergoing a radical change that is caused by the growth of energy demand, decarbonisation requirement, and a faster pace of renewable energy technology implementation. New opportunities and new challenges that are beyond precedent are presented by the changes, and are present in front of the planners of energy-related infrastructures. Historical approaches that may be based upon obsolete modelling practices and manual processes utilising human effort are poorly designed to deal with the complex nature of variable renewable generators, distributed energy resources, and the new market structures. The deluge of information issued by the contemporary electricity, gas and water systems, coupled with the multifaceted interaction between them, requires a much more sophisticated analytical hardware. In the past, simulation tools like PLEXOS have provided essential support to energy planners, who have been allowed to analyse the activity at the deepest level of generation, transmission and market performance in the energy sector. These simulations also provide important information on system behaviour in a variety of situations and thus are used to guide investment decisions, policy formulation and real time operations. However, not only the skillsets necessary for maximum use but also the amount of time consumed make classic simulations less adaptive to emerging data or unexpected events. The potentially disruptive transition offered by artificial intelligence (AI) and, more-recently, multi-agent systems (MAS) is likely to undertake sophisticated analytical work, enhance the quality of decisions, and, in general, the efficiency of energy infrastructure planning. The current study focuses on how the more sophisticated AI methods may be harnessed to build a more dynamic, intelligent, and robust platform to operate within the mazes of modernized energy structures.

1.1 Problem Statement

Some of the major limitations of the traditional planning method which is premised on manual analysis of data and fixed parameter simulation models have been put in place. First, the repetitive task of performing the scenario analysis and optimisation so vital to sound planning is highly elongated by the manual actions imperative, causing the consequences of time lapse that may degrade the quality of decisions, and generate non-optimal results. Such incompetency can contribute to losses of investment, insufficient use of investments and can fail to reorganize on time to deal with sudden shocks and policy shifts.

Moreover, the secrecy that surrounds some advanced analytical products, as well as the absence of any transparent decision-making mechanisms in traditional systems, creates significant issues, especially due to the emergence of new regulatory requirements. Indeed, the Artificial Intelligence Act (EU AI Act) of the European Union will impose strict accountability requirements on high-risk AI systems, where AI-powered energy infrastructure planning tools would probably fall. The existing approaches often neither have built-in facilities to express determinate reasons behind their suggestions nor do they leave papers focusing on their analytical judgments that can be audited. Such gaps not only create the challenges in being compliant with regulations but also discourage stakeholder confidence and impossibility to diagnose and repair the flaws properly. As such, a self-learning analysis mathematic is urgently needed, not only as a means of simplifying the process of model see-saw decision-making when analysing complex energy-systems, but also as a means of providing transparent, traceable and compliant decision-making intrinsically, minimising the inadequacies of current methods and addressing with a new approach the rapidly changing regulatory and technology environment.

1.2 Research Objectives

The present thesis examines the progress of energy infrastructure planning by developing and testing a new system of multi-agents. The primary research question is to prove the consistency and viability of the collaborative framework based on artificial intelligence that allows incorporating more sophisticated simulation and involving intelligent processing of data and transparent decision making. The following study aims to achieve the following objectives in order to achieve its overall objective:

- Development of an effective multi-agent system that will be able to organize the work of complex analytical processes in energy infrastructure planning, being modular, scalable, and tolerating faults.
- New Information science methods implemented on the multi-agent environment based on the viewpoint on automated collection, cleaning, and processing of large amounts of energy data through intelligent information and Key Performance Indicators (KPIs).
- Ensuring transparency and traceability of the AI-powered analytically driven process that reflects through mechanisms of self-reflection and generation of rationales and audit trails, complying with regulatory regulations, one of which is the EU AI Act.
- Optimization of performance and scaling of the system in a way which guarantees that the system can be applied in computationally intensive simulations and data processing and multiple analytical runs can be executed concurrently.
- Provision of a friendly interface which allows one interact intuitively with the complicated underlying models and enables the user i.e. the energy analyst to make the most of the capabilities of the system through prompt-based input and clear visualization of the results obtained

The successful attainment of these goals results in a viable theory and practically sound solution that can enhance the state-of-the-art in the efforts to plan AI-driven energy infrastructure and prospectively make the process more sustainable, efficient, transparent, and adaptable to challenges in the future.

1.3 Contributions

The thesis contributes to the four areas in artificial intelligence, multi-agent systems and the planning of energy infrastructure, as follows:

- New Multi-Agent Architecture: We introduce and apply a new multi-agent system framework that is very useful in the context of complex energy infrastructure planning. Designed on Autogen, this architecture shows how task-specific AI agents can be used in combination to tackle complex analysis problems, such as scenario generation, data analysis and compliance examinations.
- Integrated LLM-PLEXOS Framework: We are able to propose a novel integration of Large Language Models (LLMs) and the standard PLEXOS model that is used throughout the energy modeling industry. This integration makes it possible to communicate with sophisticated energy models in ordinary language, thus automating the task of synthesising and interpreting scenarios, allowing the intelligence of that tool to be bridged between an energy model and a human planner.

- Transparent and Regulatory Compliant AI System Design: The other aspect which is vital is the emphasis on transparency and compliance of an AI system with regulations. We show how the systems can be used to produce natural-language explanations of AIdecisions, support a full audit trail (e.g., Langfuse traceability), and introduce a GovernanceAgent to make sure they are compliant with a range of regulations, including the EU AI Act.
- Automated Data Analysis and Insight Generation: The study provides robust data to extract, transform, analyze and provide insights to large-scale energy simulation outputs.
- Performance and Scalability Optimization: We highlight a performance-optimized and highly scalable system that can support computational intensive simulations and run multiple analytical jobs in parallel. These are efficiency strategies on resource allocation, I/O batching and distributed data management.
- For a practical application and evaluation: the thesis puts in place a practical implemented system that addresses real world issues in energy infrastructure planning. The approach has strong testing and research processes to determine the performance, dependability, and effectiveness of the system, providing the practical data about its usefulness.

The contributions taken together thereby contribute to the state-of-the-art in AI-based decision support in critical infrastructure and outline the blueprint to future intelligent systems in complex domains.

1.4 Thesis Structure

The rest of this thesis is as follows:

