Master Thesis

An Agentic AI Framework for trend monitoring and real-time scenario generation for LCA models on CRM value chains

Student name: Zhi-Chin Ju
Student number Leiden: 4195280
Student number Delft: 6348440

First thesis supervisor:
Carlos Felipe Blanco Rocha, Institute of Environmental Sciences, Leiden University
Second thesis supervisor:
Benjamin Sprecher, Industrial Design Engineering, TU Delft

Date: 22-06-2026
Program name:
MSc Industrial Ecology, Leiden University, Delft University of Technology, the Netherlands

Abbreviation

AI	Artificial Intelligence
API	Application Programming Interface
APOC	Awesome Procedures on Cypher
CBAM	Carbon Border Adjustment Mechanism
CBR	Case-Based Reasoning
CF	Characterization Factor
CoT	Chain-of-Thought
CRAG	Corrective Retrieval-Augmented Generation
CRM	Critical Raw Material
CRMA	Critical Raw Materials Act
DLCA	Dynamic Life Cycle Assessment
EKG	Event Knowledge Graph
ERP	Enterprise Resource Planning
ESPR	Ecodesign for Sustainable Products Regulation
ESA	Event Sequence Analysis
EV	Electric Vehicle
FIT	Feed-in Tariff
GND	Global Network Density
GNN	Graph Neural Network
GraphSAGE	Graph Sample and AggregatE
GWP100	Global Warming Potential over 100 years
HITL	Human-in-the-loop
IEKG	Industrial Event Knowledge Graph

IoT	Internet of Things
IPCC	Intergovernmental Panel on Climate Change
JSON	JavaScript Object Notation
KG	Knowledge Graph
LCA	Life Cycle Assessment
LCIA	Life Cycle Impact Assessment
LCI	Life Cycle Inventory
LLM	Large Language Model
LP	Link Prediction
MAS	Multi-Agent System
ML	Machine Learning
Pa-LCA	Parametric Life Cycle Assessment
PV	Photovoltaic
PyG	PyTorch Geometric
RAG	Retrieval-Augmented Generation
ReAct	Reasoning and Acting
REE	Rare Earth Elements
RL	Reinforcement Learning
SRQ	Sub-research Question
UNCTAD	United Nations Conference on Trade and Development

Table of Contents

Executive summary

Conventional Life Cycle Assessment (LCA) is limited by its reliance on static and retrospective background datasets and by its predominantly inside-out modelling approach, where models are built strictly from internal factory operations and primary data, often missing sudden external disruptions**.** These limitations reduce their capability in context of the Critical Raw Material (CRM) value chains affected by geopolitical conflicts and climate-related disruptions. While external macro-events may alter Life Cycle Inventory (LCI) parameters, they are usually reported in fragmented and unstructured sources like daily news. Additionally, public news rarely reports direct LCA parameters or exact numerical changes, making it difficult to capture external events that genuinely impact LCA models. More recently, Large Language Models (LLMs) have emerged as a promising solution. However, without historical references and the capacity for complex reasoning or real-world interaction, LLMs are highly prone to fabricating or hallucinating missing numbers when asked to generate scenario values. This makes these events difficult to translate into structured, quantitative, and LCA-ready scenario inputs.

This study conducts a methodological exploration and develops an Event Knowledge Graph (EKG)-driven Agentic AI framework as a human-governed outside-in sensing layer for LCA scenario exploration. Because relying solely on LLMs to quantify qualitative news inevitably leads to numerical hallucinations, the objective is to examine how qualitative real-world macro-events can be reliably translated into structured, quantitative, and LCA-ready parameter-change scenarios without unconstrained AI guessing.

In the framework, EKG acts as a historical reference library, recording who, where, when, what, why, and how selected LCI parameters have changed during past disruptions. To ensure the Agentic AI focuses only on what is both LCA-relevant and observable in the news, these parameters and their trigger mechanisms are defined in advance by experts using Contribution Analysis or Global Sensitivity Analysis. Complementing this, the Agentic AI functions as a team of specialized agents that continuously monitors real-time news, matches emerging events against the EKG's historical records, and generates grounded, bounded parameter adjustments strictly within the expert-defined numerical ranges.

To improve reliability and prevent AI hallucination, the framework incorporates a Human-In-The-Loop (HITL) approach, Retrieval-Augmented Generation (RAG), and rigorous validation steps. First, human intervention establishes the operational boundaries, including the selection of LCI parameters and the setting of quantitative limits for different severity scenarios. Second, the

Agentic AI is designed to reason exclusively within a focused context via the RAG architecture, preventing it from guessing or fabricating data. For instance, LLM-based graders retrieve and filter relevant documents to build an EKG centered on historical parameter shifts, while the scenario generator only draws conclusions if current news events align with these documented historical precedents. Lastly, structured validations evaluate the performance of the RAG architecture, testing whether the grader consistently classifies document relevance and ensuring that generated scenarios strictly follow the defined methodological rules and logical processes.

The framework is demonstrated using the graphite value chain, focusing on China's electricity carbon intensity and graphite supply-chain transport routing as the key LCI parameters. Before searching related news, the LLM-based graders were validated, achieving an F1-score of 72.64% and supporting their use for filtering LCI-relevant documents. Then, the Agentic AI framework built a historical EKG from 2020 to 2025 and monitored news in 2026 to detect similar events, estimating parameter changes across best-, average-, and worst-case scenarios. After that, 78.8% of the generated scenarios successfully passed the plausibility validation, proving they were constrained within the predefined scenario boundary matrix and logical reasoning process. Finally, the system translated qualitative disruptions into bounded parameter changes. For instance, news regarding Houthi attacks on maritime routes was translated into discrete increases of 30.0%, 37.5%, and 45.0% in graphite transport distance across the best-, average-, and worst-case scenarios. Similarly, reduced renewable output and increased thermal generation in China's Q1 electricity mix were translated into discrete increases of 3.0%, 4.5%, and 6.0% in electricity carbon intensity. In the simplified impact calculation, the rerouting scenario produced discrete GWP100 increases of 0.070, 0.087, and 0.105 kg CO2e/kg for natural graphite, and 0.098, 0.122, and 0.146 kg CO2e/kg for synthetic graphite. For the electricity scenario, the respective best- and worst-case GWP100 increases were 0.139 and 0.278 kg CO2e/kg for natural graphite, and 0.575 and 1.149 kg CO2e/kg for synthetic graphite.

The main contribution of this research is a novel, validated, and reusable blueprint for the structured, low-hallucination risk translation between macro-disruptions in CRM value chains reported in real-time news and quantitative changes in LCI parameters. Crucially, the system explicitly positions Agentic AI as a supportive co-pilot, proving that human expertise remains indispensable in defining parameter scopes, setting scenario boundaries, and performing the final LCA interpretation.

1. Introduction

The global transition toward sustainability heavily relies on technologies that require critical raw materials (CRMs) (Bobba et al., 2020; Lewicka et al., 2021; Pommeret et al., 2022), which are essential for EV batteries (lithium, cobalt, nickel), electric motor magnets (neodymium, dysprosium), and advanced semiconductors (Rare Earth Elements) (Granvik et al., 2025). To quantify the environmental impacts of these technologies and ensure regulatory compliance, Life Cycle Assessment (LCA) is the gold-standard assessment framework. However, the static and retrospective nature of conventional LCA limits its ability to represent rapidly changing and uncertain CRM value chains (Zong et al., 2025), highlighting a critical need to shift LCA from a static accounting tool to a proactive scenario exploration framework.

1.1 Background

LCA has been recognized as an effective tool for assessing environmental impacts, resource criticality (Cimprich et al., 2019; Mancini et al., 2015) and supply-chain implications (Elias Mota et al., 2020; Hülagü et al., 2025). Beyond holistic environmental evaluation, LCA identifies the supply chain hotspot (Masudin et al., 2025) and compares different alternatives, which can guide supply chain vulnerability mitigation, such as recycling and substitution (Mertens et al., 2024), as well as circular design (van Gaalen & Chris Slootweg, 2025).

In addition, LCA is no longer only a voluntary sustainability reporting tool. It is increasingly embedded in mandatory environmental regulations, including the EU Carbon Border Adjustment Mechanism (CBAM) (Regulation (EU) 2023/956, 2023) , the Battery Regulation (Regulation (EU) 2023/1542, 2023), and the Ecodesign for Sustainable Products Regulation (ESPR) (Regulation (EU) 2024/1781, 2024). As a result, LCA now plays an important role in ensuring supply chain transparency, traceability, and digital integration.

Despite its significant utility, the high volatility and uncertainty driven by geopolitical conflicts and extreme weather (Qin et al., 2023) pose substantial challenges to the conventional LCA. The vulnerability of the CRM supply chain stems from a heavy geographic concentration in specific developing and mineral-rich developed nations (Britt & Czarnota, 2024; Ericsson & Löf, 2020), compounded by intersectoral competition (Bobba et al., 2020). Furthermore, Bednarski et al. (2025) emphasize that geopolitical risks are now the major threat; the “weaponization of interdependence” not only intensifies these existing supply chain vulnerabilities but also makes future disruptions highly unpredictable.

The absence of temporal and dynamic factors in LCA not only undermines the credibility of LCA results but also impairs sustainability decision-making (Zong et al., 2025). To overcome the challenges, recent studies have developed Parametric LCA (Pa-LCA) (Arias et al., 2025) and

Dynamic LCA (DLCA) to improve its flexibility. For instance, dynamic life cycle inventory can be implemented by connecting with Industry 4.0 technology to get real-time primary data in the industry system (Cornago et al., 2022). Nevertheless, these solutions share a major limitation with traditional LCA: they mostly operate through an "inside-out" approach. In reality, the most critical disruptions often occur *outside* the factory gates, driven by macro-environmental events.

Background data, which capture broader network conditions and macro-environmental changes, can strongly influence LCA results (Mendoza Beltran et al., 2020). Even a simple update to a newer version of an LCA database can drastically shift impact outcomes, sometimes flipping which technology appears more environmentally friendly (Xicotencatl et al., 2023) (Lubecki et al., 2024). Relying on a single, fixed background assumption is therefore insufficient for CRM value chains. However, unlike structured foreground data, external background information or supply-chain information are typically fragmented, unstructured, and heterogeneous, such as daily news and industry reports (Deng et al., 2023; Ponnock et al., 2025), This disorder hinders manual Life Cycle Inventory (LCI) mapping, making continuous human data extraction highly inefficient and unrealistic (Pigné et al., 2020).

1.2 Problem and Research Gap

LCA struggles to keep pace with real-time volatility in CRM value chains. As a result, there is a pressing need for an "outside-in" AI-driven external sensor that can continuously monitor real-world events and inform LCA models, enabling them to explore various disruption scenarios proactively.

However, achieving this requires overcoming the "Text-to-Insight Gap". Currently, there is a methodological lack of an external information-processing layer capable of translating chaotic, qualitative news into structured, grounded, and LCA-compatible quantitative inputs. While Large Language Models (LLMs) offer strong natural language processing capabilities, relying on a single LLM presents two major challenges. First, without historical context, a single LLM is prone to hallucinate numerical values, leading to outputs that cannot be trusted or justified. Second, a single model lacks the ability to autonomously handle complex goals or interact with external data environments. Extracting LCA parameters from unstructured text requires rigorous reasoning constraints that a single LLM simply cannot reliably execute.

As a result, there is methodological research gap: there is a lack of a reliable AI external information-processing layer that can monitor changes in real world from qualitative text and convert them into structured and grounded number and LCA-compatible inputs.

1.3 Research Questions and Objectives

To bridge this gap, this study introduces an innovative framework integrating two core technologies: Event Knowledge Graph (EKG) and Agentic AI.

EKG is a method of storing and representing real-world entities, events, and the relations between them in a graphic structure with nodes and edges (Li et al., 2025). It serves as the historical reference library of how selected LCI parameters by experts changed in the past. Rather than just recording numerical values, the EKG captures the "5W1H" context (Who, What, When, Where, Why, and How), aligning with the news extraction framework proposed by Keith et al. (2020). For example, the EKG connects the macro-trend (geopolitical conflict) and the micro-event (disruption in the Red Sea) to the actors (shipping companies) and the resulting parameter observation (a 30% increase in distance) for the parameter (shipping distance) in specific CRM value chains. Additionally, it changes how we monitor events. Instead of executing superficial searches for parameter value changes, the EKG enables the proactive and periodic monitoring of critical entities or events, such as key actors, strategic locations, or emerging macro-trends.

Meanwhile, Agentic AI moves beyond the traditional approach of relying on a single, all-purpose chatbot. It coordinates a team of specialized, task-focused agents, powered by LLMs as their "brain" and external Application Programming Interface (API) as their "tool" (Bandi et al., 2025). To ensure that agents work together effectively and efficiently, the framework uses schemas, the standardized data templates, to generate structured output and shared from one agent to the next. This helps break down complex tasks, such as searching, filtering, and scenario generation, into small, manageable, and traceable steps. By limiting each agent to a specific task and following human-defined boundaries, the system prevents the AI from overloading or hallucinating but also ensures that the immediate and final results are all grounded and easy for experts to review.

Overall, this Agentic AI framework aims to equip traditional LCA with its own knowledge base and foresight across CRM supply chain. The final objective of the Agentic AI framework is to monitor real-time qualitative news related to key trends that might trigger LCI parameters sensitivity and further generate potential change percentage in best-average-worst scenarios. This can then provide quantitative data for LCA mapping and explore the change in environmental impact. This outside-in layer is not intended to replace LCA modelling. Instead, it supports the dynamic scenario development in LCA.

Therefore, the main research question in this study is:

To what extent can the EKG-driven Agentic AI framework bridge the gap between qualitative news and quantitative LCI parameter changes to support trend monitoring and grounded real-time scenario exploration?

To comprehensively address the main research question, the following sub-research questions (SQs) are defined:

SRQ1: How can unstructured, qualitative public news be systematically extracted and transformed into a structured EKG to establish a grounded historical evidence base?

SRQ2: How can the historical EKG be integrated with LLMs to facilitate contextual reasoning, enabling the translation of monitored events into quantitative variations of LCA parameters?

SRQ3: How can the relevance of the news inputs and the plausibility of the generated outputs be systematically validated?

The study intends to achieve three contributions. First, it develops an Agentic AI pipeline for constructing an evidence-grounded EKG from unstructured CRM supply-chain documents. Second, it demonstrates how monitored real-time events can be translated into bounded best-, average-, and worst-case LCI parameter changes using mechanism mapping and historical EKG references. Third, it proposes a scalable outside-in sensing framework that connects external trend monitors with parameterized and dynamic LCA applications.

This paper is structured as follows: Chapter 2 reviews the related literature; Chapter 3 details the proposed methodology; Chapter 4 shows the results with graphite value chain as case study and end-to-end system validation; Chapter 5 discusses the limitation and future direction.

2. Literature review

This chapter reviews three bodies of literature that form the conceptual basis of this study. First, it discusses the limitations of conventional LCA and the emergence of parametric and dynamic LCA approaches. Second, it reviews Knowledge Graphs (KG) and EKG as methods for structuring heterogeneous event information. Third, it examines Agentic AI as an orchestration framework for automating multi-step information processing workflows while maintaining traceability and control.

2.1 The Limitations of Conventional LCA

The conventional LCA has faced criticism across multiple methodological dimensions. Primarily, it relies on static and retrospective LCIs (Ibn-Mohammed et al., 2023; Salla et al., 2025). The temporal and spatial parameters within these LCIs are often treated as fixed values, which inherently restrict their applicability in complex decision-making (Romeiko et al., 2024). Furthermore, conventional LCA focuses exclusively on physical flow processes, failing to incorporate socio-economic dimensions (Cimprich et al., 2019). Additionally, the data collection process remains highly time-consuming and often suffers from opaque sourcing (Saavedra-Rubio et al., 2022). In summary, conventional LCA operates on an inside-out, bottom-up approach: it utilizes static databases with fixed values to trace physical flows across a life cycle and derives environmental impacts through a LCIA characterized by static, long-term characterization factors (CFs) lacking spatial and temporal resolution (Su et al., 2022).

To address these limitations, various methodological extensions of LCA are emerging to enhance adaptivity and decision-support capabilities, including Parametric LCA (Pa-LCA) and Dynamic LCA (DLCA). Pa-LCA is a dynamic modeling and analysis method that integrates predefined variable parameters to enable the evaluation of LCI parameter sensitivity (Arias et al., 2025) and even facilitates design optimization (Reisinger et al., 2022). This approach significantly enhances assessment flexibility and provides a basis for connecting LCA with artificial intelligence (AI) through mathematical modeling and machine learning (Ibn-Mohammed et al., 2023).

DLCA is an approach that explicitly accounts for changes in systems and flows over time (Beloin-Saint-Pierre et al., 2020). Current research predominantly features partially dynamic LCAs, focusing mainly on dynamic system boundaries, dynamic system inventories, and dynamic characterization (da Costa et al., 2024; Sohn et al., 2020). Within this scope, dynamic system inventories have received the most attention, particularly through integration with sensors, the Internet of Things (IoT), and digital twins to acquire real-time primary data (Ferrari et al., 2021; Petri et al., 2025). Several studies further show the LCIA differences after including dynamic factors in LCA. For instance, Rovelli et al. (2022) showed that using monthly inventory data in steelmaking led to a ±14% variation compared with static annual averages, mainly due to changes in raw material inputs. Similarly, Hosamo et al. (2024) found that building-material impacts vary across timeframes and supply-chain scenarios. These findings indicate that DLCA requires strong data-processing and information-integration capabilities.

Taken together, these studies show that coupling LCA with parameterization and dynamic modelling can better address the complexity and volatility of real-world systems. However, these approaches still lack standardized methodological frameworks (Arias et al., 2025; Sohn et al., 2020). Researchers also face two major challenges: the lack of standardized reporting practices and the difficulty of collecting time-dependent information (Beloin-Saint-Pierre et al., 2020). Simultaneously, as the number of dynamic variables and data increases, it introduces further complexity and uncertainty into the system (Fnais et al., 2022). Consequently, the incorporation of big data requires substantial additional effort for data collection and management (Su et al., 2022). Finally, these approaches improve the flexibility of LCA models, but they do not answer how to continuously identify relevant external events and convert them into structured, quantitative, and traceable LCI parameter changes.

2.2 Event Knowledge Graph

A KG is a knowledge representation framework centered on "entities" as core nodes, connected by "relations" that represent static relationships between them (Guan et al., 2022). The fundamental unit of a KG is the structural triple: subject, predicate, and object (Ji et al., 2021). By providing a machine-readable context, KGs enable computers to process information efficiently and unambiguously (Peng et al., 2023). Consequently, KGs can effectively accumulate and transfer real-world human knowledge into AI systems (Ji et al., 2021; Peng et al., 2023)

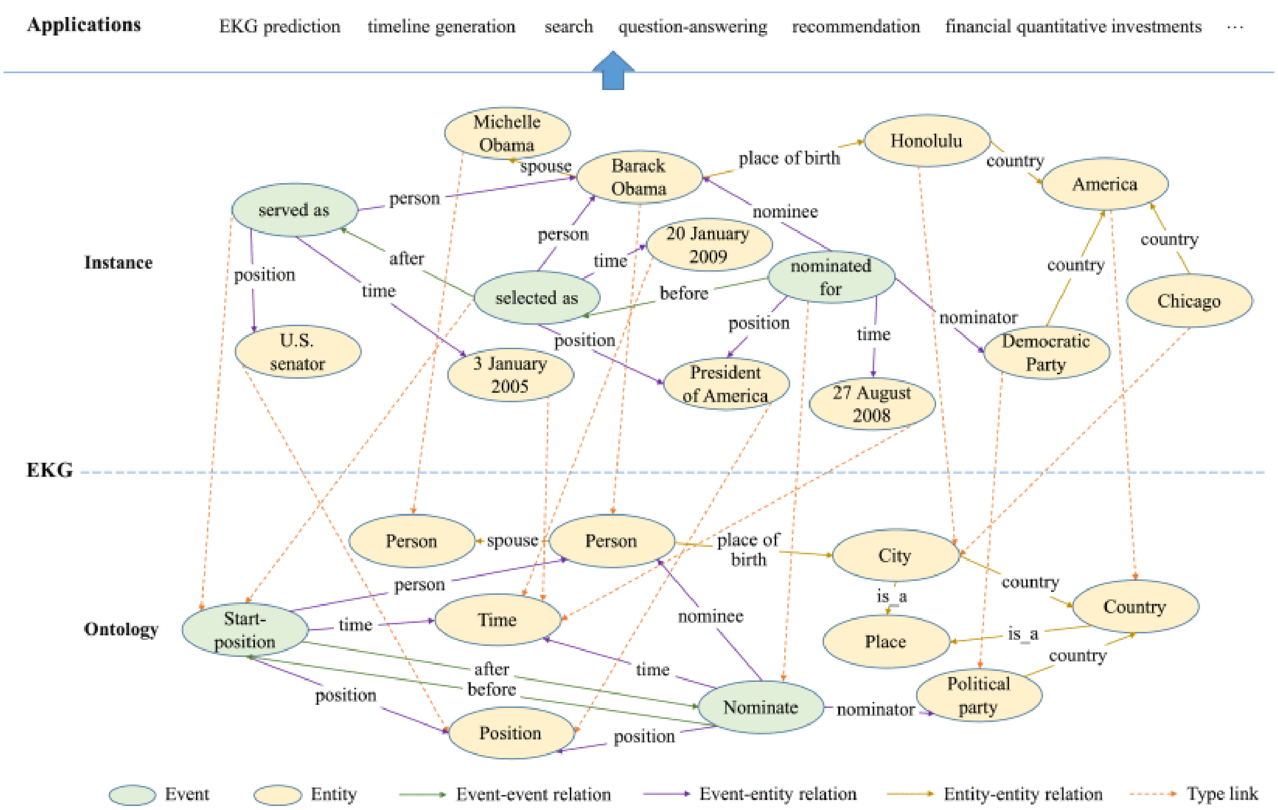


Figure 1. Example structure of the Event Knowledge Graph (Guan et al., 2022)

Despite these integration capabilities, traditional KGs exhibit significant limitations in capturing "events" characterized by temporality and causality. To address this deficiency, Rospocher et al. (2016) introduced an event-centric framework, pioneering the development of EKGs. An EKG comprises two distinct types of nodes, events and entities, and three types of directed edges representing event-event, event-entity, and entity-entity relationships as shown in Figure 1.

In this paradigm, events are elevated from mere entity attributes to independent, interconnected nodes, making EKG flexible and scalable graph capable of modeling the complex temporal and causal relationships. As a result, there are domain-specific EKGs emerging, including the construction of (Fensel et al., 2020) derived from supply chain management text corpora (Deng et al., 2023), as well as Industrial Event Knowledge Graphs (IEKG) designed for manufacturing process management and dynamic assessment models (Li et al., 2025).

To build these graphs from unstructured text, Guan et al. (2022) outlined a standard Event Acquisition Pipeline, which encompasses event extraction, event relation extraction, event coreference resolution, and event argument completion. Recent studies have demonstrated that these complex extraction steps can be significantly augmented through the integration of LLMs (Pan et al., 2024; Trajanoska et al., 2023). While LLMs excel at general language processing, they often suffer from hallucinations and poor explainability. Conversely, KGs provide high factual accuracy and structured reasoning but struggle with semantic understanding and are difficult to construct and update. To bridge this, LLMs can facilitate automated KG construction through entity extraction (Pan et al., 2024), relation extraction (Hu et al., 2025), and Link Prediction for graph completion (Shu et al., 2024) via their generalizability and reasoning ability. Zhu et al. (2024) proposed AutoKG, a multi-agent framework that assigns specialized roles to different agents to automate the entire KG acquisition pipeline.

Furthermore, transforming complex network data into KGs significantly facilitates interactive data exploration, typically utilizing graph databases such as Neo4j (Miller, 2013). In Neo4j, data is accessed and manipulated using Cypher, a language that enables users to directly query and update graph structures. Its distinct advantage lies in an intuitive query format that visually represents the relationships within the network, allowing users to search based on structural patterns rather than complex code (Francis et al., 2018). Leveraging this capability, S.-X. Li et al. (2024) analyzed supply chain networks to evaluate node centrality and structural importance, thereby uncovering potential risk pathways. Furthermore, in the context of LCA, Peng et al. (2024) successfully converted the entire ecoinvent database into a KG hosted on Neo4j, streamlining the querying, mapping, and automated recommendation of background data via Cypher.

2.3 Agentic AI

Agentic Artificial Intelligence (Agentic AI) emerges at the architectural intersection of machine learning (ML), reinforcement learning (RL), and Multi-Agent System (MAS) (Bandi et al., 2025). The foundational brain driving this autonomy is the Reasoning and Acting (ReAct) framework (Yao et al., 2022). ReAct enables AI to solve problems through a human-like cognitive process. Instead of separating decision-making from task execution, ReAct integrates them into a continuous loop. The AI “reasons” to identify missing information, “acts” by calling external tools (such as Python functions), “observes” the outcomes, and “reasons” again to adjust its plan based on new feedback.

While both conventional AI agents and Agentic AI leverage LLMs and ReAct to extend the boundaries of generative intelligence, they diverge fundamentally in terms of structural complexity.

Sapkota et al. (2025) define AI agents as modular, single-entity systems optimized for task-specific automation via prompt engineering and tool integration. In contrast, Agentic AI coordinates multiple specialized agents across a workflow. These agents can share intermediate states, call external tools, and follow conditional routes. This enables Agentic AI systems to autonomously decompose and execute long-horizon, highly complex objectives through continuous environmental interaction and adaptation (Acharya et al., 2025; Hosseini & Seilani, 2025). Furthermore, while traditional MAS also use multiple agents, they often rely on rigid, rule-based logic (Bandi et al., 2025) and struggle to maintain a shared context across interactions (Krishnan, 2025). Modern Agentic AI overcomes this by using the communication protocol and LLM's contextual awareness to handle unpredictable workflows smoothly (Bandi et al., 2025).

In practical applications, LangChain is a popular framework used to build basic LLM applications for AI agents (Topsakal & Akinci, 2023). However, it lacks the iterative loops and conditional logic among multiple agents for dynamic adaptation. To address this, modern Agentic AI frameworks integrate built-in communication protocols, memory management, and safety controls, equipping multi-agent systems with collaborative adaptability (Derouiche et al., 2025; Joshi, 2025). Existing Agentic AI frameworks are suited for distinct operational domains. For instance, Smolagents is suitable for simple, lightweight code-based systems (Gridin, 2025), whereas LlamaIndex excels in data-centric, text-heavy reading tasks (Ponnock et al., 2025). In contrast, LangGraph is highly suitable for industrial systems that necessitate human participation and strict supervision (Xavier et al., 2024).

Compared to other frameworks, LangGraph intentionally limits the degrees of freedom to trade for higher safety and reliability. It provides robust native security support through node-level validation, ensuring that an agent's output or state conforms to expected format standards before transitioning to the next node (Derouiche et al., 2025). Also, LangGraph can employ memory checkpointing, saving intermediate results also allowing directly change the state (Dong et al., 2026), the input and output objects of agents, providing a robust foundation for human interventions. Finally, LangGraph allows developers to switch LLM providers, such as OpenAI, Anthropic, and Meta AI, to node-level without modifying the underlying code or interface (Lei, 2024).

Nevertheless, new challenges within Agentic AI primarily stem from increased systemic complexity. In multi-agent architecture, an over-reliance on LLMs amplifies the risks of hallucinations, weakens causal reasoning, and accelerates error accumulation, leading to compromising system trust, explainability, and verification (Raheem & Hossain, 2025; Sapkota et al., 2026). Furthermore, inadequate shared context among agents and poorly orchestrated workflows can trigger logical contradictions, infinite loops, or resource exhaustion, significantly amplifying overall systemic uncertainty (Sapkota et al., 2026). On the socio-legal dimension, expanding systemic autonomy creates an unprecedented accountability vacuum, obscuring liability attribution among developers and the Agentic AI systems themselves (Hughes et al., 2025; Khalid et al., 2026).

RAG, HITL, and communication protocol are recognized to be effective solutions to overcome the Agentic AI challenges. RAG can search and pull relevant information from specific data stores before generating answers to ensure the generation is grounded in reliable context, significantly reducing errors and making the system much more reliable (Zhao et al., 2026). This mechanism effectively minimizes hallucinations and improves the factual accuracy of the generation (Gupta et al., 2024). Therefore, for unstructured text extraction across massive supply chain datasets, RAG is highly effective at acquiring high-quality, relevant data (Gaddala, 2023). As for HITL, it is an approach actively connecting humans and AI to blend human reasoning with AI computing power (Borghoff et al., 2025). Instead of just acting as backups when things go wrong, humans actively validate decisions, handle complex issues, and govern the system (Bin Osman & Owolabi, 2026). It is also essential for maintaining traceability and accountability in AI systems, especially when workflows involve multiple autonomous agents (Sapkota et al., 2026). Lastly, robust communication protocols and system architectures support coordination among agents by preserving shared context, aligning task objectives, and enabling clear task decomposition (Acharya et al., 2025; Huang & Hughes, 2025) .

3. Methods

This chapter presents the methodological framework developed in this study: an EKG-driven Agentic AI framework that functions as an outside-in sensing layer for LCA background-parameter scenario development in CRM value chains. The framework is designed to monitor external events, structure unstructured evidence, and translate selected event signals into bounded LCI parameter-change scenarios.

During initial exploration, a baseline approach of relying solely on an unconstrained Agentic AI framework to generate scenarios directly from news yielded poor results. Because macro-events reported in public news are qualitative, fragmented, and unstructured, directly prompting an LLM to generate exact numerical parameter changes inevitably leads to hallucination and unjustified data. To overcome these limitations and answer the SRQs, this study adopts a step-by-step evolutionary design. The proposed framework incorporates five critical elements, each specifically designed to bridge the methodological research gaps and finally answer the main research question:

1. EKG as Historical Evidence (Addressing SRQ1): To construct an interpretable and evidence-grounded foundation from unstructured documents, the EKG transforms fragmented news into a structured, event-to-parameter graph. Instead of superficially searching for what a parameter value is, the EKG enables a deep investigation into why and how parameters change by monitoring strategic targets via centrality measures. This structured historical knowledge base provides essential support for parameter matrix definition and the subsequent Graph RAG implementation.

2. HITL as Scoping and Numerical Safeguard (Addressing SRQ2): To translate qualitative news into bounded, quantitative parameter variations without hallucinations, a core methodological philosophy of this framework is to take the numerical decision-making power away from the AI. Human experts conduct LCA scoping to define a scenario boundary matrix. This matrix sets strict user-defined constraints, trigger mechanisms, and percentage-change boundaries, ensuring that the AI operates strictly under human-led rules rather than freely inventing values.

3. Schemas for Structural Control (Addressing SRQ1): To further preserve source traceability and prevent structural hallucinations during extraction, LangChain is utilized to build predefined schemas, which are set of data templates and rules that the model must follow or return, while LangGraph is employed for multiple-task planning and communication. This restricts the LLMs to producing structured, traceable JSON format (JavaScript Object Notation), a commonly data format used for data storage and transfer, rather than ambiguous text paragraphs.

4. Two-Tier RAG for Reliable Reasoning (Addressing SRQ1 & SRQ2): To ensure evidence-based extraction and plausible scenario generation, the framework deploys dual RAG architecture to make the generation result grounded:

- Document-level RAG: LLM-based graders filter incoming news to ensure that only relevant, reliable, and sufficient information is used to build the EKG, thereby maintaining a high-quality evidence base.

- Graph-level RAG: An LLM-based scenario generator evaluates newly monitored events based on hidden relations discovered via LP. This Graph RAG approach translates these structural historical pathways into plain text and feeds them to the LLM alongside the original news document. By reasoning over this combined structural and textual context, the LLM avoids blind guessing, maximizing the traceability and plausibility of the generated scenarios

5. End-to-End System Validations (Addressing SRQ3): Since there are no correct answers for future scenario exploration, this study validates system compliance by measuring how well LLMs follow predefined schemas to generate expected outputs. The validation framework assesses the two main LLM applications in the RAG architecture across two stages: graders and scenario generator. To avoid self-validation, Stage 1 requires a gold-standard dataset separate from the EKG construction data. Stage 2 then reviews reasoning processes and results and validates the plausibility of the final generated scenarios from the Agentic AI framework.

Figure 2 provides a methodological roadmap to show how human-led inputs, Agentic AI framework phases, and validation stages are processed in order.

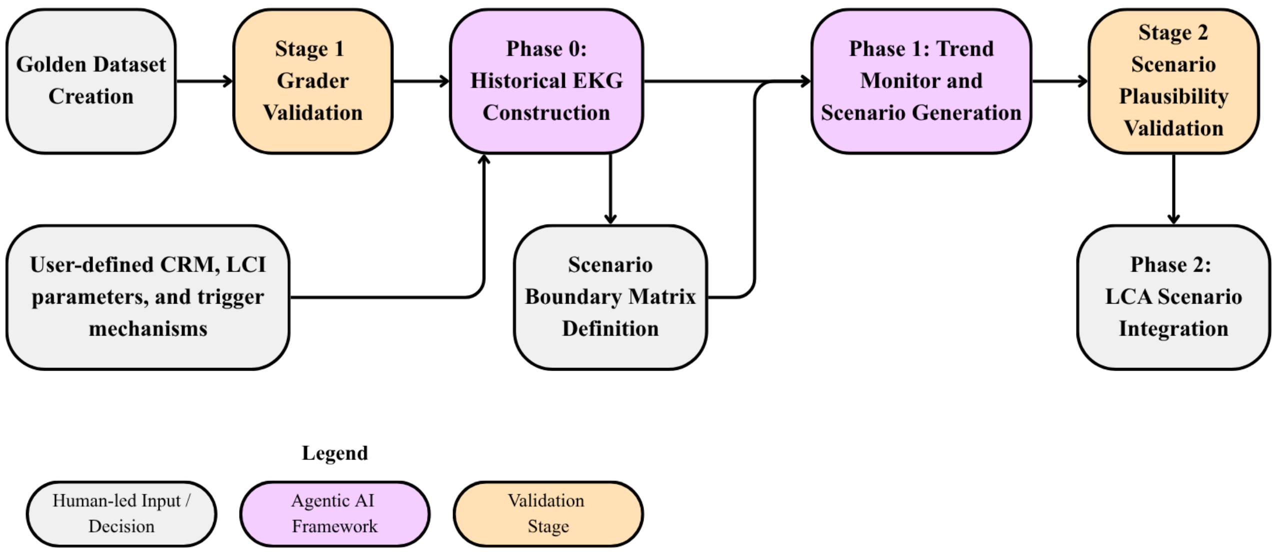


Figure 2. Methodological roadmap of the proposed EKG-driven Agentic AI framework

To clarify the positioning of this framework, it does not directly implement a full Pa-LCA or DLCA model. Instead, it operates as a human-governed, outside-in information-processing layer. The Agentic AI is not treated as a fully autonomous decision-maker; human judgement is required to define the LCA-relevant scope, formulate the scenario boundaries, and evaluate the final outputs.

The chapter is organized as follows:

- Section 3.1 introduces the overall framework design, including its inputs, outputs, phases, and reliability controls.
- Section 3.2 describes the user-defined input and reusability of the framework.
- Section 3.3 explains the reason why choose the graphite value chain as case study, and the selected LCI parameters and trigger mechanisms.
- Section 3.4 explains the construction and analysis of the historical EKG (Phase 0).
- Section 3.5 explains the purpose and construction of the scenario boundary matrix.
- Section 3.6 describes the real-time monitoring and scenario generation workflow (Phase 1).
- Section 3.7 presents the validation approach used to evaluate document graders, scenario plausibility, and boundary compliance.