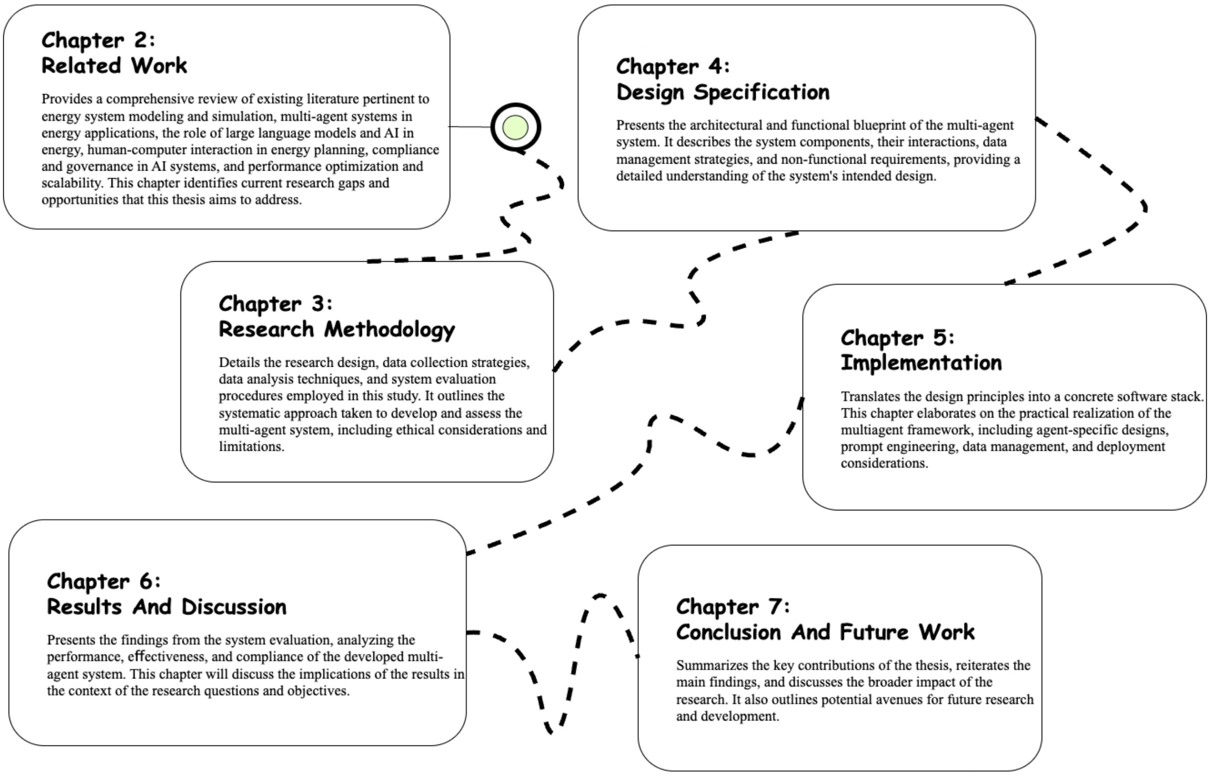


Figure 1: Thesis roadmap with structure of the chapters

2 Related Work

The topic of modeling and decision-support in energy systems has been developing in quite diverse ways over the last few decades, including optimization-based approaches as well as artificial intelligence-based tools. The chapter examines the state of the art with reference to various aspects relating to this thesis, which include simulation tools like PLEXOS and PYPSA which are used in infrastructure planning, multi-agent systems application in energy management.

2.1 Energy System Modeling and Simulation

Simulation and modeling of the energy system form the pillars of how we conceive and plan the intricate energy base. The tools allow analysts to make energy demand predictions, automate resource distribution optimization, and provide an assessment of how different policy interventions can affect the outcome. PLEXOS is one of the most famous and powerful combined energy models, which has widely been used in the electricity market analysis and long-term energy planning (Sosef, 2017)(Proquest.com, 2023)(Hamdi, El Salmawy and Ragab, 2024). PLEXOS offers the ability to specify generation, transmission and/or market operations and to independently evaluate the full system through extensive modeling analysis (www.energyexemplar.com, n.d.). Researchers have used PLEXOS to study energy system performance at large percentages of renewably generated energy with consideration of operational limits in long-term planning (Hamdi, El Salmawy and Ragab, 2024). As an example, Hamdi et al. (2024) ran the PLEXOS power system simulation to simulate the Egyptian power system for a highly renewable penetration (Hamdi, El Salmawy and Ragab, 2024). Following the same line of thought, Bagga (2023) used PLEXOS to model the production costs to evaluate how an expanded parameterization of combustion turbines would affect grids with different renewable energy fractions (Proquest.com, 2023).

In addition to PLEXOS, the other common simulation tools also include EnergyPLAN used to conduct a holistic energy system analysis looking into all the sectors of energy (Widl et al., 2022). Selection of the modeling tool is largely dependent on a given research question, level of detail needed, and scope of energy system being considered. Optimization tools are often combined with these models to determine which avenues of energy system development may be defined as the most cost-effective or environmentally friendly.

2.2 Multi-Agent Systems in Energy Applications

Multi-Agent Systems (MAS) provide a disaggregated and adaptable management of a complex system, and are therefore, quite favored to manage modern dynamic and interconnected energy infrastructures. The MAS entail the use of several autonomous agents which assist one another in attaining a shared objective, and so they provide that degree of autonomy and collaboration that was not possible in earlier energy optimization methods (SmythOS, 2024). Such systems are changing the nature in which we should treat energy management, such as smart grids, real-time energy management and control systems (Feng et al., Fawaz et al., 2025). As an example, in a study by (Feng et al., 2025), the combination of MAS and enhanced AI algorithms in the implementation of self-healing subway power supply systems were discussed, and the importance of these algorithms regarding fault diagnosis and isolation in this application was described (Feng et al., 2025). In a similar way, Pinto et al. (2020) invented MARTINE, a real-time multi-agent infrastructure focused on energy, using a synthesis of AI methods to manage and operate energy infrastructure (Pinto et al., 2020).

Other areas which use MAS in energy will encompass energy routing protocols within the confines of Energy Internet in which intelligent agents can be used to calculate power flow and maximize security (U.S. Department of Energy, 2024). The fact that decentralized structure is adopted in MAS allows the system to become resistant since when a single agent breaks down, the system does not easily break down as well. In essential infrastructure like energy grids where resiliency and reliability are a priority, this can be very handy. Additionally, MAS allows decentralization of decision-making thus providing local optimality and global effciency of the systems.

2.3 Large Language Models and AI in Energy

The application of Large Language Models (LLM) and, more generally, AI technologies to the field of energy is rapidly changing. Predictive capabilities that are powered by AI are increasing grid resilience when they predict and control outages (Huang et al., 2024). LMs, and more recently, LMs with high natural language understanding and generation capabilities, are being considered to take part in the analysis of data, the generation of prompts in power systems, among others (Husom et al., 2024). Huang et al. (2024) studied how the detailed analysis of data in the power systems is possible with the large models techniques reliant on the prompt generation and information on data to ensure the effcient management of the energy (Husom et al., 2024). The adoption of LLMs, however, brings into consideration other factors, in this case the energy consumption related to the use of the model. Initial investigations of the energy cost of prompting, profiling the energy consumption of LLM inference, and examining diאffenent prompt features and their effect on energy cost are ongoing (You, 2025, Pressley, 2024). This raises the importance of energy solutions to AI, in particular, to sustainable energy infrastructure. In addition to LLMs, the use of AI in grid planning is altering the way grids are planned by utilizing the wealth of available data collected by smart meters, weather forecasting, and historical behaviour to find the optimal grid behaviour (AI, 2025). Predictive maintenance (reduced downtime, operational efficiency in energy infrastructure management) is also made possible by IA (Department of Computer Science, 2024). This has been suggested as a major motivation towards adoption of AI because the capability of AI to save the cost of design, licensing, installations, operations, and maintenance of energy infrastructure by hundreds of billions of dollars (European Commission, 2021).

2.4 Human-Computer Interaction in Energy Planning

Human-Computer Interaction (HCI) is a critical factor in successful use of complex energy planning tools and Artificial Intelligence. The issue of cognitive load on the users, effective visualization of multi-dimensional data, and intuitiveness of the interaction patterns should be taken into account to design user interfaces of such systems. With AI systems increasingly becoming an intrinsic part of decision-making systems, user interactions with the systems and especially through prompt-based interfaces have become critical. This capability to transform natural language queries into the usable commands to a simulation model and data analysis tools can truly increase the usability and accessibility by non-expert users. Effective HCI in energy planning would also include presentation of results. Raw simulation results usually tend to be voluminous and hard to interpret. Thus, these outputs need to be transformed into useful and best understandable insights in terms of visualizations, dashboards, and natural language summaries.