3.1 Framework Design for External Trend monitoring and LCA Scenario Generation

Figure 3 presents the EKG-driven Agentic AI framework and its interaction with the LCA framework. It starts with user-defined LCA inputs: the selected CRM value chain, the product system of interest, the sensitive LCI parameters, and the mechanisms through which external events may influence these parameters. The framework consists of three connected phases:

- Phase 0 builds a historical EKG from documents related to predefined LCI parameters and their trigger mechanisms, as well as identifying monitoring targets.
- Phase 1 uses these monitoring targets to detect new external events and generate three scenario-based percentage changes compared with the similar cases in the historical EKG.
- Phase 2 is conceptually aligned with Pa-LCA and DLCA because it evaluates how changes in selected LCI parameters influence LCIA results. Instead of implementing a full LCA, this study applies parameter changes into a simplified formula to estimate change in climate-change impact compared with the baseline.

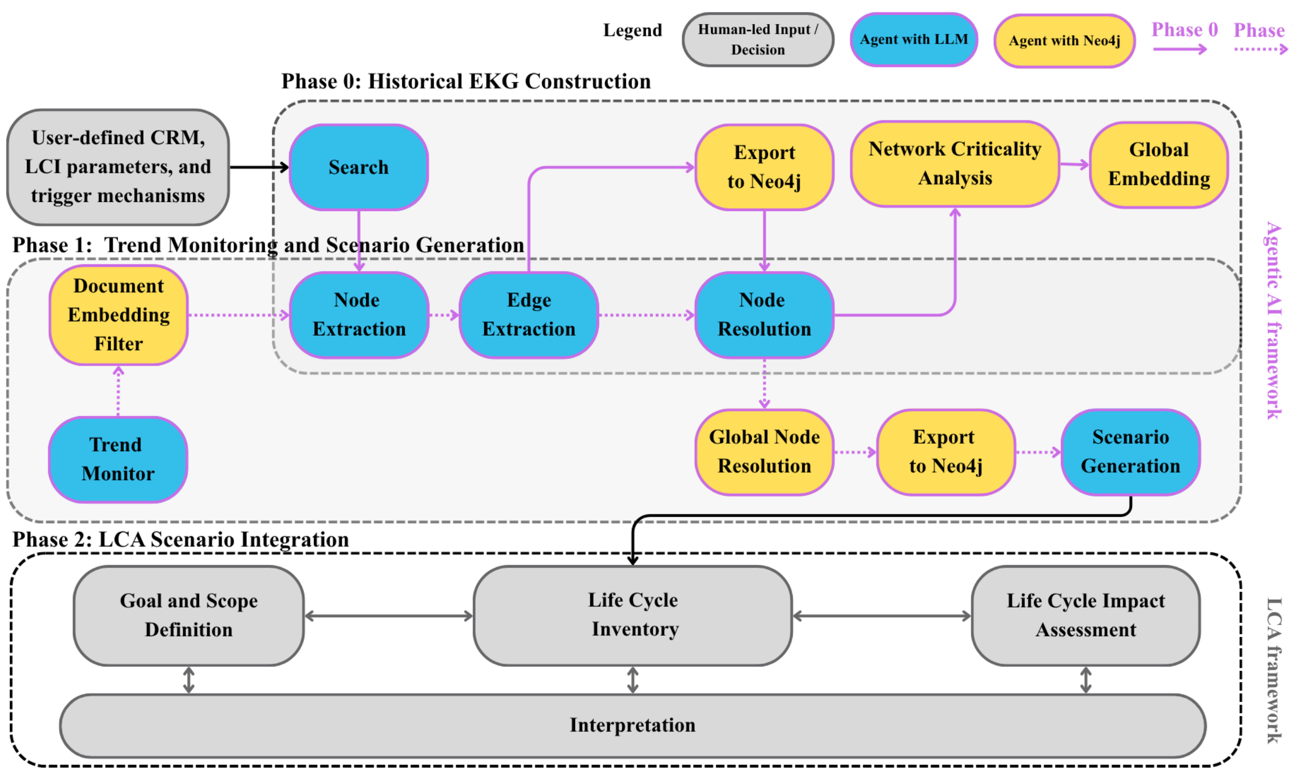


Figure 3. EKG-driven Agentic AI framework for monitoring external trends and generating dynamic scenarios in LCI parameters for CRM value chains.

The framework produces three main outputs. The first output is a historical EKG from Phase 0, which is fundamentally composed of its historical profile and parameter trajectories, which map the event-to-parameter pathways within the network. Because the EKG represents a complex system, the framework applies network centrality measures from a systematic perspective to detect critical nodes. This approach identifies strategic nodes that possess high systemic influence, meaning any changes to them would trigger significant cascading effects. For example, within the graphite value chain, a specific geographic location, such as China or specific province, emerges as a critical node due to its high degree centrality (Pandey et al., 2026). Because it interconnects with numerous other entities and events, it becomes an essential target for continuous monitoring. Ultimately, these critical nodes are designated as the monitoring targets for Phase 1 and provide insight for construction of the scenario boundary matrix.

The second output is a set of best-, average- and worst-case percentage changes for selected LCI parameter from Phase 1, while the third output is the estimated change in LCA results after applying these scenario values from Phase 2.

This study uses LangChain and LangGraph as implementation infrastructure for the Agentic AI workflow. In simple terms, LangChain structures what each individual agent produces, while LangGraph structures how agents communicate and proceed through the workflow. Together, they allow this study to combine the reasoning ability of LLMs with predefined data templates, making the output more controllable and traceable. More working logics, details and example codes about LangChain and LangGraph are provided in Appendix A.

3.2 CRM, LCI Parameters, and Trigger Mechanisms Definition

Before Phase 0 begins, the selected CRM value chain, sensitive LCI parameters, and their possible trigger mechanisms should be defined in advance. To establish clear system boundaries for the Agentic AI framework and focus on LCI parameters that significantly affect LCA results and are prone to external shocks, users need to identify which LCI parameters are most "sensitive". This selection can be reliably based on Contribution Analysis or Global Sensitivity Analysis from existing LCA studies. Specifically, the selection of a target parameter can be guided by three key criteria:

1. *Large Expected Contribution*: The parameter accounts for a significant share of the overall environmental impact baseline. Examples include regional electricity carbon intensity, and the power generation mix for energy-intensive CRM value chains (e.g., graphite, aluminum, silicon). Even marginal variations in these values will significantly alter the final LCA results.

2. *High Variability*: The parameter is known to fluctuate significantly or frequently over time, making traditional static average values unreliable for accurate assessments. Examples

include transnational transport distances, logistical routes, and the geographic sourcing mix of raw materials, all of which could shift rapidly within a short period.

3. *High External Dependency*: The parameter is susceptible to, or indirectly driven by, macro-events outside the factory gates, such as geopolitical conflicts, shifting trade policies (e.g., sudden export controls or tariffs), or extreme weather conditions.

In addition, this step is necessary because the framework relies mainly on public news, yet not all public information is relevant to LCA, nor can every LCI parameter be monitored through public sources. Many process-specific or engineering parameters, such as specific heat, are only reported in technical studies or primary inventory datasets rather than in news articles. It is also important to note that this study differs from general CRM supply-chain risk monitoring. Rather than aggregating all news related to general supply disruptions or demand fluctuations (Cheng et al., 2024), this framework strictly focuses on external events that impact specific LCI parameters.

Furthermore, simply searching for an LCI parameter name is insufficient for trend monitoring. Differences in reported parameter values may result from different modelling assumptions or data sources, rather than from real external events. Therefore, the framework must monitor the external mechanisms that could plausibly change the parameter, not only the parameter itself.

For these reasons, expert-defined trigger mechanisms are required to force Agentic AI to focus on what is both LCA-relevant and observable in the public news. They specify the physical, logistical, regulatory, or geopolitical pathways through which external events may affect the selected LCI parameters. By combining parameter names with trigger mechanisms, the agents can retrieve documents that explain how and why a parameter may change. This makes later scenario generation more LCA-relevant. It also prevents the LLM from freely deciding which parameters or events are important based only on general knowledge.

Beyond mitigating AI hallucinations, defining these inputs highlights the reusability of the established framework. Because the core Agentic AI framework is already built and is not tied to a specific CRM value chain, applying it to a new case simply requires updating these initial definitions and slightly calibrating the prompts. However, when defining the scope for a new application, it is currently recommended to focus on a simplified value chain for a single material, such as graphite anodes. Targeting a complex, multi-component product like a complete lithium-ion battery would introduce excessive noise and computational burden at this stage of development.

3.3 Case Study: Graphite Value Chain

This study uses the graphite value chain as the case study. Graphite was selected because it is a key material for lithium-ion battery anodes and because its value chain is highly geographically concentrated. Although graphite itself is geologically abundant, battery-grade graphite production and processing are dominated by a limited number of regions (Park et al., 2025), especially China. China plays a dominant role in both upstream graphite supply and anode-material processing. Previous study reports that China accounts for approximately 79% of natural graphite supply and 78% of synthetic graphite supply, while also holding a major share of midstream anode-related processing capacity (Ritoe et al., 2022).

This concentration makes the value chain sensitive to geopolitical tensions and logistics disruptions. As a result, this case is relevant because external events may affect background LCI parameters from an LCA perspective. In addition, policy developments such as the EU Critical Raw Materials Act and the EU Battery Regulation increase the demand for transparent supply-chain monitoring and carbon-footprint disclosure (Surovtseva et al., 2022). Therefore, graphite provides a suitable case for testing the proposed framework under high uncertainty.

Within this framework, two high-sensitivity LCI parameters are selected: the carbon intensity of the electricity mix in China and graphite supply chain transport distance and routing, as shown in Table 1. These two parameters were selected because they are both environmentally influential and externally observable.

Carbon intensity of the electricity mix in China

The carbon intensity of China's electricity mix was then selected as one of the monitored LCI parameters because electricity use is a major contributor to the environmental impact of graphite processing. Battery-grade graphite production involves energy-intensive processing steps, and therefore the carbon intensity of the electricity used during production can strongly influence the resulting carbon footprint.

This is particularly relevant for China, where graphite processing is highly concentrated and where reported emissions for natural graphite processing are higher than those reported for many non-Chinese production routes. For example, Park et al. report an average of 9.6 kg CO_2e/kg for natural graphite processing in China, compared with approximately 3-7 kg CO_2e/kg for non-Chinese production routes (Park et al., 2025). Therefore, this study monitors China's electricity carbon intensity as a background LCI parameter that is both environmentally significant and externally observable.

The carbon intensity of electricity is not fixed. It can change due to shifts in electricity generation, weather-related disruptions, infrastructure development, and broader policy or geopolitical conditions. This study therefore defines four trigger mechanisms for monitoring this parameter.

First, fuel switching may occur when electricity shortages increase reliance on coal- or gas-fired backup generation (Auffhammer et al., 2021; Dong et al., 2018). Second, extreme weather events, such as droughts, may reduce hydropower availability and increase the share of fossil-based generation (Li et al., 2024). Third, infrastructure and policy changes, such as renewable-energy integration, grid expansion, and storage deployment, may affect the electricity mix over time (Dong et al., 2018). Fourth, international policy and trade tensions may indirectly affect clean-energy investment, technology access, or deployment speed, thereby influencing the future electricity mix (Attílio, 2026). These mechanisms provide a focused search scope for identifying external events that may plausibly affect the carbon intensity of electricity used in graphite-related LCA scenarios.

Graphite supply chain transport distance and routing

Graphite supply-chain transport distance and routing were selected as another monitored LCI parameters. In graphite-related LCA models, these transport assumptions might be treated as fixed background parameters. However, in practice, they can change substantially when external events affect global logistics.

Recent disruptions show that shipping routes can change quickly due to geopolitical tensions, regulatory barriers, and climate-related events. These disruptions may increase transit distance, extend delivery time, or shift transport modes (Notteboom et al., 2024). For example, during the Suez Canal disruption, container vessels were forced to reroute via the Cape of Good Hope. This longer route increased total shipping fuel-cycle emissions, meaning emissions from both marine fuel production and onboard fuel use, by approximately 48.6% (Yue et al., 2024). At the same time, the emissions for a Singapore-Rotterdam round trip could increase by around 70% under such rerouting conditions (UNCTAD, 2024a). These examples are not used to represent graphite shipping directly, but to demonstrate how route changes can significantly affect transport-related emissions.

Transport disruptions can also create indirect effects across the logistics network. A disruption in one corridor may cause vessels to reroute, and generate congestion or delivery delays elsewhere (Ghosh & Sourav, 2026). In addition, policy changes such as export controls, tariffs, and efforts to reduce dependency on Chinese graphite may shift sourcing patterns toward alternative mining regions or domestic processing hubs (Park et al., 2025). Such sourcing shifts can change origin-destination pairs and therefore alter the transport distance and route assumptions used in LCA models.

For these reasons, this study monitors graphite supply-chain transport distance and routing through three trigger mechanisms. The first is route disruption, where a shipping corridor becomes unavailable or less attractive and vessels are forced to take longer routes. The second is port congestion spillover, where disruption in one part of the network causes delays or rerouting through other ports. The third is sourcing shift, where changes in trade policy or supply-chain strategy alter the geographical origin or processing location of graphite.

Table 1. Key LCI parameters selected for the graphite value chain case study

LCI Parameter name	Description
Carbon intensity of electricity mix in China	The carbon emission factor of the regional power grid used by midstream processing plants (specifically for Spheroidization, Purification, and Graphitization). This parameter fluctuates based on (1) fuel switching, (2) extreme weather, (3) evolving infrastructure policies, and (4) geopolitics spillover.
Graphite Supply Chain Transport Distance and Routing	Measures changes in the physical transport distances or routing of graphite and anode materials. Changes are driven by (1) forced geographical route disruptions, (2) port congestion spillover, or (3) strategic sourcing shifts.

3.4 Phase 0: Building the Historical Event Knowledge Graph

Since the supply chain is complex to understand the causal and temporal relationship, the pre-built database is essential for supply chain downstream applications. Sprecher et al. (2017) conducted actor-centric Event Sequence Analysis (ESA) to evaluate resilient mechanisms and calculate indicators for the REE supply chain toward the 2010 crisis. Cheng et al. (2024) conducted schema induction on the library to compare the similarity of current news to predict supply chain interruption.

Building upon these concepts, this phase constructs a dedicated knowledge base to map the networks driving key background parameter changes in the past. While inspired by the (Zhu et al., 2024) who divided KG construction into sub-tasks managed by chatbot-like agents, the agentic workflow in Phase 0 integrates LLM reasoning, external APIs, and Neo4j Cypher to directly save, query, and update EKG, moving beyond simple chatbot interactions to achieve autonomous EKG construction. To avoid terminological confusion with EKG nodes, the functional nodes within LangGraph are hereafter termed “agents.”

The EKG organizes the 5W1H information of each event: who was involved, where it occurred, what material or parameter was affected, when it occurred, why it happened, and how it influenced the selected LCI parameter. Therefore, nodes are classified into two broad categories: Entities (parameter, actor, location, Material/Product) and Events (Macro-Trend, Micro-Event, Parameter Observation). The specific nodes are parameter observation, which record the parameter shift over time instead of only storing in the attributes inside the parameter nodes or edges. The interaction of nodes and edges is presented in Figure 4.

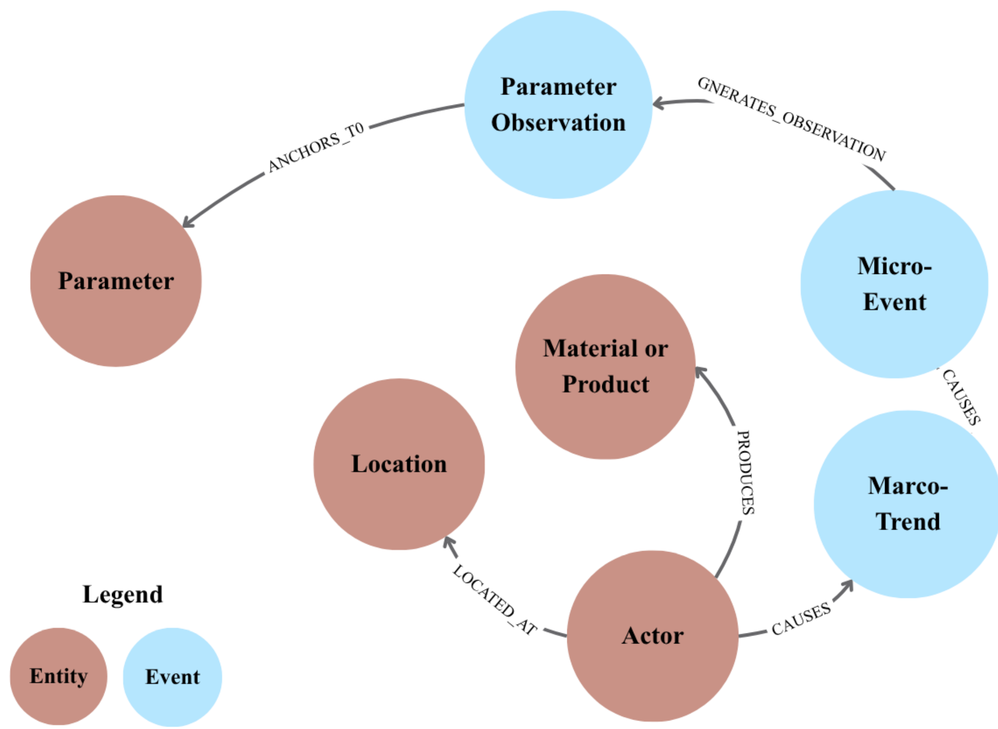


Figure 4. Interaction example of node and edge types in the proposed EKG schema

Phase 0 Starts with the **Search** agent in Figure 5, which retrieves historical news articles and public reports related to predefined parameters and mechanisms and filters out irrelevant news with two graders. Specifically, the document grader evaluates the text to ensure source reliability, relevance, and traceability to parameter variations. Following this, the answer grader decides whether the document contains how and why context regarding the trigger.

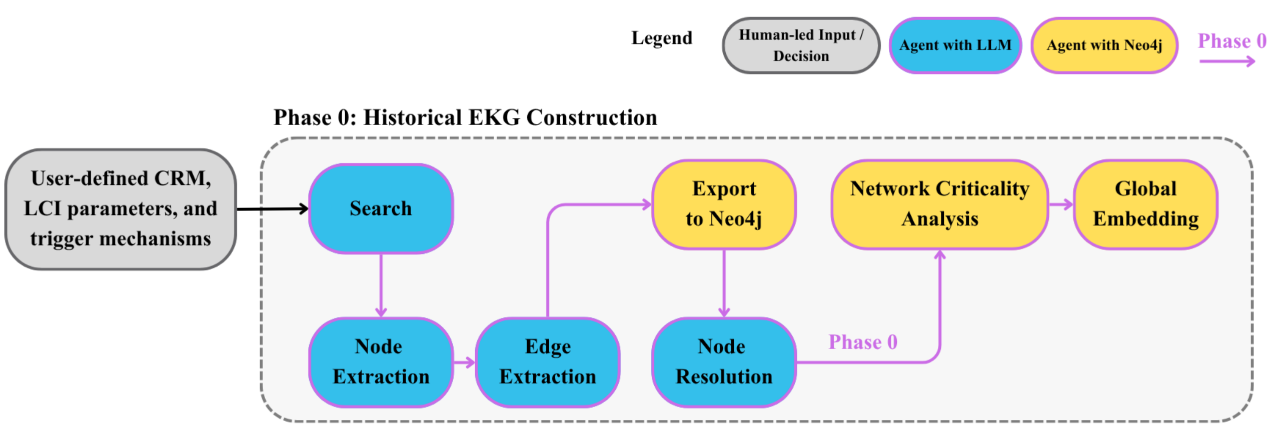
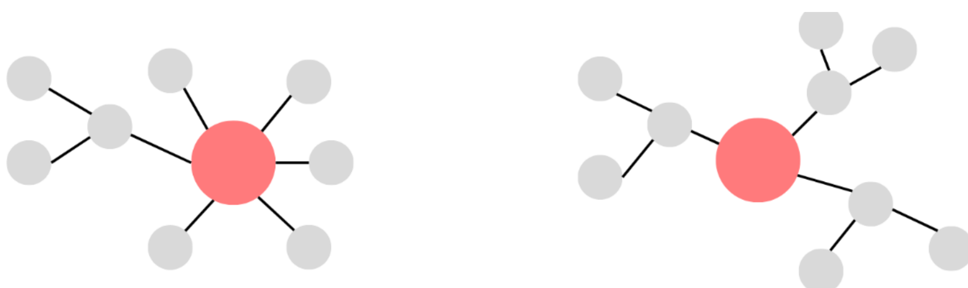


Figure 5. Overview of the Phase 0 workflow

These filtered documents are then passed to the **Node Extraction** agent, which extracts relevant entities and events concepts according to the predefined node schemas. Next, the **Edge Extraction** agent reasons for the relationships among the extracted nodes based on the original documents.

After initial graph is ready, the **Export to Neo4j** agent imports it into Neo4j to create the initial graph. The **Node Resolution** agent then refines the graph by merging nodes with similar or overlapping meanings and reconnecting their relationships, allowing the graph to converge into a coherent network-based EKG.

The refined EKG is subsequently processed by the **Network Centrality Analysis** agent, which analyzes the graph structure, calculates centrality, and identifies parameter-related event trajectories. Following centrality definition by Pandey et al. (2026) as shown in Figure 6, this agent uses two network centrality measures to identify system hubs and chokepoints as the monitoring targets.



(a) A node with a high degree centrality score

(b) A node with a high betweenness centrality score

Figure 6. Centrality-based identification of a System Hub and a Chokepoint in the EKG

Finally, the **Global Embedding** agent assigns embedding attributes to each node in the EKG. Embedding refers to translating real-world objects into mathematic index/vector, so that computer can understand it. These embeddings support semantic and structural understanding in the EKG and further reasoning for agents in Phase 1.

An overview of each agent's specific role and assigned toolsets is summarized in Table 2. For a comprehensive technical breakdown, including the conceptual and implementation models for embeddings, please refer to Appendix B.

Table 2. Agent role and its tools in Phase 0

Agent name	Role	LLM application	External API/tool
Search	Search historical documents related to the causality and participants involved in change in predefined LCI parameters	Query Generator, Document Grader, Answer Grader	Exa, Sentence Transformer
Node Extraction	Extract entities and events from the following document scheme for node	Node Extractor	
Edge Extraction	Extract relation from the document following scheme for edge	Edge Extractor	
Node Resolution	Conduct entity resolution and event coreference to merge and clean the EKG	Node Resolver	
Export to Neo4j	Connect with Neo4j and export the final EKG JSON file into it		Neo4j Cypher
Network centrality Analysis	Conduct the Cypher to analyze (1) Graph Profile (2) Strategy Nodes (3) Parameter trajectories	Report Generator	Neo4j Cypher
Global Embedding	Conduct whole graph embedding and add computed results to Neo4j		Neo4j Cypher

3.5 Scenario Boundary Matrix Identification

The boundary matrix serves as a critical numerical safeguard during Phase 1 scenario generation. Its primary purpose is to constrain the LLM's numerical reasoning within commonsense, physically plausible limits, preventing it from freely inventing unrealistic figures. The numerical ranges within this matrix are defined mainly by human expert judgments, supported by evidence extracted from the EKG, supplementary industry reports, and academic literature.

Many news documents do not report directly usable parameter changes; instead, they often describe trends or sentiments rather than absolute numbers. To address this limitation, this study employs a mechanism-oriented, semi-quantitative classification approach. Rather than forcing the LLM-based scenario generator to calculate exact values from incomplete text, the framework prompts it to identify the underlying disruption mechanism and classify the event's severity.

These mechanisms are categorized into specific "disruption patterns," each linked to predefined severity levels and numerical boundaries within a boundary matrix. This matrix serves as a critical numerical safeguard, constraining the LLM's reasoning within common sense, physically plausible limits. The ranges are derived from EKG evidence, literature reviews, and human expert judgments, and they dynamically adjust based on the classified severity.

For example, as illustrated in Figure 7, if the LLM processes news about a heatwave limiting hydropower generation, it identifies the underlying mechanism as an "increasing coal share in the electricity grid" and classifies the event as "Level 3: Severe." The boundary matrix then restricts the resulting carbon intensity of electricity increase to a plausible range of [3%, 6%] based on similar cases in EKG and IEA report (IEA, 2026). This safeguard prevents the AI from freely inventing an extreme scenario, such as a 100% increase in carbon intensity, which contradicts the inherent stability of regional power grids, ensuring that the simulated environmental impacts remain physically grounded and practically realistic. Furthermore, in cases where absolutely no prior information is available, the matrix functions as an exploration and transition mechanism, allowing the framework to conduct exploratory "what-if" analyses and uncover hidden vulnerabilities in uncertain environments.

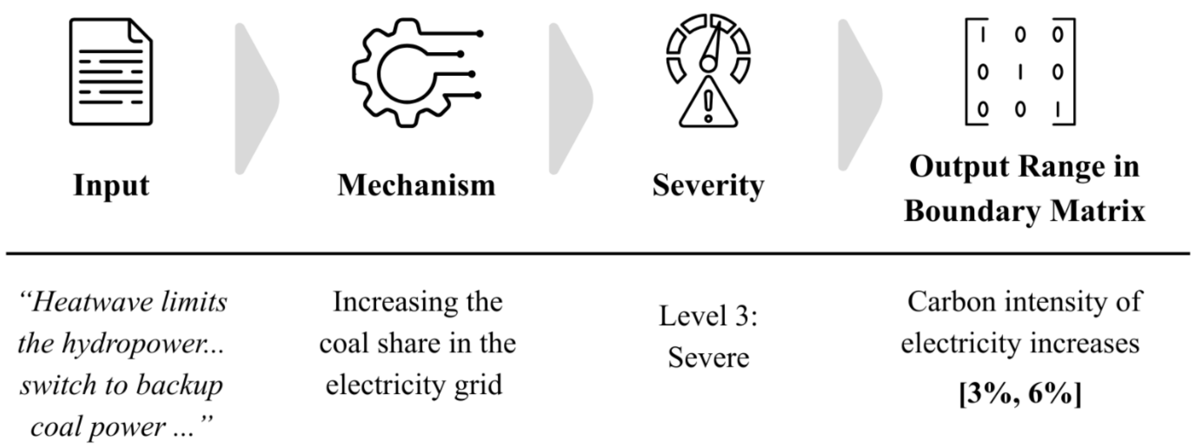


Figure 7. Text-to-number pipeline in scenario boundary matrix construction

This scenario boundary matrix should be interpreted as a case-specific and exploratory constraint rather than as a validated forecasting model. Although the ranges are informed by EKG-derived evidence, relevant literature, and expert judgement, the final numerical boundaries still partly depend on case-specific interpretation. Therefore, the selected percentage ranges and time windows should not be treated as universally valid thresholds. Instead, their main purpose is to prevent the LLM from generating arbitrary or unbounded scenario values, making scenario generation more transparent, controlled, and reproducible. The subjectivity and limited generalizability of the boundary matrix are therefore acknowledged as methodological limitations.

3.6 Phase 1: Trend monitoring and Scenario Generation

Phase 0 and Phase 1 employ distinct state structures for communication and operate at different temporal frequencies to suit their specific architectural roles. As a real-time monitoring and scenario generation engine, Phase 1 requires highly frequent, periodic execution (e.g., daily or tri-weekly). In contrast, Phase 0 updates at longer intervals to maintain the foundational database and redefine monitoring targets, ensuring the system remains adaptively aligned with broader trends while optimizing computational efficiency.

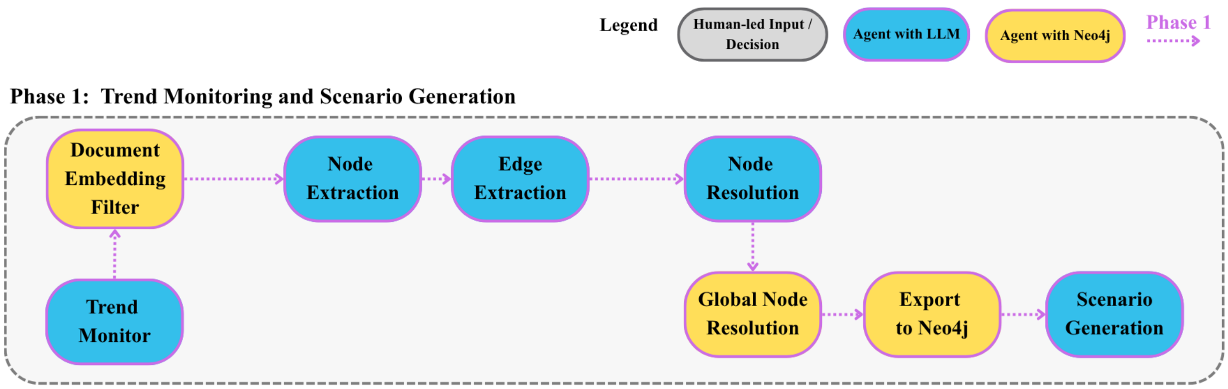


Figure 8. Overview of the Phase 1 workflow

Phase 1 performs real-time monitoring and scenario generation by updating the historical EKG with newly observed events. As shown in Figure 8, The workflow begins with the **Trend Monitoring** agent, which uses the strategic nodes identified in Phase 0 as search anchors to retrieve recent news related to emerging trends or changes. The retrieved documents are then processed by the **Document Embedding Filter** agent, which removes duplicate or highly similar documents to prevent redundant graph updates.

The filtered documents are subsequently transformed into a mini-EKG JSON file using the same agents with slightly modification in Phase 0: Node Extraction agent, Edge Extraction agent, and Node Resolution agent. The **Global Node Resolution** agent then aligns the mini-EKG with the original historical EKG by resolving overlapping or semantically similar nodes across the two graph structures. After this integration step, the **Export to Neo4j** agent imports file including the newly generated nodes and edges into Neo4j.

Finally, the **Scenario Generation** agent evaluates if there are hidden relations between the newly added events and the parameter nodes in the EKG. It examines whether these new events show structural similarity to historical patterns. If there is potential relation, the source news of the events, similar historical reference (parameter observation), and the scenario boundary matrix will be assembled and help LLM contextual reasoning to generate bounded parameter changes under three scenario cases.

Table 3 summarizes the overviews of each agent's specific role and assigned toolsets in Phase 1. Detailed descriptions of the agent architecture, please refer to Appendix C.

Table 3. Agent role and its tools in Phase 1

Agent Name	Role	LLM	External API/Tool
Trend Monitoring	Search related news about strategy nodes	Monitoring Query Generator, Document Grader, Answer Grader	Exa
Document Embedding Filter	Filter out the similar monitored news existing in the original EKG		Neo4j, OpenAI embedding model
Global Node Resolution	Conduct entity resolution and event coreference to merge the updated EKG and original EKG	Global Node Resolver	Neo4j Cypher
Scenario Generation	Based on the LP between updated events and parameters and the news, generate the best, average, and worst scenarios within the parameter scenario matrix	Scenario Generator	Neo4j Cypher

3.7 End-to-End System Validation

Validation is necessary to assess the truthfulness, reliability, and task alignment of the proposed LLM-based framework. Although LLMs can generate fluent and coherent text, they may also produce unsupported or fabricated information (Wang et al., 2023), commonly referred to as hallucination, which limits their use in domain-specific and evidence-based applications (Huang et al., 2025). Therefore, this study validates the LLM-based applications at two critical stages of the workflow: the initial document-retrieval stage, which determines the quality of the evidence entering the system, and the final scenario-generation stage, which determines the validity of the generated parameter-change scenarios.

The prompts used by the graders and scenario generator were iteratively refined as part of system development. Changes in validation results were used to assess the effects of prompt refinements and to ensure that the output was sufficiently reliable, consistent, and suitable for integration into the agentic workflow. Overall, the validation was conducted outside the Agentic AI framework using independent evaluation scripts and, for reasoning quality only, an independent LLM-as-a-Judge, used as a supplementary evaluator and not as the sole basis for determining scenario validity.

3.7.1 Golden Dataset and Validation Criteria

Given the current scarcity of benchmarks within the Agentic AI (Acharya et al., 2025), especially in domain for CRM value chains and LCA context, this study created a customized golden dataset for validation. The dataset was designed to evaluate whether the LLM-based graders could stably classify the positive and negative cases.

The golden dataset focused on parameter-related news published between 1 January 2026 and 31 May 2026. This period was selected as an out-of-period validation set because Phase 0 used historical documents from 2020 to 2025 for EKG construction.

To avoid using the system to evaluate itself, dataset construction separated candidate collection, human annotation, and model evaluation. The Search Agent was used but only served as an advanced search engine to retrieve a diverse pool of candidate documents. It did not assign the final labels and was not used to determine correctness.

The golden dataset creation followed three steps. First, the Search Agent retrieved a diverse set of potentially relevant news documents from the target period. In this step, the Search Agent was used only as an advanced search tool for candidate collection and did not assign final labels. Second, the candidate documents were manually annotated by the researcher according to the predefined ground truth metrics. These criteria determined whether each document provided sufficient evidence of a plausible parameter-level effect. Third, the final dataset was balanced with 26 positive cases and 26 hard negative cases to test whether the graders could distinguish physically relevant events from superficially related news. The ground truth metrics and examples for the hard negative cases are provided in Appendix D.

3.7.2 Stage 1: Grader Validation

Stage 1 evaluates the classification performance and robustness of the LLM-based graders used in the Search and Trend monitoring agents. From an LCA perspective, this validation asks whether the graders can correctly identify documents that provide a plausible basis for changing selected CRM or LCI parameters.

This step is necessary because the EKG depends on the quality of its input documents. Irrelevant or weakly supported sources may introduce noisy event nodes and unsupported parameter links. The document grader is tasked with filtering out document sources that were not credible, not relevant to the target parameter, or lacking factual traceability. Next, answer grader decides the document can contain sufficient information about the parameter and trigger mechanism. These criteria were implemented through a structured schema but are reported here as three screening dimensions: source credibility, parameter relevance, and factual traceability.

In simple terms, Stage 1 checked whether the graders made the same classification as the human annotation. Each document had a human label, which served as the ground truth, and a grader label, which was produced by the document grader and answer grader. If the grader label matched the human label, the classification was counted as correct. If the two labels differed, the classification was counted as an error. The comparison was summarized using a confusion matrix (Sathyanarayanan & Tantri, 2024). As illustrated in Figure 9, True Positives (TPs) indicate that both the human annotator and the graders identified the document as positive. True Negatives (TNs) indicate that both identified it as negative. False Positives (FPs) occur when the graders classified a human-labelled negative document as positive. False Negatives (FNs) occurs when the graders rejected a human-labelled positive document.