2.5 Research Gaps and Opportunities

Although there has been substantial research and development in the areas of energy system modeling, multi-agent system, and applications of AI, there are still some research gaps and opportunities that have been taken up by this thesis; Integrating LLMs into Complex Simulation Frameworks; that is, deep integration with highly specialized and computationally-intensive energy simulation models like PLEXOS, especially in the context of scenario generation and scenario interpretation, remains a nascent area. The majority of current research is centred either on language-level processing or direct control of lower-powered models. This thesis will therefore attempt to fill this gap by showing a sound model to such integration. End-to-End Transparency and Traceability in AI-driven Energy Planning; in which individual machine learning components may provide a degree of explainability, end-to-end transparency and traceability across multi- agent systems performing complex analytical workflows are currently a challenge. This plays a very critical part in high risk applications as far as regulatory compliance is concerned. The proposed research aims to develop a holistic solution to the generation of auditable trails and explanatory rationales in natural-language of each individual step within the analysis procedure. Automated Data Analysis-Flexible Code Execution: here, flexibility in code execution in response to user requests and simulation outcomes is highly useful, yet untapped, potential in energy modelling. Current models usually prescribe analytical procedures or much human effort. This thesis discusses how AI makes it possible to perform data analysis in a flexible on-the-fly manner and respond to unplanned analytical requirements. This study can help because it integrates governance agents and technical safeguards to respond to such issues, in advance. This thesis seeks to showcase a system that is able to satisfy demanding performance expectations in real life. By targeting these gaps, the study would be of benefit to developing smarter, more efficient, transparent, and compliant AI systems in critical energy infrastructure planning.

3 Research Methodology

The research methodology used in the experiment consists of the creation and testing of the multi-agent system to plan energy infrastructure. This is the scientific methodology that includes research design, methods of collecting data, the analytics procedure, and intensive system testing. The approach employed relies on the values of transparency and reproducibility, making it possible to clearly see how the questions are answered, how the conclusions are reached.

3.1 Research Design

The research design adopted for this thesis is primarily an applied research approach, focusing on the development and evaluation of a practical, AI-driven solution to a real- world problem in energy infrastructure planning.

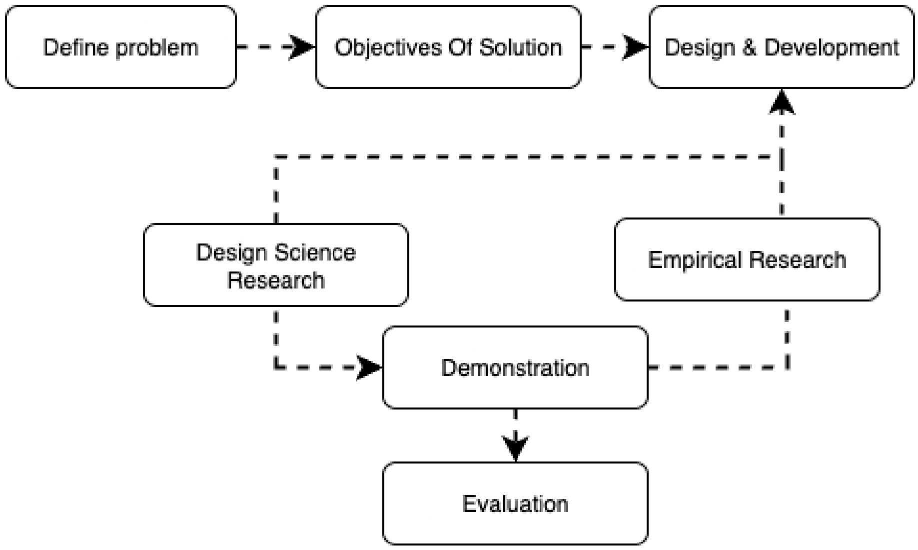


Figure 3: Holistic Design

3.2 Data Collection & Analysis

The data collection in the present research is a multi-faceted approach that involves primary data produced by the multi-agent system and secondary data used in system training, validation, and situation establishment activities. The main sources of data are the results of the PLEXOS models and they are managed and orchestrated by the SimulationAgent. The data to be extracted initially is stored in SQLite format, but is then processed under control of the DataExtractorAgent as specified in Section 3.3 of the Design Specification. This conversion guarantees that data is converted into a high-performance / columnar form that can then be subjected to subsequent analytical transformation. Secondary data sources are part and parcel of the system run and include; Natural-language prompts and baseline templates. These data are supplied by the users defining the analytical tasks and the initial conditions to be simulated. They are preprocessed and mapped into high-dimensional vectors using text-embedding- 3-large and indexed in QuadrantDB, so that they can be recalled and used across sessions (Section 4.1, Design Specification). External data sets can take the form of past energy demand, fuel price estimates, renewable energy production schedules and policy inputs. These datasets will be important to both building meaningful simulation scenarios and training the ParamsForecaster sub-agent (Section 3.7, Design Specification). their awareness and relevance of these external data sets are thoroughly checked so they can be used in the research. Data collection automation is implemented wherever possible, interrupted with manual actions to define and substantiate key entries to start with. Data collection is also provided by the system internal logging systems, i.e., logging mechanisms handled by Logger Handler and Vector-Embedding Pipeline (Section 4.5, Design Specification) that captures operational logs and system events. The logs offer good information on how the system is behaving, performing and any anomaly that it may have experienced during executions, which can serve as a rich source of diagnostic and evaluation information. The whole data collection process is instilled with maximum robustness, which guarantees data quality, uniformity.

Data analysis in this research will largely be governed by the AnalysisAgent and its sub-agents, to carry out deterministic analysis on the raw and processed data of simulation runs. Our analytical procedure is multi-layered that integrates automated tasks with strict validation and error lock procedures. The ultimate goal of data analysis is to turn unrefined simulation data into insights and Key Performance Indicators (KPIs) that can be directly used to answer research questions. Specific analysis techniques and approaches used comprise the KPI Calculation; where the PostProcessingAgent (Section 3.4, Design Specification) calculates numerous KPIs applicable to the energy infrastructure planning process. This entails adoption of Pandas 2.2 to undertake ETL processes. LLM requests and the responses are streamed to Langfuse, and a provenance graph rendering of the analysis process is generated, which guarantees traceability of the analytical process. The counter check sums are provided along with the result bundles, so as to ensure bit-exact reproduction of the produced output. This is crucial to research scientific integrity in terms of traceability and reproducibility, it is the backbone of scientific integrity by which the research results can be verified and so much more - faith in what the system produces can be bequeathed. In short, the data analysis method is highly detailed and integrates automatic processing procedures with intelligent checking and unfixed analytical impulses. It guarantees that the knowledge extracted on the basis of a multi-agent system is accurate, reliable and directly serving the research purposes.

4 Design Specification

This Design Specification describes the architectural and functional system design of the multi-agent system, which has been developed to optimise energy infrastructure planning. The major goal of this document is to be as complete and clear in the description of the system, how the components are integrated with each other and how they relate to the general principles that were used to build the system. It acts as the primary basis of implementation because it will have a guide which will ensure that the system which is developed relates to the research objectives and clearly solves the problem statement formulated. This specification follows a modular design philosophy of promoting principles of reusability, maintainability, and scalability, as well as the principles of transparency and adherence to applicable regulatory frameworks.