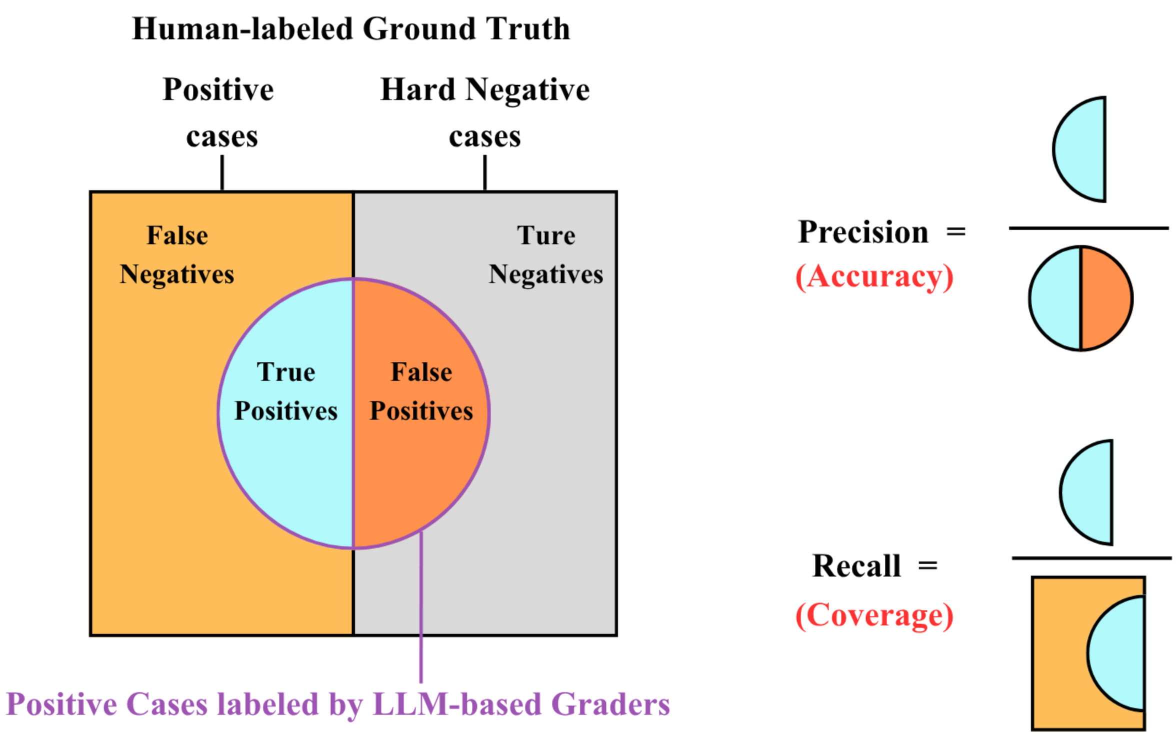


Figure 9. Visual representation of classification outcomes (TP, FP, TN, FN) and evaluation indicators (precision and recall)

Subsequently, precision, recall, and F1 were calculated from TP, FP, and FN below (Equation 1, 2, and 3) as Naidu et al. (2023) defined and illustrated in Figure 8. In this study, precision measures how many documents accepted by the graders were truly relevant, filtering out FP. High precision score indicates that the graders avoid introducing irrelevant or misleading evidence, while recall measures how many truly relevant documents were successfully captured. High recall indicates that the system does not miss important events that may affect LCI or CRM parameters. F1-score is the harmonic means of precision and recall and represents the balance between avoiding FP and avoiding FN.

$$Precision = \frac{TP}{TP+FP} \quad \text{(Equation 1)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{(Equation 2)}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad \text{(Equation 3)}$$

To assess robustness, the evaluation was repeated across multiple runs after the grader prompts were fixed. The mean and standard deviation of precision, recall, and F1-score were then reported. A predefined threshold of F1-score > 0.70 was used as the acceptable quality gate for retaining graders for subsequent application in the graphite value chain case study.

3.7.3 Stage 2: Scenario Plausibility Validation

Stage 2 evaluates logical plausibility of the scenarios generated by the Scenario Generation agent. Since future scenarios do not have an absolute ground truth, this stage does not evaluate predictive accuracy. Instead, it evaluates whether the generated scenarios are logically plausible, numerically valid, and consistent with expert-defined constraints.

First, numerical consistency checks whether the generated best, average, and worst values follow the expected logic for the selected parameter. The expected ordering depends on the direction of the parameter change and its effect on the LCA result. For example, when the scenario represents an increase in grid carbon intensity, the best case should correspond to the smallest increase, the average case to a moderate increase, and the worst case to the largest increase. Second, boundary compliance checks whether the generated values remain within the expert-defined scenario boundary matrix. These two criteria were evaluated using rule-based checks implemented in independent deterministic code, as shown in Figure 10.

```
# [Partial Extraction] Deterministic Guardrail Script for Monotonicity and Boundary Checks
f_best = float(best)
f_avg = float(avg)
f_worst = float(worst)
tolerance = 0.002  # Relaxed to prevent false negatives from floating point rounding

# 1. Numerical Monotonicity Check
if direction_clean == "Increase":
    if not (f_best > 0 and f_avg > 0 and f_worst > 0):
```

```
        eval_result["reason"] = f"Violation: values must be positive."
        final_evaluation_statistics.append(eval_result)
        continue

    if not (f_best <= f_avg <= f_worst):
        eval_result["reason"] = f"Mapping Violation: expected best <= avg <= worst."
        final_evaluation_statistics.append(eval_result)
        continue

# ... [Code for "Decrease" and "None" directions omitted for brevity] ...
# ... [Logic for fetching expert-defined lower_bound and upper_bound omitted] ...

# 2. Boundary Compliance Check
if direction_clean == "Increase":
    # Ensure generated values do not violate the Scenario Boundary Matrix
    if f_best < (lower_bound - tolerance):
        eval_result["reason"] = f"Boundary Violation: Best case ({f_best}) is below
Tier lower bound ({lower_bound})."
        final_evaluation_statistics.append(eval_result)
        continue

    if f_worst > (upper_bound + tolerance):
        eval_result["reason"] = f"Boundary Violation: Worst case ({f_worst}) exceeds
Tier upper bound ({upper_bound})."
        final_evaluation_statistics.append(eval_result)
        continue

# If all deterministic checks pass
eval_result["boundary_passed"] = True
```
Figure 10. Example code of numerical consistency check and boundary compliance check in stage 2 validation

The last criteria, reasoning quality was evaluated using an independent LLM-as-a-Judge (Gu et al., 2026). The judge assessed whether the scenario explanation was consistent with the cited evidence, the identified mechanism, and the selected severity level. Each explanation was scored on a scale from 1 to 5, where higher scores indicate stronger evidence, support and better logical coherence. A scenario required a reasoning score greater than 3 to pass this component of the validation. These validated scenarios were then passed to Phase 2 for LCA scenario integration. The scenario plausibility metrics and score sheet for LLM-as-a-Judge is provided in Appendix D.

4. Results

This chapter presents the results according to the sequential input-output workflow of the proposed Agentic AI framework. As shown in Figure 11, the result-processing structure moves from human-labelled data and validation to historical EKG construction, event-based scenario generation, scenario plausibility validation, and final LCA scenario integration. This workflow is used to clarify how each output is produced from the previous stage and how human-led inputs, Agentic AI framework processing phases, and validation stages interact throughout the study.

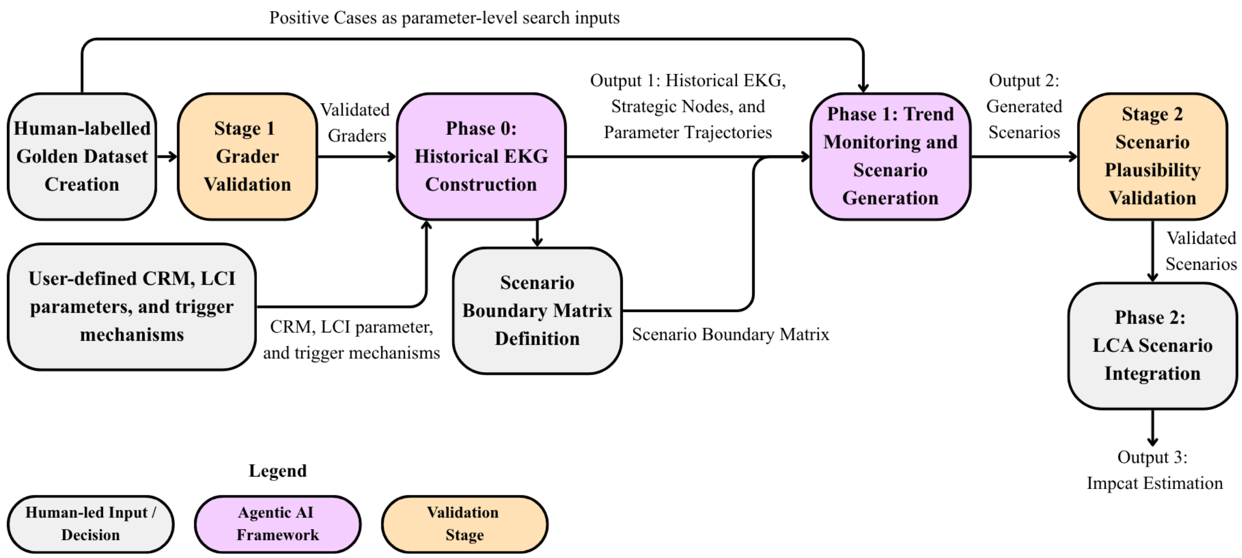


Figure 11. Result-processing workflow and input-output structure of the proposed framework

4.1 Stage 1: Grader Validation

The validation used the golden dataset consisting of 26 positive and 26 negative cases. For each round, five independent validation runs were conducted to calculate the confusion matrix (TP, TN, FP, FN), precision, recall, and F1-score. The mean and standard deviation are shown in Figure 12.

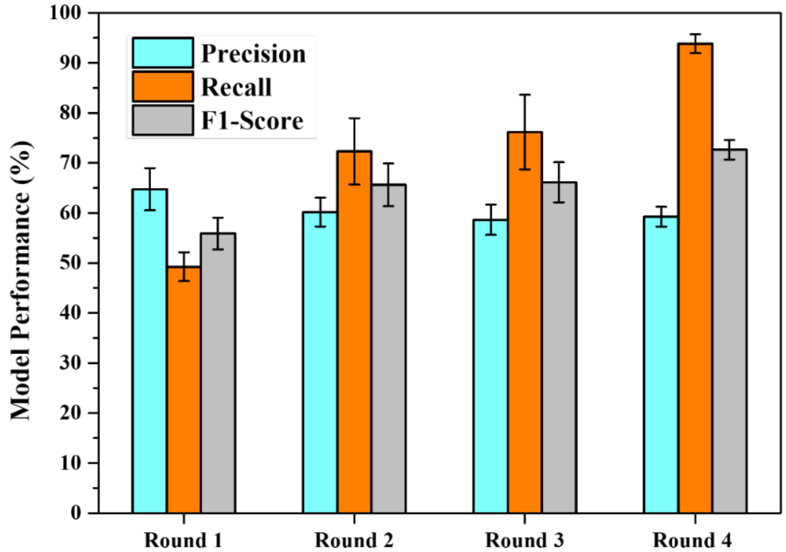


Figure 12. Grader validation results across four prompt-tuning rounds

Round 1 tested the grader without providing it with any previous examples. Although the model produced stable results across runs, it missed many relevant impact events and accepted a considerable number of false positives. This showed that the baseline prompt was not sufficient for noisy real-world supply chain news, where causal links are often implied rather than explicitly stated.

Round 2 relaxed the Answer Grader by adding physical mechanisms and examples as mentioned in Section 3.3, such as fuel switching for change in carbon intensity of electricity grid. This improved recall by allowing the model to accept documents with physically plausible parameter links, even when the causal explanation was incomplete. However, this also increased false positives, showing a precision and recall trade-off. Round 3 then tightened the Document Grader by adding stricter conditions to the factual traceability criterion in Document Grader and clearer negative criteria (exclusion rules) to the Answer Grader. For example, early-stage announcements of potential sourcing shifts in the graphite value chain, or purely retrospective reviews of China's 2025 coal usage, are now strictly classified as negative cases, as they represent speculative goals or past events rather than ongoing, concrete physical actions. Similarly, news reporting a severe drought in the United States, when the target parameter is China's electricity grid, would be immediately rejected due to a clear geographic mismatch.

Round 4 introduced domain-specific refinement. The definition of source credibility was revised to prioritize official reports and third-party news sources, while excluding unsupported private analyst opinions. The definition of factual traceability criterion was expanded to include authoritative assessments of structural issues supported by past data. Additionally, common-sense domain rules were also added in Answer Grader, such as if the country faces the import limit of gas or oil as energy source, the coal power will be back up, which should be treated as trigger mechanism for change in electricity carbon intensity. The final simplified codes for Document Grader and Answer Grader are illustrated in Figure 13 and Figure 14, respectively. The complete codes are provided in Appendix E.

```python
from pydantic import BaseModel, Field
from langchain_core.prompts import ChatPromptTemplate

# 1. Schema: Enforcing structured, standardized outputs
class DocumentGrade(BaseModel):
    is_credible: bool = Field(description="True if from a valid third-party outlet, government domain, or corporate press release.")
    is_relevant: bool = Field(description="True if the text contains physical supply chain events altering target LCA parameters.")
    is_factual: bool = Field(description="True if the event physically occurred, is an ongoing macro-shift, or an empirical study. False if purely speculative.")
    rationale: str = Field(description="Brief explanation of the grading decision.")

# 2. Prompt: Setting the logical boundaries and rules
document_grader_prompt = ChatPromptTemplate.from_messages([
    ("system", """You are an "Intelligence Filter" for a supply chain monitoring pipeline. Evaluate incoming documents based on THREE strict criteria to prevent noise:

[CRITERIA 1: CREDIBILITY] *Added in Round 4
- PASS: Government domains, corporate press releases, or industry news portals.
- FAIL: Anonymous rants or purely subjective fiction.

[CRITERIA 2: RELEVANCE & BOUNDARY MATCH]
- PASS: Contains events disrupting the target LCA parameter: '{parameter}'.
- FAIL: Thematic mismatch (e.g., shipping vs. grid), geographic mismatch, or purely commercial noise.

[CRITERIA 3: FACTUALITY & ACTION STATE] *Added in Round 3 and Round 4
- PASS: Realized physical shocks, official enactments, or empirical scientific studies.
- FAIL: General speculations without physical action."""),
    ("human", "Content Snippet: {context}\nURL: {url}")
])

# 3. LangChain Pipeline: Connecting the logic, LLM, and schema
document_grader = document_grader_prompt | llm.with_structured_output(DocumentGrade)
```

Figure 13. The example code of Document Grader, consisting of the output schema, prompt, and LangChain workflow

```
answer_grader_prompt = ChatPromptTemplate.from_messages([
    ("system", """You are a Quality Assurance Evaluator for an Event Knowledge Graph
(EKG).Your ONLY task is to determine if the text contains SUFFICIENT CAUSAL EVIDENCE
to alter the target LCA parameter: '{parameter}'.

[CORE DEFINITION: PHYSICAL vs. COMMERCIAL]
To output TRUE, the event MUST trigger or be HIGHLY EXPECTED to trigger a "Physical
Structure Alteration" that STRICTLY MATCHES the Geography and Domain.
- EXCLUDE Purely Commercial Fluctuations: Do NOT output TRUE if the text ONLY
describes freight rates or price hikes WITHOUT physical grid/routing changes.

[TEMPORAL HORIZON & ACTION STATE]
- TRUE: Active/Ongoing Shocks or Authoritative Predictive Warnings
- FALSE: Early-Stage/Aspirational goals

[VALID MECHANISMS BY PARAMETER (Common-sense domain rule)] *Added in Round 2 and
Round 4
- Apply parameter-specific logic (e.g., Route Disruption for Transport; Fuel
Switching for Grid Carbon; Gas Import Limit/Gas-to-Coal)
- Apply logical assumptions ONLY IF geography matches (e.g., "The Grid Seesaw": if
clean energy drops, assume fossil fuels increase)

[RED FLAGS FOR IMMEDIATE FALSE (STRICT EXCLUSION)] *Added in Round 3
Output FALSE if the text contains:
1. Geographic or Thematic Mismatch
2. Analyst/Expert Speculation
3. Purely Downstream Commercial Consequences."""),
    ("human", "Documents: {context}")])
```

Figure 14. The example code of Answer Grader prompt

The final Round 4 results reached an F1-score of 72.64% ± 1.95%, with standard deviations below 2%. Recall increased to nearly 94%, while precision slightly decreased. This trade-off is acceptable for supply chain risk sensing because missing relevant early-warning signals is more costly than passing some additional candidates for later filtering. Since subsequent validation stages include additional evidence filtering, graph anchoring, boundary compliance checks, and LLM-as-a-Judge validation, the grader was intentionally tuned toward higher recall at this stage. Overall, the validation shows that the graders can serve as reliable document gatekeepers for Phase 0 and evidence filters for later scenario generation.

4.2 Historical EKG and Scenario Boundary Definition

4.2.1 Graph composition and evidence coverage

Phase 0 was not only a graph-building step. It provided the evidence base and structural foundation for Phase 1 scenario generation. During Phase 0, news searches covering the period from 2020 to 2025 were conducted for two predefined LCI parameters: carbon intensity of the electricity mix in China and graphite value chain transport distance and routing. The LLM-based query generator created search queries using parameter synonyms and related mechanism keywords. These queries were then used to retrieve documents through the Exa API as search engine.

After the search and grading process, 80 highly relevant documents were collected, such as such as reports on the Sichuan power crunch increasing coal reliance in 2022 and the Panama Canal drought disrupting shipping routes in 2023. More samples with URL links are provided in Appendix F. The grading step was important because not every retrieved document contained useful information for LCA parameter reasoning. The graders were evaluated with both positive and negative cases in Section 4.1. Through iterative prompt tuning, the number of successfully extracted parameter observations increased from 7 in the initial version to 85 in the final version. This improvement shows that the graders were able to filter out irrelevant or weak sources and identify documents that contained enough information to explain why and how the selected LCI parameters may change.

This result still indicates that the node resolution process did not simply reduce graph size; it improved the structural coherence of the EKG. The final EKG profile is illustrated in Table 4 and Figure 15.

Table 4. Final EKG profile after grading and node resolution

Node Type	Parameter	Actor	Location	Material or Product	Macro Trend	Micro Event	Parameter Observation
Number	2	129	133	43	125	243	85

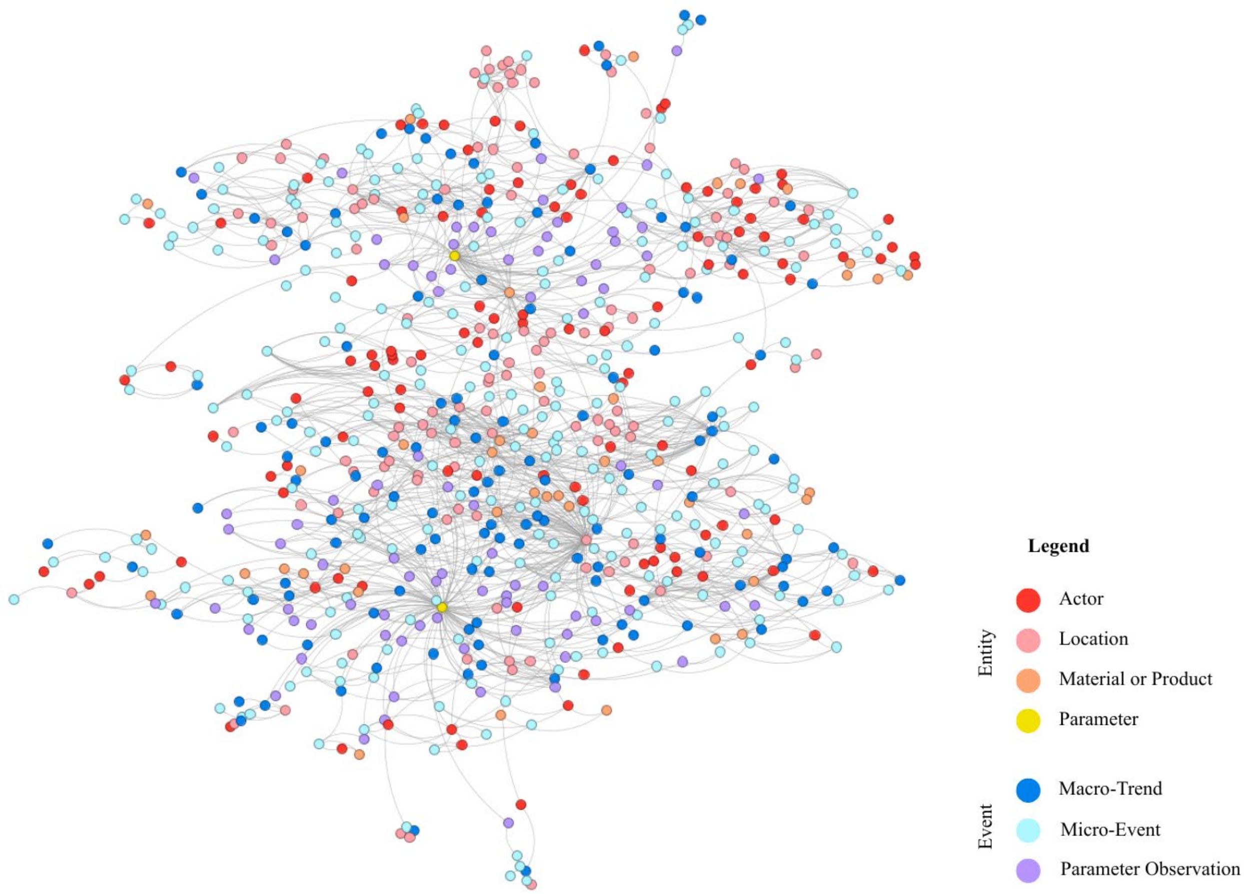


Figure 15. Global view of the historical EKG constructed in Phase 0

4.2.2 Parameter trajectories

This section illustrates how real-world events were captured and linked to parameters in the EKG. Two parameter trajectories are selected as examples: the first trajectory is about 2022 compound heatwave and drought in Southwest China, as shown in Figure 16. The EKG shows that the event disrupted hydropower generation, which could be substituted by coal-fired electricity. This substitution generated the observation of increased electricity carbon intensity and was anchored to the grid-carbon parameter.

The second trajectory is about Red Sea crisis, as shown in Figure 17. The EKG shows that the crisis disrupted the Suez Canal route and led to Cape of Good Hope rerouting. This route substitution generated the observation of higher transit time and emissions, which was then anchored to the transport-distance parameter.

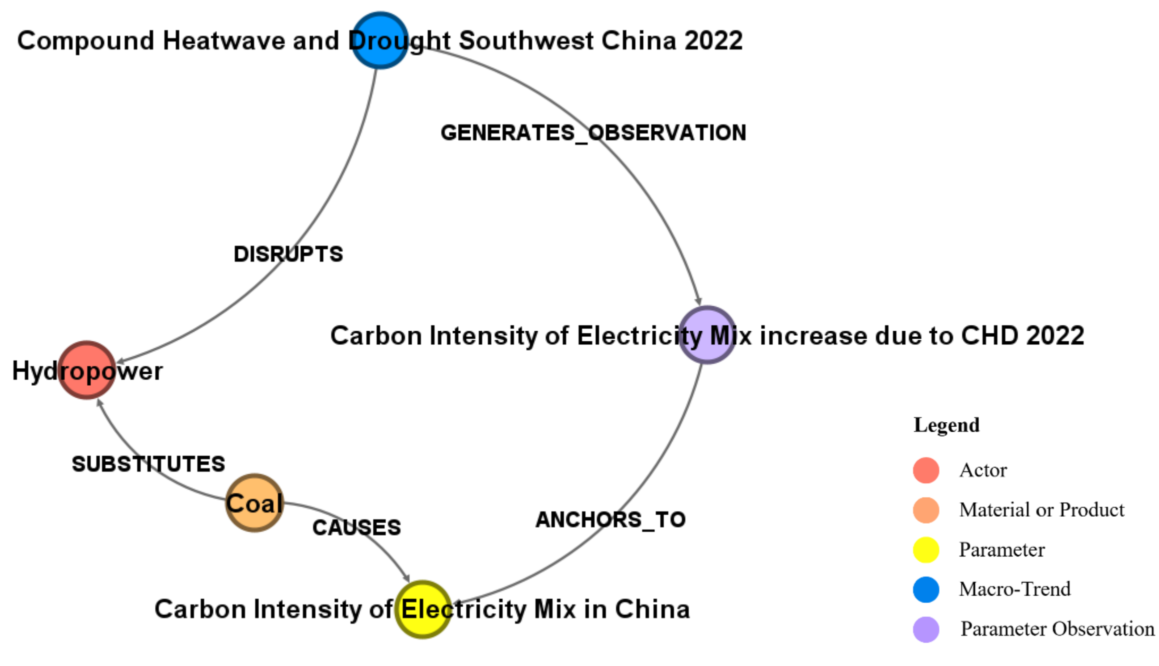


Figure 16. EKG trajectory linking the 2022 Southwest China compound heatwave and drought to grid carbon intensity

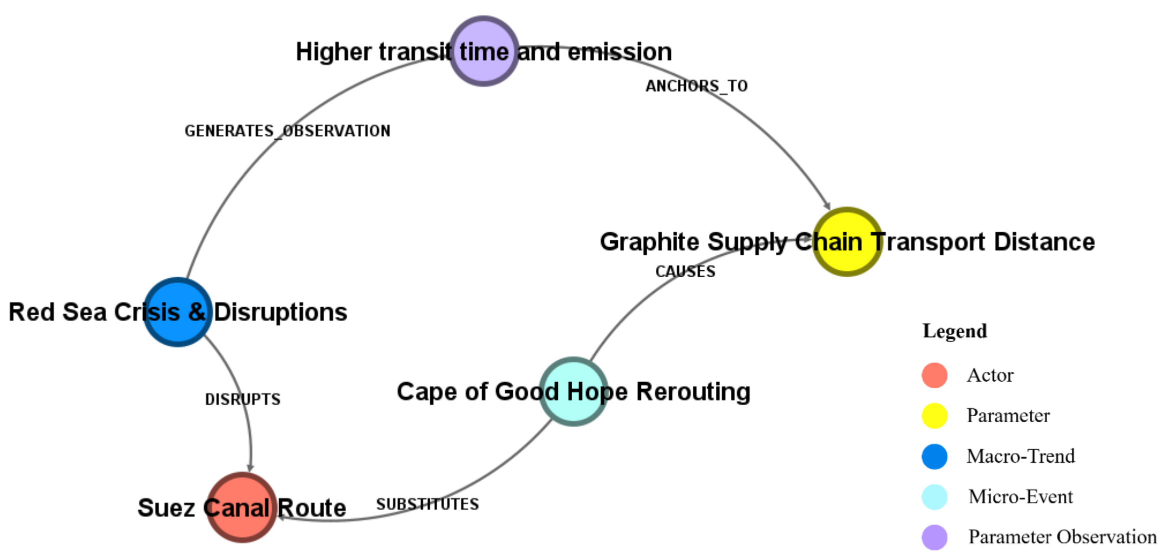


Figure 17. EKG trajectory linking the Red Sea crisis to graphite supply chain transport distance

These trajectories demonstrate how the EKG converts external events into structured LCI parameter-relevant pathways. This trajectory links an event to a disrupted system component, a substitution mechanism, an observation, and finally an affected LCI parameter. The EKG also makes it possible to inspect the potential impact radius of an event, meaning whether one event affects only one parameter or propagates across multiple parameters. In the current cases, no strong cross-parameter impact radius was observed.

4.2.3 Strategic nodes: System Hubs and Chokepoints

As illustrated in Figure 18, after filtering out transient micro-events and target parameters, a notable topological divergence emerges system hubs and chokepoints. They capture two backgrounds: the energy and electricity system, and the transport system in real-world.

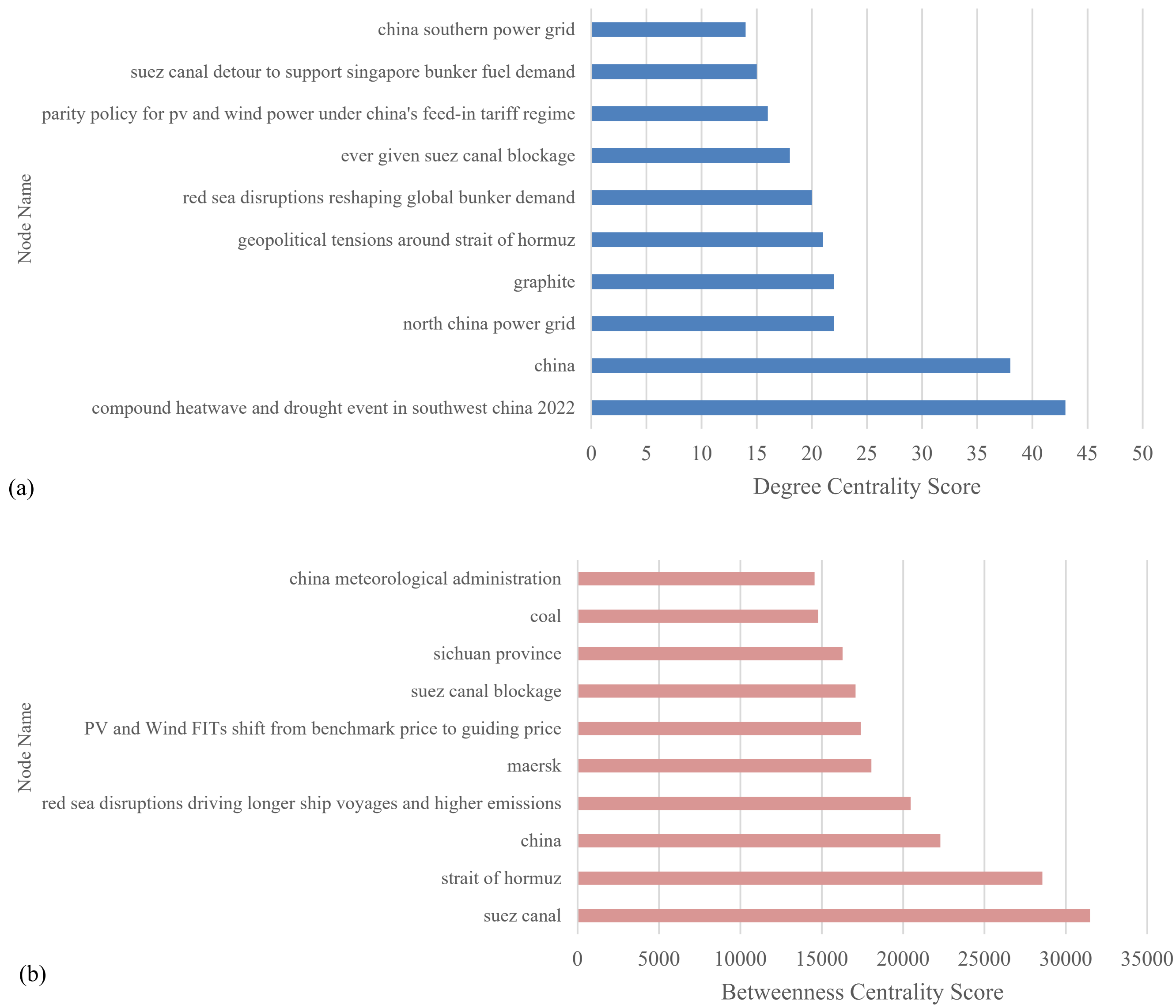


Figure 18. Quantitative ranking of the top 10 EKG nodes based on centrality. (a) Top 10 system hubs identified by degree centrality, dominated by predefined LCA parameters and macro-energy

drivers. (b) Top 10 chokepoints identified by betweenness centrality, highlighting severe geographic and maritime bottlenecks.

The system hub results show that the EKG captures two main groups of external drivers. The first group relates to China's energy system and climate stress, including compound heatwave and drought events, national and regional power grid actors, and renewable electricity policy shifts. The second group relates to maritime transport and geopolitical disruption, including the Strait of Hormuz, and Ever-Given Suez Canal blockage. These nodes are mainly relevant for graphite transport distance and routing by being associated with longer transport distances and fuel demand increase.

The chokepoint results further show how external shocks are transmitted to LCI parameters. Maritime corridor nodes, such as the Suez Canal, Red Sea disruptions, and Maersk, act as bridges between geopolitical or climate-related disruptions and change in transport parameters. In contrast, energy-system nodes, such as China, Sichuan Province, and coal, connect weather extremes, hydropower availability, coal dispatch, and electricity carbon intensity. Policy-related nodes, such as the shift from feed-in tariffs to guiding prices for PV and wind power, indicate that market design can also mediate the relationship between renewable expansion and grid carbon intensity.

4.2.4 Scenario boundary matrix definition

Table 5 and Table 6 detail the four-level scenario boundary matrices for the two selected LCI parameters, which are subsequently combined into a single integrated matrix for use in Phase 1.

For grid carbon intensity, changes in electricity carbon intensity are usually much smaller than route-distance changes, around 1.0-1.5% in China annually (IEA, 2026), which set as the mild scenario boundary. Although short-term weather and dispatch effects may be larger, such as the 21.64% variation reported by Zhang et al. (2025), this study does not model hourly electricity-mix fluctuations. Instead, it focuses on average changes that are suitable for simplified LCA scenario integration. For this reason, the main scenario boundary matrix uses a conservative extreme range of 0.10 for grid carbon intensity. The 0.20 value is retained only as an optional sensitivity case for future high-resolution dynamic LCA applications.

For transport-distance boundaries, UNCTAD (2024b) and Notteboom et al. (2024) report that Suez Canal and Red Sea-related disruptions led to rerouting via the Cape of Good Hope, increasing transport distance and transit time by approximately 30% or more. Additional EKG evidence suggests that Cape routing can add around 6,000 miles and increase voyage distance by more than 40%. Accordingly, the extreme scenario boundary is set at 0.45.

Table 5. Scenario boundary matrix for carbon intensity of electricity mix in China (P1)

Severity level	**Severity interpretation**	**Mechanism**	**Representative EKG evidence**	**Δ P1**
Level 1: Mild	Localized or short-term adjustment	Minor renewable integration or local grid balancing	Local wind/solar integration; short-term low-carbon supply improvement	[0.005 ,0.015]
Level 2: Moderate	Seasonal or regional fluctuation	Seasonal coal-share fluctuation; moderate fossil backfills	Coal-share fluctuation from approximately 5 0%; regional drought pressure	[0.015, 0.030]
Level 3: Severe	Sustained regional energy stress	Hydropower shortage and coal backfill	Sichuan or Southwest China heatwave and drought; hydropower shortage	[0.030, 0.060]
Level 4: Extreme	Structural or system-level shift	Large-scale fossil capacity expansion or compound climate-energy stress	106 GW coal project approval; prolonged compound heatwave and drought	[0.060, 0.100]

Table 6. Scenario boundary matrix for graphite value chain transport distance and routing (P2)

Severity level	**Severity interpretation**	**Mechanism**	**Representative EKG evidence**	**Δ P2**
Level 1: Mild	Localized operational disturbance	Port congestion or minor route adjustment	General port congestion; short-term operational delays	[0.01, 0.05]
Level 2: Moderate	Regional infrastructure constraint	Canal draft limit or partial rerouting	Panama Canal drought restrictions; capacity constraints	[0.05, 0.15]
Level 3: Severe	Major corridor disruption	Major canal blockage or	Suez Canal blockage; prolonged canal restrictions	[0.15,0.30]

		prolonged route disruption		
Level 4: Extreme	Full route substitution	Full maritime artery avoidance	Red Sea/Suez rerouting via Cape of Good Hope; over 40% voyage-distance increase	[0.30, 0.45]

4.3 Scenario Generation, Stage 2: Plausibility Validation, and Scenario-Based Climate-Change Impact Estimates

In Phase 1, 36 newly monitored news documents in 2026 were collected based on the 20 strategic nodes identified in Phase 0. Since the current Trend Monitoring agent focuses on predefined monitoring targets rather than unrestricted global parameter-level searching, the 26 positive cases from the golden dataset were also included as representative parameter-level event cases. In future applications, the monitoring interval could be shortened to support higher-frequency trend detection and more timely scenario updates.

After extracting the mini-EKG from them and updating the graph, the EKG contained 131 updated event nodes, while 52 events were identified as relevant to the target LCI parameters and were passed to the Scenario Generation agent for numerical inference.

Stage 2 validation evaluated whether the expanded EKG could support plausible scenario generation and bounded parameter inference. The validation focused on whether the generated scenarios followed the predefined reasoning steps, selected an appropriate mechanism and severity level, and remained within the constraints of the scenario boundary matrix.

After prompt refinement and schema strengthening, the validated scenario rate increased from approximately 40% in the initial experiment to 78.8% in the final validation. The improved schema required the LLM to first identify the mechanism and then select the corresponding scenario severity level, which improved the consistency among evidence extraction, mechanism mapping, severity selection, and numerical inference.