4.1 System Architecture

The framework over which the system is developed is Autogen, which enables building and coordinating conversational AI agents. The architecture has a number of important layers

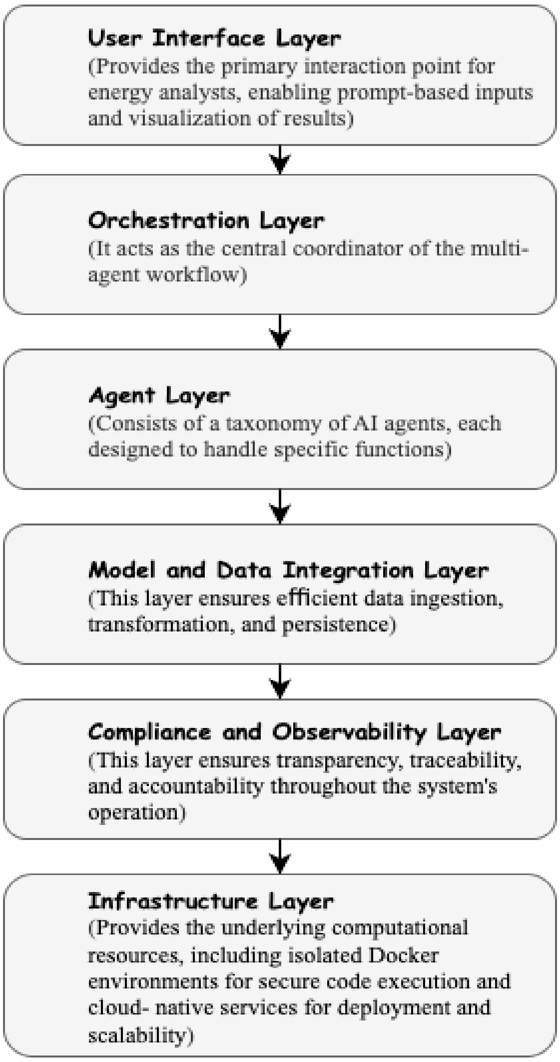


Figure 4.1: System Architecture

4.2 Agent-Specific Design

Agent	Role	Key Responsibilities	LLM Backend
OrchestratorAgent	Central coordinator	Prompt decomposition, task delegation, result synthesis	GPT-4o
SimulationAgent	PLEXOS interaction	Scenario synthesis, simulation monitoring, error handling	GPT-4 Turbo
DataAnalysisAgent	Data processing	SQLite → Parquet, KPI calc, charts/visuals	GPT-4 Turbo
PostProcessingAgent	Final stage reporting	KPIs, visualization, summaries	GPT-3.5 Turbo
InteractionAgent	User interface	Classify prompts, DAG expansion	GPT-4 / GPT-3.5
Analysis Sub-Agents	Deterministic analytics	FormulaResolver, ParamsForecaster, CodeExecutor, ErrorSolver	Mixed (GPT-4o, deepseek-Chat)

Table 4.2: Resource Requirement chosen based on the features specified in the official documentation of OpenAI and henceforth.

4.3 Data Management and Persistence

Data management is a key to the performance, reliability and traceability of the system. The design has a multi modal data storage and solid data handling strategy.

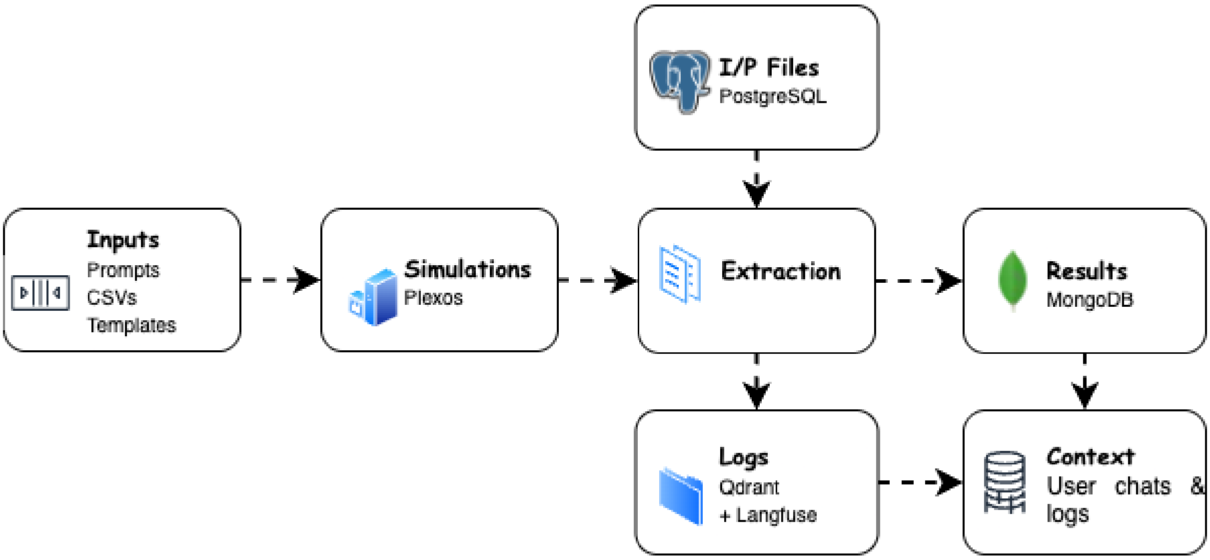


Figure 4.3: DB Architecture

4.4 Deployment and Infrastructure

The system is cloud-native and built around the AWS services to allow scalability, reliability, and security. The deployment process has integrated the continuous integration/continuous delivery (CI/CD) approaches.

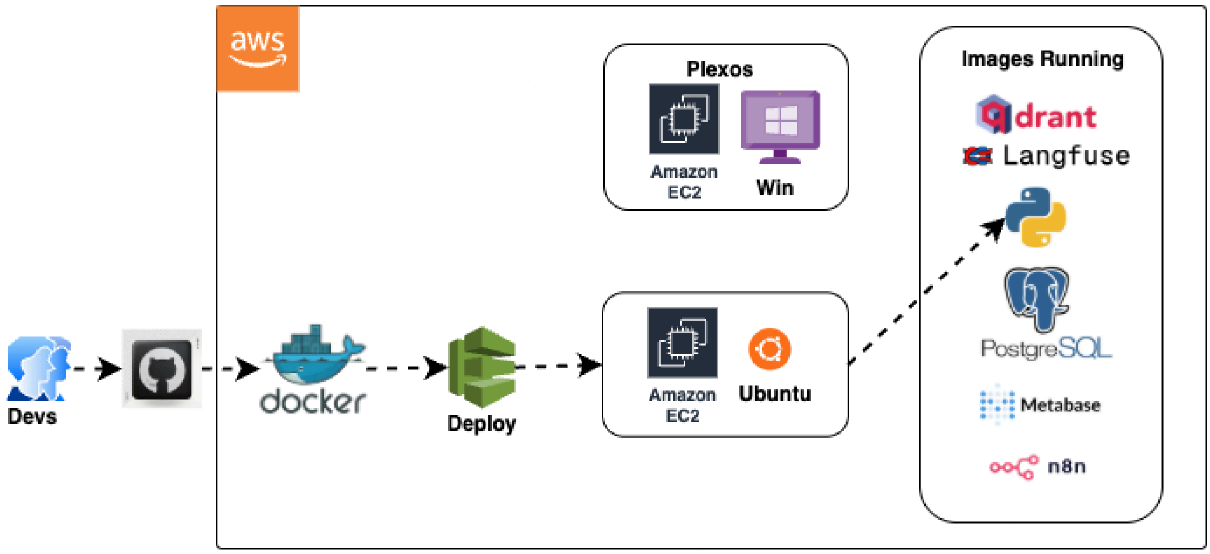


Figure 4.4: Deployment Architecture

5 Implementation

This chapter maps the system design onto an actual, multi-agent software stack based on the Autogen framework. The implementation highlights versatility, trackability and standards validity, where the parts of the implementation can work autonomously and as part of the larger workflow.

The chapter sets out the design principles outlined in Section 3.2 to a with-Autogen-v0.2 concrete multi-agent software stack. The orchestration layer divides or decomposes the responsibilities into five specialised agents A best-fit LLM back-end identified by the LLM Router (Section 3.4.5.1) is encapsulated by each agent. Table 5.1 gives a consolidated view.