As shown in Table 7, two representative scenarios with high plausibility validation scores were selected to illustrate the conversion of EKG-derived events into LCI parameter changes. The first scenario links Houthi attacks and Cape of Good Hope rerouting to increases in graphite supply chain transport distance and routing, with inferred changes of +30.0%, +37.5%, and +45.0% under the best, average, and worst cases. The second scenario links increased thermal generation and reduced renewable output in China to higher electricity carbon intensity, with inferred changes of

+3.0%, +4.5%, and +6.0%. This shows Agentic AI can monitor the relevant events and translate them into bounded parameter changes.

Table 7. Representative EKG-derived scenarios for LCI parameter

Trigger Event	**Target Parameter**	**Historical / evidence cases in EKG**	**Best Case**	**Avg Case**	**Worst Case**	**Strategic Rationale**
Houthi attacks forcing Cape route rerouting	Graphite supply chain transport distance & routing	Red Sea / Suez disruption; Cape of Good Hope rerouting; reported longer sailing distances	+30.0%	+37.5%	+45.0%	Maritime security risks force vessels to avoid the Red Sea/Suez corridor and reroute via the Cape, increasing route length, transit time.
Q1 2026 thermal generation up; renewables down	Carbon intensity of electricity mix in China	Thermal generation rebound; lower renewable output; fossil backfill under grid stress	+3.0%	+4.5%	+6.0%	Higher thermal generation and lower renewable output increase the average carbon intensity of electricity.

The selected validated scenarios were further passed to Phase 2. This phase does not aim to reproduce a complete graphite LCA model or to redefine the full system boundary. Instead, it uses a simplified fixed reference calculation to demonstrate how parameter changes can be translated into climate-change impact estimates. The calculation includes only the parameter positions affected by the selected scenarios, namely electricity-related emissions and transport-related emissions. Other life-cycle processes are held constant and are not re-estimated in this phase.

This simplified calculation in Equation 4 and 5 demonstrates how external-event evidence identified by the agentic AI framework can be translated into bounded parameter changes and further connected to climate-change impact estimates. Climate-change impact is expressed as global warming potential over a 100-year time horizon (GWP100), based on IPCC 2021 characterization factors.

$$\Delta GWP_{Total} = GWP_{electricity,base} \times \Delta P1 + GWP_{transport,base} \times \Delta P2 \quad \text{(Equation 4)}$$

$$GWP_{electricity,base} = \alpha \times I_c \quad \text{(Equation 5)}$$

ΔGWP_{Total} represents the scenario-induced change in climate-change impact relative to the baseline, expressed in kg CO_2e per kg of graphite. It is calculated as the sum of the electricity-related change and the transport-related change. The baseline impact is presented in Table 8.

$\Delta P1$ and $\Delta P2$ represent the relative scenario change rates for electricity carbon intensity and transport distance, respectively. Figure 19 reports the estimated changes in GWP100 caused by the two representative EKG-derived scenarios. The values represent additional climate-change impacts relative to the baseline, rather than total product-level GWP100 values.

Table 8. Baseline Impact of electricity consumption and transport in graphite value chain

Baseline climate-change impact in GWP100	$\boldsymbol{GWP_{electricity,base}}$ (kg CO_2e/kg graphite)			$\boldsymbol{GWP_{transport,base}}$ (kg CO_2e/kg graphite)
	$\boldsymbol{\alpha}$ Specific electricity consumption (kWh/kg graphite)	$\boldsymbol{I_c}$ Carbon intensity of electricity in China (CO_2e/ kWh)	Result	
Natural graphite	8.7	0.532	4.628	0.015-0.45
Synthetic graphite	36		19.152	0.3-0.35
Data source	(Bhattacharyya et al., 2026)	(IEA, 2026)		(Surovtseva et al., 2022)

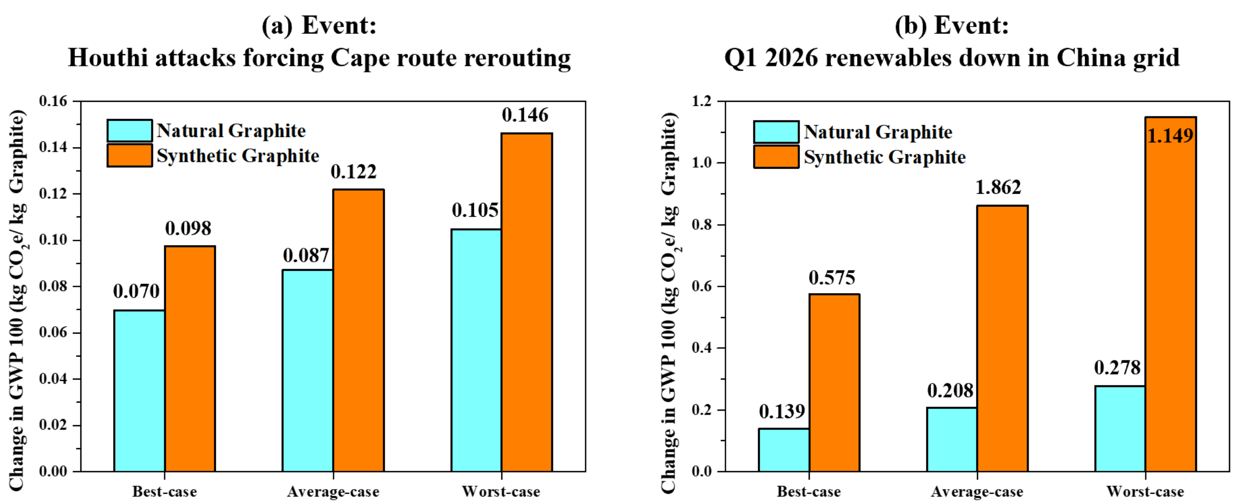


Figure 19. Estimated scenario-induced changes in GWP100 relative to the baseline triggered by 2026 events: (a) Houthi attacks and (b) renewables down in China

The results show that the magnitude of the final GWP100 change depends not only on the relative scenario change rate, but also on the baseline contribution of the affected LCI parameter. The Houthi-related rerouting scenario generates a large relative increase in transport distance, ranging from 30.0% to 45.0%. However, the resulting GWP100 change remains comparatively small because transport accounts for a limited share of the baseline climate-change impact. Consequently, the estimated GWP100 increases for natural graphite are discrete values of 0.070 kg CO2e/kg (best case), 0.087 kg CO2e/kg (average case), and 0.105 kg CO2e/kg (worst case). For synthetic graphite, the corresponding estimated changes are 0.098, 0.122, and 0.146 kg CO2e/kg across the three respective scenarios.

In contrast, the electricity carbon-intensity scenario produces larger changes, particularly for synthetic graphite. This is because synthetic graphite production has a substantially higher electricity consumption coefficient than natural graphite. Under the Q1 2026 thermal-generation increase and renewable-output decrease scenario, the estimated GWP100 changes for natural graphite are 0.139 kg CO2e/kg in the best case and 0.278 kg CO2e/kg in the worst case. For synthetic graphite, the corresponding changes are 0.575 kg CO2e/kg and 1.149 kg CO2e/kg, respectively. This result demonstrates that even moderate relative changes in carbon intensity of the electricity can generate significant impacts when the affected process is electricity intensive.

5. Discussion

This chapter discusses the broader implications of the proposed framework. It first examines the main limitations related to EKG completeness, data availability, dynamic LCA standardization, and system engineering. It then outlines future research directions for improving data integration, graph structure, and parameter monitoring. Finally, it clarifies the role of HITL governance and positions the framework as a supportive outside-in sensing layer rather than an autonomous LCA decision-making system.

5.1 Limitations

While this study demonstrates the potential of Agentic AI and EKGs for supporting supply-chain risk monitoring and dynamic LCA, several limitations remain.

1. EKG Constraints in Completeness and Black Swan Events

This study follows the standard event acquisition pipeline proposed by Guan et al. (2022), but event argument completion is excluded from the thesis scope. Therefore, the resulting EKG should not be interpreted as a complete causal representation of the graphite supply chain. Instead, it serves as an evidence-informed reference structure that links newly monitored events to historical events, mechanisms, and parameter-relevant nodes. Its purpose is to support plausible event-to-parameter reasoning, rather than to fully reconstruct historical causality or predict unprecedented disruptions. Furthermore, this study explicitly excludes “black swan” events from its scope.

Defined as exceptionally rare and unpredictable macro-disruptions with massive consequences (Taleb, 2007), black swans in CRM value chains represent completely unprecedented shocks, such as a sudden global trade halt. Because these events share no similarity with past structural patterns, they create a 'zero-shot' scenario where the EKG cannot reliably infer bounded parameter changes.

A further limitation concerns the quality of edge extraction and node resolution. These were among the most challenging steps in constructing the EKG. Although the extraction and node-resolution agents were iteratively refined through repeated prompt adjustments and code-level modifications, the final graph may still contain some residual errors, such as fragmented event chains, duplicated entities, and missing edges. As a result, the graph does not fully conform to the intended schema in all cases. This may reduce the clarity of event-to-parameter pathways and limit the extent to which downstream scenario generation can rely on the EKG as a complete evidence structure.

Finally, there is a trade-off between graph density and interpretability. The calculation and threshold for graph density are provided in Appendix E. While sparse connections limit graph coverage and can make the EKG appear as disconnected event sequences rather than a coherent network, increasing graph density can make the graph difficult to interpret visually. In early experiments, reconnecting orphan nodes was considered to reduce sparsity, but this strategy risked producing hallucinated connections and undermine interpretation. Therefore, this study did not apply aggressive orphan-node rescue or graph densification. In short, for LCA-oriented scenario generation, causal traceability, physical plausibility, and interpretability should be prioritized over network density and breadth.

2. Lack of Expert Panel and Subjective bias

A second critical limitation of this study is its reliance on a single researcher's judgment, which introduces subjective bias. Because the current framework is primarily designed as a foundational proof of concept, the core methodological decisions, including the LCA scoping, the numerical thresholds defined within the Scenario Boundary Matrix, and the human-labeled ground truth used for validation, have not yet undergone rigorous review by an independent expert panel. Consequently, the assigned parameter ranges and decision rules currently represent a subjective baseline rather than an objectively consensus-driven standard. Furthermore, while employing an "LLM-as-a-Judge" for scenario validation offers high operational efficiency, it cannot fully substitute for the critical scientific verification that human domain experts provide.

3. LCA Application Gap

Since the only data resource of the current framework is public news, it limits the LCI parameter selection. Initial experiments attempted to monitor six graphite-related LCA parameters, including specific heat and ore grade. However, these parameters were not well suited to monitoring through public news sources, because news articles rarely provide detailed engineering or process-level information that directly affects environmental impacts per functional unit. This reflects a broader mismatch between supply-chain disruption data and LCA data requirements. General news sources often report on mine capacity expansion, demand changes, or price fluctuations.

In addition, the Pa-LCA and DLCA approaches still lack fully mature methodological standardization and decision-support tools (Arias et al., 2025), let alone the integration with external AI sensor. Therefore, even when external information is retrieved, there remains limited consensus on how to align such information with international LCA standards, temporal system boundaries, dynamic inventories, and time-dependent characterization methods (Beloin-Saint-Pierre et al., 2020). Future work should therefore explore how event-based parameter values generated by the system can be integrated with parametrized LCA, dynamic inventory structures, and evolving system-boundary assumptions.

4. System Stability, and Environmental Costs

At the system architecture level, the proposed Agentic AI workflow introduces computational robustness concerns that are relevant for its practical use as an LCA decision-support tool. Unlike a static LCA model, the workflow depends on multiple sequential agent calls, external API connections, and intermediate file exchanges steps. Therefore, long-running executions remain vulnerable to API instability and network interruptions. Without appropriate safeguards, a temporary failure in one step could force the entire workflow to be restarted, reducing reproducibility and increasing computational cost. Finally, the environmental footprint, specifically the energy and carbon costs associated with continuous Agentic AI operations, remains unassessed.

5.2 Recommendations for Future Research

Future research should address the limitations identified in this study. First, future work should develop a clearer, three-layered ontology to organize the EKG. This ontology would systematically separate the network into three levels: Layer 1 (L1) for supply-chain topology (e.g., companies involved in CRM value chain, unit processes and their chemicals and resource uses, mining locations); Layer 2 (L2) for LCI parameters (e.g., carbon intensity of electricity); and Layer 3 (L3) for macro-events (e.g., geopolitical disruptions and climate extremes). While the current framework primarily connects L3 to L2, integrating the L1 would allow the system to trace exactly how external shocks physically alter material flows and sourcing routes. This layered approach would reduce irrelevant connections and make the AI's reasoning much easier to interpret.

Second, the system should overcome its reliance on public news through multi-source data integration. Because public news rarely reports on highly technical variables, future frameworks should incorporate specialized industry reports, market databases, and CRM-specific platforms. Broadening these data inputs will provide more precise, engineering-level evidence to support LCA-relevant scenario generation.

Lastly, it is highly recommended to establish an expert panel comprising specialists and stakeholders in LCA and CRM value chains to perform revalidation or recalibration. With diverse domain expertise, an expert panel can define more objective parameter boundaries, establish reliable ground-truth datasets for testing, and perform rigorous cross-validation. Strengthening this HITL oversight is essential to transition the framework from a conceptual prototype into a reliable decision-support tool for real-world industries.

6. Conclusions

This research serves as a methodological exploration to address the static, retrospective limitations of conventional LCA in highly volatile CRM supply chains. Translating qualitative news into quantitative LCA inputs presents challenges. First, public news rarely reports direct LCA parameters or exact numerical changes, making it difficult to capture external events that genuinely impact LCA models. Second, when forced to generate scenario values without historical references, LLMs frequently fabricate or hallucinate these missing numbers. To overcome these barriers, this study developed an EKG-driven Agentic AI framework that functions as an innovative outside-in sensing layer.

Rather than relying on unconstrained AI generation, the framework incorporates extensive human oversight and decision-making. This study uses predefined data templates, task-specific agents with a clear division, and an expert-defined scenario boundary matrix to strictly control numerical parameter generation based on historical knowledge base. Demonstrated through a graphite value chain case study, the framework translated complex qualitative events, such as maritime rerouting due to Houthi attacks and fluctuations in China's electricity mix, into bounded, physically plausible parameter changes to estimate their change in climate-change impacts. While ongoing challenges remain regarding data availability and graph completeness, this study provides a novel architectural blueprint for proactive scenario exploration.

Appendixes

Appendix A: The working concept in LangChain and LangGraph

LangChain is used to build the individual LLM applications used by each agent. In this study, an LLM application refers to a task-specific unit that combines a prompt, a language model, and a predefined output schema. The prompt tells the LLM what role to perform, the language model performs the text-based reasoning, and the schema defines how the output should be structured. This allows the framework to use the reasoning ability of the LLM while still producing outputs that can be processed by later steps. The language model used in this study is OpenAI's GPT-5 Nano model, accessed through the API. The example code for LLM application via LangChain in this study is shown in Figure 13.

The predefined output schemas are implemented using Pydantic, which is a Python library that allows users to create a model class and description for an object and can serve as the data validation method (Narayanan, 2024). In this study, Pydantic schemas function as data templates for LLM outputs. Instead of allowing the LLM to return only free-form text, the schema requires the output to follow a predefined format. This makes the output more consistent, easier to check, and easier to pass from one agent to another.

Among all LLM-supported applications, the document graders and the scenario generator are especially important. The graders determine which documents enter the EKG and scenario reasoning process, while the Scenario Generation agent translates monitored events into bounded LCI parameter-change scenarios. Because these two components directly affect downstream results, they are selected for validation in this study, as described in Section 3.7.