Agent Class	Responsibility	Default LLM	Token Budget
OrchestratorAgent	Global planner; delegates authority; collates results	GPT-4o (128 k)	16 k/turn
Simulation Agent	Scenario synthesis; PLEXOS invocation	GPT-4 Turbo	8 k
DataAnalysis Agent	Parse out databases; normalise to Parquet	GPT-4 Turbo	4 k
PostProcessing Agent	KPI calculation; chart generation	GPT-3.5 Turbo	8 k

Table 5.1 Instantiated Autogen agents and their default model back-ends.

The system uses layer by layer prompting strategy to minimise the occurrence of hallucination and to rate convergence faster:

- Guardrails that run a combination of regex verification and JSON Schema verify tool parameters.
- Reflective feedback- The orchestrator interrogates itself after the fact and provides an explanation of each action that is presented in natural language via the user interface

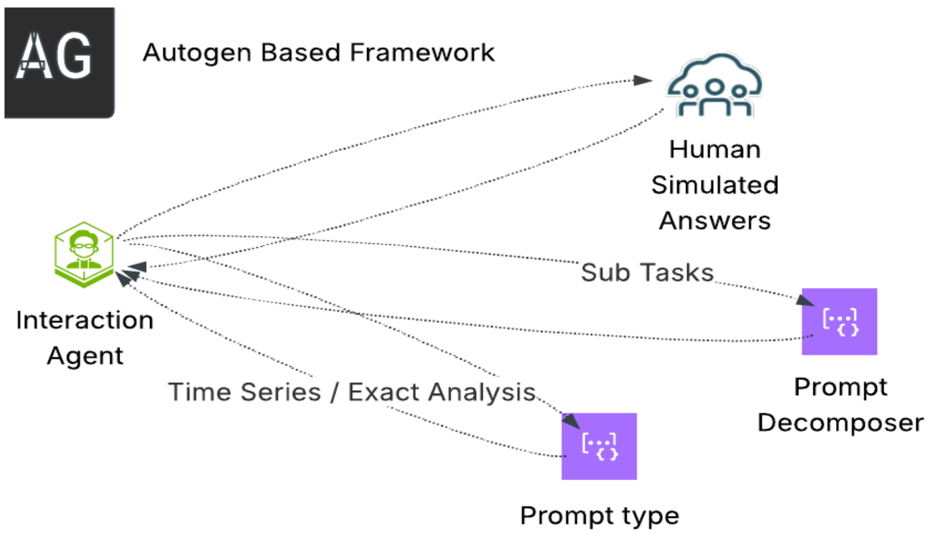
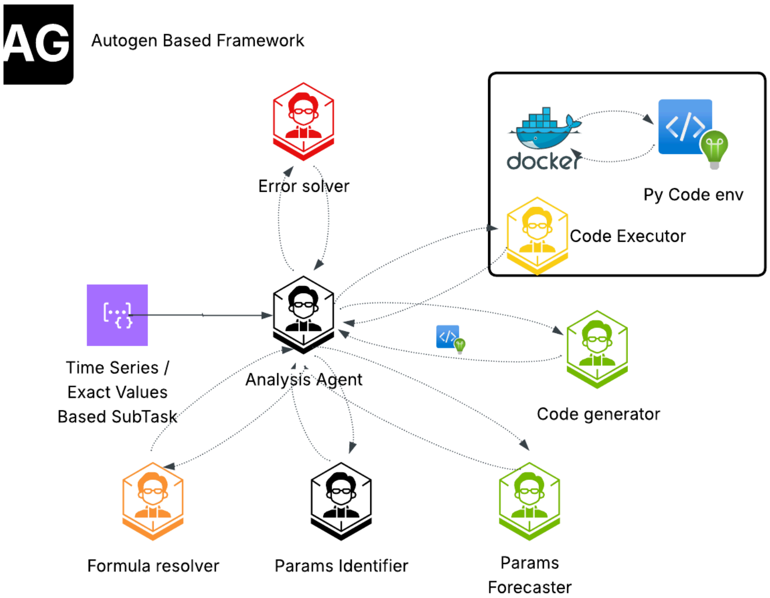
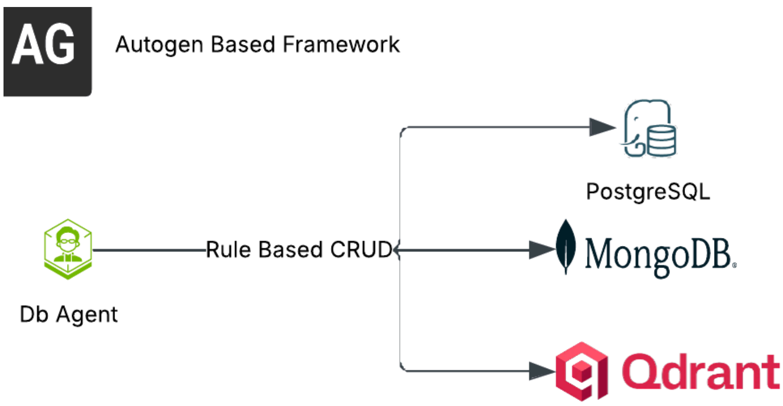


Figure 5.1: Illustrates the control flow for handling user prompts (Interaction Agent & Prompt Decomposer)

The ***AnalysisAgent*** is the heart of the deterministic analytics within the system, in which it combines a group of more specific sub-agents to analyze simulation outputs. It breaks analytical operations into components and distributes them, assigning them to specialized agents Formula Resolver, Params Identifier, and Code Generator. Dynamic code execution occurs in separate Docker sandboxes to both provide security and reproducibility. This modular solution enables the framework to be flexible to address a wide variety of analytical requirements with the robustness of error solving and validation.

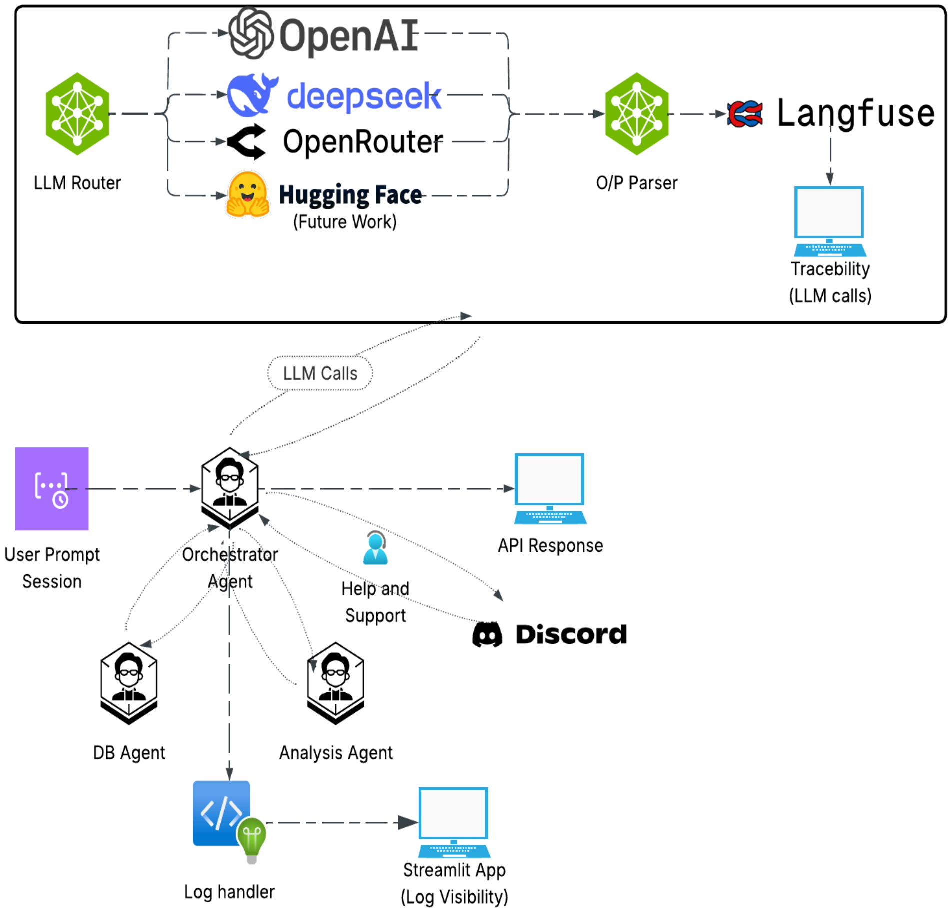


Data Management and PLEXOS Integration module is the element of the agentic framework that establishes a smooth communication cycle between user inputs on the one hand, and the simulation and the results of the analysis on the other hand. Natural-language prompts, CSV datasets, and baseline templates are initialised to high-dimensional vectors to allow semanticisation of previous scenarios. These inputs are sent through to a containerized



PLEXOS Runner micro-service that takes advantage of HPC cores, I/O batching, and gRPC stream-based simulation execution. Raw SQLite databases are used to generate simulation outputs which are extracted, transformed into optimized Parquet files, and partitioned by location and type. A built-in DbAgent manages data access on PostgreSQL (structured metadata), MongoDB (semi-structured artefacts) and Qdrant (vector embeddings) and applies security and consistency rules via Open Policy Agent. This combination guarantees that not only does the framework enable the underlying simulations to be accurate and at scale, but additionally the structured data they produce is traceable and query-ready supporting downstream analysis and visualization.

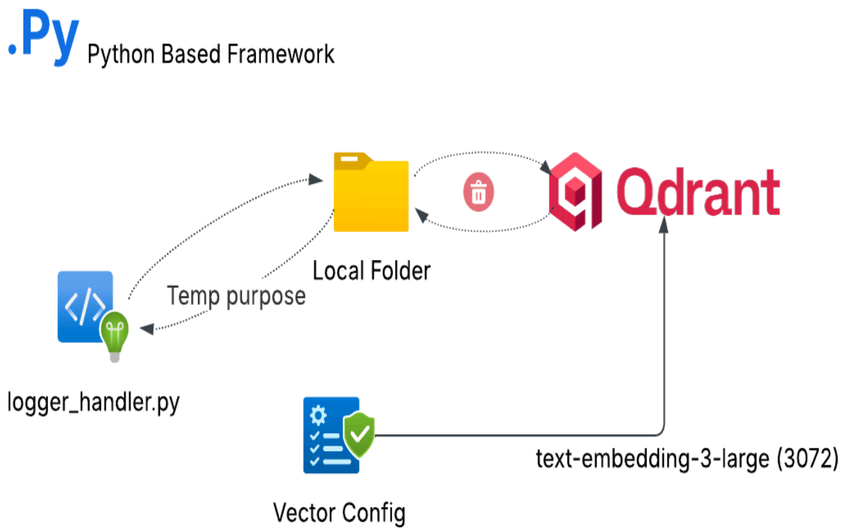
The central coordinator of the multi-agent workflow is the ***OrchestratorAgent*** that controls



the exchange of tasks among users, specialized agents and also external services. It processes instructions of high-level cognitive tasks of the user, splits them into manageable sub-tasks and directs them to components like the AnalysisAgent and DbAgent. The Orchestrator also connects to a Log Handler and Streamlit App, giving visibility of runtime events as well as providing the analyst real-time monitoring capability. One of its most important components is the LLM Router which dynamically switches between multiple backends (OpenAI, DeepSeek, OpenRouter, and potentially Hugging Face) in order to balance cost, latency, and capability. The router is coupled with an Output Parser, which checks responses before funneling the response into downstream workflows. As a measure of compliance and transparency, all LLM calls have been streamed and stored in Langfuse creating a full traceability graph that can be used as an audit trail as required by the EU AI Act. There are also external interfaces including API, Discord integration to facilitate live analyst support, through which Orchestrator has the possibility of connecting to other platforms in real-time.

The sequencing of this routing, monitoring, and traceability guarantees reliability, auditability, and user-friendliness of the system.

The Vector-Embedding Pipeline and the Logger Handler give a traceable or observable characteristic across the agentic structure. The stream of all logs is written to a buffer where they are encoded into high-dimensional vectors and indexed in Qdrant where the semantic recall and the detection of anomalies can be retrieved. A cron based sweeper with a six-hour



retention policy is in place to balance efficiency and auditability with respect to local buffers. The mechanism guarantees that all actions, system interactions, and system events can be re-constructed and analyzed which is important to make compliance, debugging, and long-term monitoring.

The Compliance, Risk and Governance layer incorporates regulatory compliance and moral protection into the core of the multi-agent framework. A particular society is classified as high-risk under the EU AI Act (Article 6 & Annex III) through a dedicated Governance Agent, and thereby automatically generates fundamental-rights impact assessments that are thereafter approved by human oversight. Technical security.. Even to build further confidence and security, technical measures are deployed throughout the pipeline: content filters intercept and block malicious or otherwise non-compliant prompts, differential-privacy security measures are used to anonymize granular dispatch data prior to its publication and all audit logs are stored immutably within AWS, assuring a verifiable and unmodifiable record of operations. Cumulatively the mechanisms guarantee that the framework is transparent, responsible, and in line with the regulator emerging requirements of critical infrastructure AI systems.

6 Evaluation

This assessment of the proposed multi-agent framework will aim to show that it is technically sound, as well as of practical use of the energy infrastructure planning. As there is currently, to the best of our knowledge, no directly comparable system available, the evaluation consists of (i) case studies consisting of realistic planning tasks, (ii) expert validation, consisting of managerial reviews, as well as (iii) comparative positioning, both in comparison to existing and available workflows such as manual PLEXOS pipelines and the newly introduced PLEXOS Intelligence module. Such a combination of methods makes the evaluation comprehensive, paying much attention both to the quantitative outcomes of the evaluation and to the qualitative issues of using the developed software,including its usability and interpretability as well as decision-making support.

6.1 Case Studies & Experiments Performed

The performance of the system was evaluated in three case studies that resulted under various circumstances: generation of scenarios, comparative analysis and explanatory reasoning.

6.1.1 Automated Scenario Generation (Hydrogen & Renewables Pathways)

The system in this case study was charged with the task of automatically creating transition scenarios of the energy system in the Dutch with high penetration of renewables and differing levels of the adoption of hydrogen. The SimulationAgent created the design-of-experiments matrix and created several runs of the PLEXOS Runner micro-service. The outcomes contained renewable shares, marginal cost lines and hydrogen use pathway. In comparison to the manual configuration, the automated pipeline resulted in scenario outputs taking less time, saving considerable amount of time and resulting in circumstantial reproducibility.

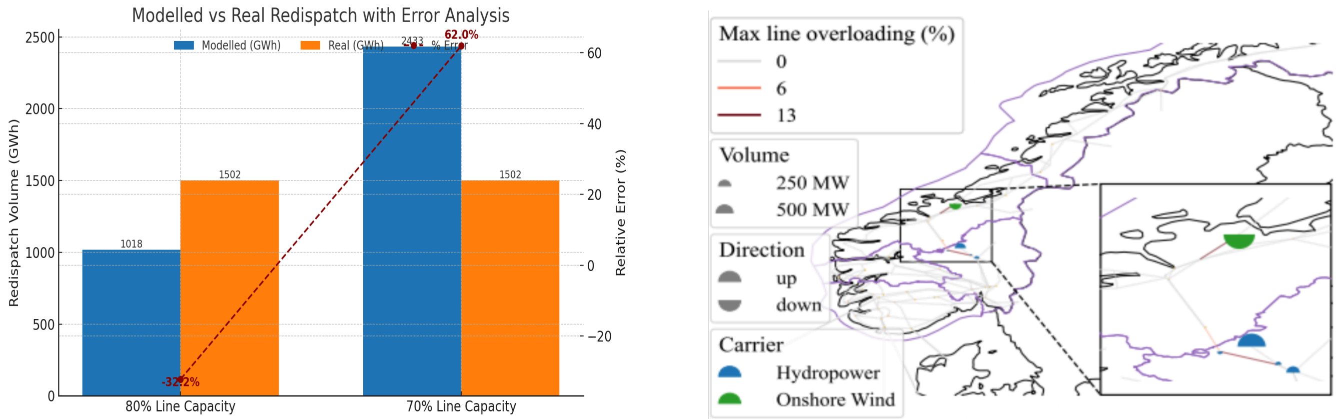


Figure 6.1.1: Redispatch analysis comparing modelled vs. real volumes with error evaluation (left) and a spatial snapshot of redispatch flows, line overloading, and carrier contributions (right).

6.1.2 Explanatory Reasoning

The third case study assessed the explanation of the system with the question: "***What is the load factor of a generator?***"

Rather than running an entire simulation, the InteractionAgent marked the query as explanatory and used Qdrant semantic recall. The explanation is that it reduces offshore wind due to offshore wind overcapacity, a lack of cross-border interconnections and storage capacity.
A domain expert was involved in the review of the response, and its plausibility and accuracy were confirmed.

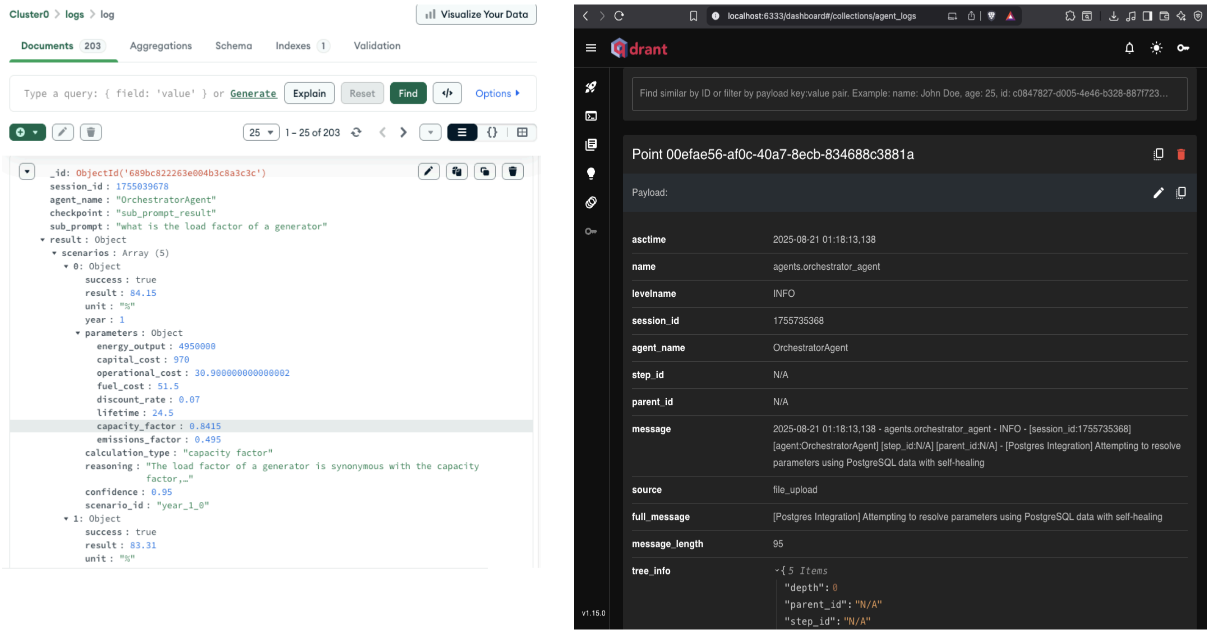


Figure 6.1.2: Example of system-generated explanatory reasoning captured in logs (QdrantDB) and eventually the final results (MongoDB).

6.2 Expert Validation

The results of the case studies were read out to a high energy systems manager as a check. The professional vouched:

- Quantitative outputs were within the level of internal reference scenarios.
- Rational thinking was in agreement with the conventional knowledge on planning.
- The manual loads were minimized, and the reproduction increased.

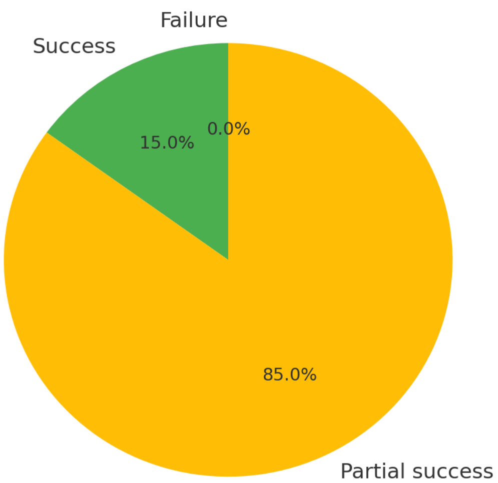


Figure 6.2.1: Outcome distribution for 20 agent-orchestrated tasks (success, partial success, failure). Partial successes reflect minor presentation or rounding issues rather than substantive computation failures.

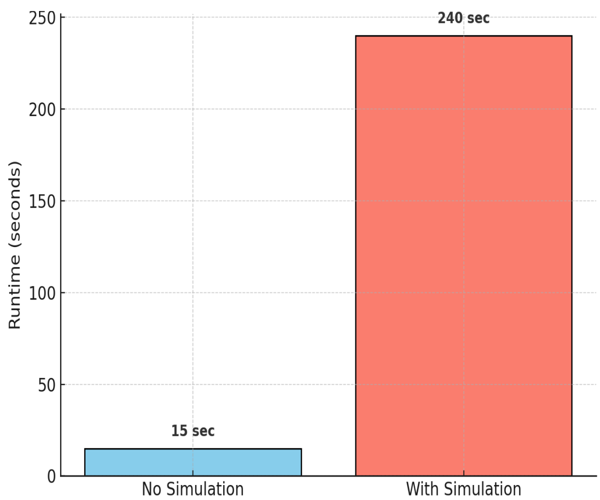


Figure 6.2.2: Average runtime of subtasks with and without Plexos simulation. The orchestrator-only workflow completes in ~15 seconds, whereas including Plexos simulation extends the runtime to ~240 seconds on average, highlighting the cost of detailed modelling.

The validation that involves such form is a golden-standard dataset, where formal benchmarks are absent, and gives a confidence in the realized accuracy of the system outputs, both technically and practically.

6.3 Comparative Positioning

Other evaluation steps were to compare the proposed framework with the current workflows.

- Manual workflows in PLEXOS: Scenarios are usually created manually, raw outputs are exported and post-processing done in Excel or R. This method is cumbersome, inaccurate and hard to repeat. These actions are automated under the proposed system, which not only decreases the number of work hours, but also increases its steadiness.
- Energy Exemplar has recently released PLEXOS Intelligence, which is an AI-enhanced feature to inquire questions in natural language on the platform. Nonetheless, this tool is reactive and only can answer questions about current outputs. In comparison, the proposed system provides proactive orchestration: generation of scenarios, simulation execution and comparison analysis.

Criterion	Manual PLEXOS Workflow	PLEXOS Intelligence (Energy Exemplar)	Proposed Multi-Agent Framework
Scenario Creation	Manual configuration of inputs; time-consuming	Limited – scenarios must already exist	Automated scenario synthesis via SimulationAgent
Simulation Execution	Manual run setup; no orchestration	Runs within GUI; user-triggered	Automated execution via containerised micro-service (PLEXOS Runner)
Data Extraction	Raw SQLite → manual ETL (Excel/R, custom scripts)	Provides summary metrics inside platform	Automated extraction & transformation to Parquet + KPIs
Comparative Analytics	Performed manually in external tools	Not natively supported	Built-in comparative analysis (multi-scenario KPIs & charts)
Explanatory Reasoning	Requires domain expert interpretation	Provides Q&A over existing outputs	Natural-language explanations via InteractionAgent & LLMs
Traceability	Minimal; manual notes/logs	Basic query logs	Full Langfuse traceability, audit trails, reproducibility
Usability	Requires advanced PLEXOS expertise	More user-friendly (chat-like interface)	Intuitive prompt-based UI + visual outputs
Reproducibility	Low – dependent on manual steps	Moderate – query reproducible but workflows not	High – automated pipelines, checksums, embedded logs
Role of AI	None	Reactive assistant for queries	Proactive orchestrator driving full workflow (scenario → sim → analysis → explanation)

Table 6.3: Comparative positioning (Manual Workflow vs PLEXOS Intelligence vs Proposed Framework)

6.4 Discussion

The review shows that the designed system can:

- Saving large amounts of time in the automation of the scene production and analysis.
- Interpretation and reporting of comparative analytics with reproducibility.
- Presenting explanations in terms of natural language that give a significant increase in interpretability.
- This is in the sense that it functions as an orchestration layer to augment rather than substitute PLEXOS.

Limitations also need to be introduced. The case studies are illustrative and not large scale benchmarking exercises. Besides, there are no similar systems, and the extent of quantitative benchmarking is limited. Lastly, the reasoning process based on the LLMs is likely to be irregular, and in cases where decisions must be made with serious consequences, such reasoning requires human control.
Future work could use longer datasets, studies of more than one analyst, and formal benchmark development. The results proposed, however, present a solid case evidencing that the system elevates the level of practice in energy infrastructure planning.

7 Conclusion and Future Work

The current thesis included a description of a multi-agent framework that is aimed at planning energy infrastructure through the use of Large Language Models (LLMs) and simulation software (e.g. PLEXOS). Unlike traditional manual-based workflows and vendor-specific assistants (e.g., PLEXOS Intelligence), the proposed system provides a complete orchestration pipeline that automates the process of constructing the scenario, running simulations, extracting the data, and performing analytics, yet is transparent and explainable. The framework minimizes the human input with its layered architecture that includes the use of specialized orchestration, simulation, analysis, and compliance agents. In this way, it makes the research highly reproducible and provides proactive decision support instead of reactive support. Next steps include more generalizable applicability to use alternative simulators (e.g., PyPSA, TIMES), with real-time data feeds to support operational needs, validation using domain experts as a gold standard and testing scale of application in enterprise systems. Beyond the block, other improvements, like fine-tuning domain specific LLMs, reinforcements in explainability with causal and visual computation, embedding regulatory modules in line with the EU AI Act could push the system closer toward an industry-grade decision-support platform adding a missing link between advanced computational models and the real life of planners and policymakers.

References

Sosef, M.N. (2017) *Modeling local energy systems in PLEXOS – coupling electricity and heat demand in search for improved flexibility*. Utrecht University. Available at: *https://studenttheses.uu.nl/handle/20.500.12932/27762* .

ProQuest (2023) *Assessing the impact of increased parameter modeling of combustion turbines in a grid with varying renewable energy penetration using PLEXOS*. Available at: [*https://search.proquest.com/openview/18d8fabd6fa8ae0867b406c15174e1e7/1?pq-origsite=gscholar\&cbl=18750\&diss=y*](*https://search.proquest.com/openview/18d8fabd6fa8ae0867b406c15174e1e7/1?pq-origsite=gscholar&cbl=18750&diss=y*) .

Hamdi, M., El Salmawy, H.A. and Ragab, R. (2024) 'Incorporating operational constraints into long-term energy planning: The case of the Egyptian power system under high share of renewables', *Energy*, 300, 131619. *doi:10.1016/j.energy.2024.131619*.

Energy Exemplar (n.d.) *PLEXOS | Energy market simulation software*. Available at: *https://www.energyexemplar.com/plexos* .

Widl, E., Cronbach, D., Sorknæs, P., Fitó, J., Muschick, D., Repetto, M., Ramousse, J. and Ianakiev, A. (2022) 'Expert survey and classification of tools for modeling and simulating hybrid energy networks', *Sustainable Energy, Grids and Networks*, 32, 100913. *doi:10.1016/j.segan.2022.100913*.

SmythOS (2024) *Multi-agent systems in energy management*. Available at: *https://smythos.com/developers/agent-development/multi-agent-systems-in-energy-management* .

Feng, J., Yu, T., Zhang, K. and Cheng, L. (2025) 'Integration of multi-agent systems and artificial intelligence in self-healing subway power supply systems: Advancements in fault diagnosis, isolation, and recovery', *Processes*, 13(4), 1144. *doi:10.3390/pr13041144*.

Fawaz, A., Mougharbel, I., Al-Haddad, K. and Kanaan, H.Y. (2025) 'Energy routing protocols for energy internet: A review on multi-agent systems, metaheuristics, and artificial intelligence approaches', *IEEE Access*, pp. 1–1. *doi:10.1109/access.2025.3546620*.

Pinto, D., Gomes, L., Faria, P., Sousa, F., Vale, Z. and Pinto, T. (2020) *Demonstration*, *IFAAMAS Proceedings*, 3, pp. 2114–2115. Available at: *https://www.ifaamas.org/Proceedings/aamas2020/pdfs/p2114.pdf* .

U.S. Department of Energy (2024) *Artificial intelligence for energy*. Available at: *https://www.energy.gov/topics/artificial-intelligence-energy* .

Huang, X., Huang, L., Zhou, Y., Li, H., Yu, Q., Yang, F. and Sui, J. (2024) 'Application of large model technology based on prompt generation and data insights in power system data analysis', *2024 4th International Conference on Intelligent Power and Systems (ICIPS)*, pp. 736–743. *doi:10.1109/icips64173.2024.10900178*.

Husom, E.J., Goknil, A., Shar, L.K. and Sen, S. (2024) *The price of prompting: Profiling energy use in large language models inference*. arXiv preprint. Available at: *https://arxiv.org/abs/2407.16893* .

You, J. (2025) *How much energy does ChatGPT use?* Epoch AI. Available at: *https://epoch.ai/gradient-updates/how-much-energy-does-chatgpt-use* .

Pressley, J.P. (2024) *AI is revolutionizing grid planning in the energy and utilities sector*. BizTech Magazine. Available at: *https://biztechmagazine.com/article/2024/10/ai-revolutionizing-grid-planning-energy-and-utilities-sector* .

International Energy Agency (IEA) (2025) *AI and energy security – Energy and AI*. Available at:

https://www.iea.org/reports/energy-and-ai/ai-and-energy-security .

University of Chicago, Department of Computer Science (2024) *How artificial intelligence can transform U.S. energy infrastructure*. Available at: *https://cs.uchicago.edu/news/how-artificial-intelligence-can-transform-u-s-energy-infrastructure/* .

European Commission (2021) *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*. Available at: *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206* .

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004) ‘Design science research methodology for information systems research’, *Journal of the Association for Information Systems*, 5(7), pp. 75–105. Available at: *https://aisel.aisnet.org/jais/vol5/iss7/5/* .