On the other hand, LangGraph is used to connect these individual LLM applications into a multi-step workflow. Each processing agent is represented as a node (see Figure 20), and the connections between node are represented as edges or conditional edges (see Figure 21). A central feature of LangGraph is the use of a shared state (see Figure 22). In this study, the state functions as structured memory that records the information needed for agents to communicate and continue the workflow. Like the output schemas used in LangChain, the state is predefined as a data template. It can store direct outputs and references to intermediate files, such as file paths where retrieved documents, extraction results, or reports are saved.

```
import os
import json
import time
```

```python
import uuid
import glob
from graph.state import Phase0State
from graph.chains.final_node_extractor import node_extractor
from graph.configuration import BASE_OUTPUT_DIR_PHASE0
from langchain_text_splitters import RecursiveCharacterTextSplitter

text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=10000,     # Set to 10000 to prevent exceeding single-request token limits
    chunk_overlap=1000,   # 1000-character overlap to preserve contextual causal relationships
    separators=["\n\n", "\n", "。", ".", " ", ""] # Ensures sentences are not cut off mid-thought
)

def node_extraction(state: Phase0State):
    print("\n" + "="*55)
    print("🧠 [Phase 0 - Node 2: Node Extraction] Initiating Parameter-Driven Intelligence Extraction Engine")
    print("="*55)

    output_dir = os.path.join(BASE_OUTPUT_DIR_PHASE0,"extracted/nodes") #"./output/phase0/extracted_V18/nodes"
    os.makedirs(output_dir, exist_ok=True)
    extracted_nodes_paths = state.get("extracted_nodes_paths", [])

    existing_files = glob.glob(os.path.join(output_dir, "*.json"))

    if existing_files:
        print(f"  ⏩ [Cache Hit] Found {len(existing_files)} existing 'Node Extraction' files!")
        print("  ⏩ [Skipping Extraction] To save LLM token costs, reading the existing files directly into the next phase.")
        for file_path in existing_files:
            normalized_path = os.path.normpath(file_path)
            if normalized_path not in extracted_nodes_paths:
                extracted_nodes_paths.append(normalized_path)
        return {
            "extracted_nodes_paths": extracted_nodes_paths,
            "status": "success"
        }

    corpus_paths = state.get("corpus_file_paths", [])
    if not corpus_paths:
```

```
        print("❌ [Error] No document paths received in State.")
        return {"status": "error"}

    print(f"📁 Preparing to process {len(corpus_paths)} knowledge files...")

    for file_path in corpus_paths:
        original_filename = os.path.basename(file_path)
        out_filename = f"nodes_{original_filename}"
        out_filepath = os.path.join(output_dir, out_filename)

        if os.path.exists(out_filepath):
            print(f"  ⏩ [Cache Hit] Already extracted, skipping:
{original_filename}")
            if out_filepath not in extracted_nodes_paths:
                extracted_nodes_paths.append(out_filepath)
            continue

        print(f"\n🔍 Extracting from: {original_filename}")

        try:
            with open(file_path, 'r', encoding='utf-8') as f:
                docs = json.load(f)

            if not docs: continue

            metadata = docs[0].get("metadata", {})
            seed_parameter = metadata.get("seed_parameter", "Unknown_Parameter")
            product = metadata.get("product", "Unknown_Product")
            crm = metadata.get("crm", "Unknown_CRM")

            all_extracted_entities = []
            all_extracted_events = []
            all_extracted_edges = []

            # The global record here can be retained as metadata at the file level
            global_source_url = [d.get("url") for d in docs if "url" in d]
            global_source_title = [d.get("title") for d in docs if "title" in d]

            for doc_index, doc in enumerate(docs):
                article_text = doc.get("content", "")
                # 🚨 Accurately extract the title and URL of "the current article"
here.
                current_title = doc.get("title", "Unknown Title")
                current_url = doc.get("url", "Unknown URL")
```

```
                if not article_text: continue

                print(f"  📖 Reading article {doc_index+1}/{len(docs)}:
《{current_title[:30]}...》")

                if len(article_text) > 50000:
                    article_chunks = text_splitter.split_text(article_text)
                else:
                    article_chunks = [article_text]

                for chunk_idx, chunk in enumerate(article_chunks):
                    try:
                        response = node_extractor.invoke({
                            "seed_parameter": seed_parameter,
                            "crm": crm,
                            "product": product,
                            "article_text": chunk,
                            "source_node_name": "None",
                        })

                        # ==========================================
                        # 🚨 Imprint the Title and URL onto each node
                        # ==========================================
                        if hasattr(response, 'entities') and response.entities:
                            for entity in response.entities:
                                ent_dict = entity.dict() if hasattr(entity, 'dict')
else entity
                                ent_dict["node_id"] = f"ent_{uuid.uuid4().hex[:6]}"
                                # Force write to the source array (using a List is to
conform to the format of subsequent deduplication and merging)
                                ent_dict["source_title"] = [current_title]
                                ent_dict["source_url"] = [current_url]
                                all_extracted_entities.append(ent_dict)

                        if hasattr(response, 'events') and response.events:
                            for event in response.events:
                                evt_dict = event.dict() if hasattr(event, 'dict')
else event
                                evt_dict["node_id"] = f"evt_{uuid.uuid4().hex[:6]}"
                                evt_dict["source_title"] = [current_title]
                                evt_dict["source_url"] = [current_url]
                                all_extracted_events.append(evt_dict)

                        if hasattr(response, 'edges') and response.edges:
                            for edge in response.edges:
```

```python
                                edge_dict = edge.dict() if hasattr(edge, 'dict') else
edge
                                all_extracted_edges.append(edge_dict)

                        # 💡 protect API
                        time.sleep(1)

                    except Exception as e:
                        print(f"      ❌ Node extraction failed for this chunk:
{e}")

                print(f"      [Immediate Report] The current documents are extracted
with {len(all_extracted_entities)} entities, {len(all_extracted_events)} events in
total")

            entity_count = len(all_extracted_entities)
            event_count = len(all_extracted_events)
            edge_count = len(all_extracted_edges)

            if (entity_count + event_count) > 0:
                print(f"  ✅ Successfully extracted {entity_count} entities,
{event_count} events!")

                extraction_result = {
                    "metadata": {
                        "source_file": original_filename,
                        "seed_parameter": seed_parameter,
                        "product": product,
                        "crm": crm,
                        "source_title": global_source_title,
                        "source_url": global_source_url,
                    },
                    "entities": all_extracted_entities,
                    "events": all_extracted_events,
                    "edges": all_extracted_edges
                }

                with open(out_filepath, 'w', encoding='utf-8') as f:
                    json.dump(extraction_result, f, indent=4, ensure_ascii=False)

                if out_filepath not in extracted_nodes_paths:
                    extracted_nodes_paths.append(out_filepath)
            else:
                print("  ⚠️ Warning: No entities or events were extracted.")
```

```
        except Exception as e:
            print(f"  ❌ Error occurred while processing {original_filename}: {e}")

    return {"extracted_nodes_paths": extracted_nodes_paths, "status": "success"}
```

Figure 20. Example code of an agent/node in the framework
(using the Node Extraction agent as an example)

```
from dotenv import load_dotenv
load_dotenv()
from graph.nodes import search, node_extraction, edge_extraction, node_resolution,
export_to_neo4j, network_criticality_analysis,  global_embedding
from graph.consts import SEARCH, NODE_EXTRACTION, EDGE_EXTRACTION, NODE_RESOLUTION,
EXPORT_TO_NEO4J, NETWORK_CRITICALITY_ANALYSIS, GLOBAL_EMBEDDING
from langgraph.graph import START, END, StateGraph
from graph.state import Phase0State

def route_after_process(state: Phase0State):
    status = state.get("status")
    if status == "success":
        return "next step"
    else:
        return "fail"


workflow = StateGraph(Phase0State)

# Knowledge extraction
workflow.add_node(SEARCH, search)
workflow.add_node(NODE_EXTRACTION, node_extraction)
workflow.add_node(EDGE_EXTRACTION, edge_extraction)
workflow.add_node(NODE_RESOLUTION, node_resolution)
workflow.add_node(EXPORT_TO_NEO4J, export_to_neo4j)
workflow.add_node(NETWORK_CRITICALITY_ANALYSIS, network_criticality_analysis)
workflow.add_node(GLOBAL_EMBEDDING, global_embedding)

workflow.add_edge(START, SEARCH)

workflow.add_conditional_edges(
    SEARCH,
    route_after_process,
    {
        "next step": NODE_EXTRACTION,
        "fail": END
    }
)
```

```
workflow.add_conditional_edges(
    NODE_EXTRACTION,
    route_after_process,
    {
        "next step": EDGE_EXTRACTION,
        "fail": END
    }
)

workflow.add_conditional_edges(
    EDGE_EXTRACTION,
    route_after_process,
    {
    "next step": EXPORT_TO_NEO4J,
       "fail": END
    }
)
workflow.add_conditional_edges(
    EXPORT_TO_NEO4J,
    route_after_process,
    {
    "next step": NODE_RESOLUTION,
       "fail": END
    }
)

workflow.add_conditional_edges(
    NODE_RESOLUTION,
    route_after_process,
    {
    "next step": NETWORK_CRITICALITY_ANALYSIS,
       "fail": END
    }
)

workflow.add_conditional_edges(
    NETWORK_CRITICALITY_ANALYSIS,
    route_after_process,
    {
    "next step": GLOBAL_EMBEDDING,
       "fail": END
    }
)

workflow.add_edge(NODE_RESOLUTION, END)
```

```
phase_0_app = workflow.compile()
```

Figure 21. Example code of the multiple agents/nodes in a phase (using the Phase 0 as an example)

```
class Phase0State(TypedDict):
    status: str
    error_message: Optional[str]
    seen_urls: List[str]
    corpus_file_paths: List[str]
    indicator_file_paths: List[str]

    extracted_nodes_paths: List[str]
    master_graph_path: str
    final_graph_path: str

    embedding_graph_path: str
    GraphSAGE_weights_path: str

    criticality_metrics_path: str
    final_insight_report_path: str
    neo4j_driver: Any
    GraphSAGE_model:Any

    original_density: float
    original_node_count: int
    original_edge_count: int

    final_density: float
    final_node_count: int
    final_edge_count: int
```

Figure 22. Example code of the shared state in a phase (using the Phase 0 as an example)

This design allows each agent to focus on a specific task and pass its results to the next agent in a traceable way. For example, one agent may save retrieved documents to a designated file location, while the shared state records the file path as a string. The next agent can then read this path from the state and access the saved documents for the next action. As a result, intermediate outputs can be stored, inspected, and reused, rather than hiding within a single LLM interaction. This improves accountability because each agent has a defined task, recorded input, and inspectable output.

Although LangChain and LangGraph provide the basic infrastructure, they do not define the methodological logic of the framework. The developer must still design task decomposition, agent roles, prompts, schemas, routing rules, tools, and traceability controls. Therefore, by defining task-specific LLM applications, structured output templates, shared workflow states, and conditional routing rules, this study adapts generic LLM orchestration tools into a specialized system for monitoring external events and generating bounded LCA-relevant scenarios in CRM value chains.

Appendix B: The design details of agents in Phase 0

Search Agent

This agent is designed to collect historical documents related to the triggers of pre-defined parameters because building an EKG requires a solid foundation of real-world evidence. So, the agent used an LLM to generate expansive search queries, capturing various synonyms for parameter and trigger mechanisms, and executed them via the Exa API to obtain a broad collection of external news and reports.

However, raw search results often contain noise or irrelevant information. The agent needs to strictly filter these retrieved documents because only credible, relevant, and sufficiently detailed texts can serve as reliable building blocks for the graph. Thus, two specialized LLM-driven Graders, inspired by advanced information retrieval concepts like Self-RAG (Asai et al., 2024) and Corrective RAG (CRAG) (Yan et al., 2024), are used as quality-control gatekeepers. Specifically, the document grader evaluates the text to ensure source reliability, relevance, and traceability to parameter variations. Following this, the answer grader evaluates whether the content provides enough how and why context regarding the trigger. After graders, this agent can produce documents with higher relevant and quality to the next agent, which significantly minimizes the generation of irrelevant or false nodes in the next agent.

Nodes and Edges Extraction Agents

The two agents are designed to extract specific entities, events (nodes), and their logical connections (edges) from the filtered documents because the unstructured text must be translated into a structured network to assemble the EKG. Therefore, they utilize the predefined schema-based extraction method (Guan et al., 2022). The predefined schemas give the LLM strict templates on what to look for, capturing the defined nodes and their attributes, effectively mitigating hallucinations.

EKG organizes the 5W1H information of each event: who was involved, where it occurred, what material or parameter was affected, when it occurred, why it happened, and how it influenced the selected LCI parameter. Therefore, nodes are classified into two broad categories: Entities

(Parameter, Actor, Location, Material/Product) and Events (Macro-Trend, Micro-Event, Parameter Observation). Entities possess a location scope, whereas Events include temporal attributes, event types, observation values, and sentiment scores reflecting parameter changes. Among them, Parameter Observation is the most important node type since it presents how the LCI parameters change in sentiment aspect, and even relative or absolute number, providing foundation for the scenario boundary matrix. Based on the predefined attribute schemas detailed in Table 9, the LLM can extract the corresponding information and outputs structured JSON file.

Table 9. Node types and attributes in the EKG schema

Node	**Entity**	**Event**
Node Basic attribute	Node id, node name, node category, node description	
Node Type	Parameter, Actor, Location, Material or Product	Macro Trend, Micro Event, Parameter Observation
Node Attributes	Location scope	Event type, start date, complete date, observation value, observation sentiment

Edges are classified into entity-entity, entity-event, and event-event three categories, with semantic types strictly dictated by the connected nodes (e.g., Actor LOCATED_IN Location, or Macro-Trend DISRUPTS Material/Product).

Following Hu et al. (2025), event-event edges explicitly capture both causal and temporal dependencies. Additionally, a dedicated GENERATES_OBSERVATION edge is implemented to link a Macro-Trend or Micro-Event directly to a Parameter Observation node, which is then ANCHORS_TO Parameter. The complete scheme governing these edge types is detailed in Table 10.

Table 10. Edge types and relation categories in the EKG schema

Edge	**Entity-Entity**	**Entity-Event**	**Event-Event**
Edge Type	'USES', 'PRODUCES', 'CONSUMES', 'SUPPLIES', 'PARTNER_OF', 'LOCATED_IN', 'SUBSTITUTES', 'TRANSFORMS_INTO', 'RECYCLED_FROM'	'DISRUPTS', 'AFFECTS', 'DELAYS', 'HALTS', 'OCCURS_AT', 'DRIVEN_BY', 'ANCHORS_TO'	'CAUSES', 'PRECONDITION', 'TEMPORAL_PRECEDES', 'EVIDENCES', 'GENERATES_OBSERVATION'

Export to Neo4j Agent
This agent is designed to export the extracted JSON data into the Neo4j graph database first to help next agent to implement node resolution directly using Cypher.

Node Resolution Agent
The node resolution process is designed to merge duplicate or semantically equivalent nodes to do entity resolution (Obraczka et al., 2021) and event coreference (Knez & Žitnik, 2023). This is because extracting text inevitably generates nodes that represent the exact same real-world concept but with different naming variations. Leaving these redundant nodes unmerged will severely fragment the network and dilute network analytical results. Furthermore, relying exclusively on an LLM to compare every single node across the entire graph is computationally expensive and highly inefficient.

As a result, this agent uses a hybrid approach combining semantic embeddings, targeted LLM reasoning, and database components in Neo4j. First, nodes are clustered using semantic embeddings. These embeddings transformed textual attributes into dense vector representations, facilitating the evaluation of semantic similarity through spatial proximity (Mikolov et al., 2013). Simply put, nodes with short spatial distances indicated similar meanings and are grouped together. Embedding-based methods are often combined with a predefined similarity threshold. If the distance between two embedded representations is smaller than the threshold, the corresponding items are treated as semantically similar. The embedding model used in this study is OpenAI's text-embedding-3-small.

Next, the LLM evaluates in the clusters to determine which nodes should be merged and to select the most standard canonical node name. Finally, the actual merging commands are executed directly within the graph using Neo4j APOC (Awesome Procedures on Cypher) components, which is much faster than re-writing the edges in JSON file using code.

Crucially, a HITL pause is integrated into this step. Nevertheless, node merging based on this criterion may create several challenges. It can lead to an overly dense graph that is difficult to interpret, and important node attributes or source information may be lost during the merging process. Human experts need to inspect the graph to determine whether it has genuinely converged into a dense network and assessed its overall quality, adjusting the extractor, resolver, or rerun if necessary. By combining clustering, LLM judgment, Neo4j, and human oversight, the agent can produce a cohesive graph where fragmented relations are reconnected, all while maintaining all node attributes and source references.

Network Centrality Analysis Agent
This agent is designed to evaluate the profile and carry messages from EKG. More importantly, to identify the most critical monitoring targets with high centrality in the network perspective. This

is because monitoring every node or every news is computationally impossible and time-consuming.

Following centrality definition by Pandey et al. (2026), this study uses two network centrality measures to identify strategic nodes in the EKG. Although the original EKG contains directed relationships, degree and betweenness centrality are calculated on an undirected graph projection. This means that the analysis focuses on the structural importance of each node rather than the direction of flow. Degree centrality therefore indicates how strongly a node is connected to the overall knowledge structure, while betweenness centrality indicates whether a node acts as a bridge between different clusters.

Therefore, Neo4j are used calculate two distinct network topology metrics: degree centrality to identify “system hubs”, where highly connected nodes with broad potential influence across the graph and betweenness centrality to identify “chokepoints”, where critical bridges vulnerable to single-point failures.

As a result, the agent can extract the top 10 System Hubs and top 10 Chokepoints in Neo4j query, using only persistent Entities and Macro-Trends. Micro-Events, Parameter Observations, and Parameters are excluded because centrality analysis was to identify monitoring targets rather than one-time observations. Finally, results are exported in two formats: structured JSON files for downstream computational use and a written report for human interpretation.

GraphSAGE-supported LP for Event-to-Parameter Mapping

Before Phase 1 scenario generation, the historical EKG needs to be prepared for LP, which is a technology to do graph completion (Rossi et al., 2021). The primary objective of scenario generation in this study is not deterministic future forecasting, but rather the exploratory mapping of potential pathways that could trigger parameter alterations. Therefore, LP is employed to support exploratory event-to-parameter mapping by identifying potential relationships between newly monitored events and historical parameter-related pathways in the EKG.

Traditional embedding models, such as TransE, are transductive and require the whole graph to be embedded again when new nodes are added (Gu et al., 2026). This makes it difficult to predict links for unseen event entities. To address this limitation, this study is inspired by the Case-Based Reasoning (CBR) approach of EvCBR (Shirai et al., 2023) and uses GraphSAGE (Graph Sample and AggregatE), an inductive Graph Neural Network (GNN), as the core LP model by learning the graph structure (Hamilton et al., 2017).

GraphSAGE generates node embeddings by aggregating information from a node’s own features and its local neighborhood. Unlike transductive embedding methods, which learn fixed embeddings for nodes already present in the graph, GraphSAGE can generate embeddings for newly added nodes based on their surrounding graph context. This property makes it suitable for the proposed framework because Phase 1 continuously introduces new event nodes extracted from recent news. Since these new nodes were not part of the original historical EKG, the framework requires an inductive method that can represent unseen nodes without retraining the entire graph whenever new evidence is added.

In Phase 0, GraphSAGE is trained on the historical EKG. Through this training, the model learns how different types of nodes, such as events, actors, locations, materials, parameter observations, and LCI parameters, are structurally connected. In Phase 1, when new event nodes are added to the EKG, the trained GraphSAGE model generates embeddings for these nodes based on their features and neighboring nodes. These embeddings are then compared with historical parameter observation nodes and parameter-related pathways to identify possible hidden event-parameter links.

Finally, the predicted links are not interpreted as causal proof and are not used to directly calculate parameter-change values. Instead, they provide graph-based contextual evidence for the Scenario Generation agent. The Scenario Generation agent then combines the predicted link, the source document, historical EKG evidence, trigger mechanism classification, and the scenario boundary matrix to generate bounded best-, average-, and worst-case parameter changes. In this way, GraphSAGE supports traceable and scalable event-to-parameter linking, while the final numerical scenario generation remains constrained by the expert-defined scenario boundary matrix.

Global Embedding Agent

The Global Embedding agent assigns two types of mathematical representations to every node in the historical EKG: semantic embeddings and GraphSAGE embeddings. These two embeddings support different tasks in the framework.

Semantic embeddings represent the textual meaning of each node based on its name and description. They place all nodes in a shared vector space according to semantic similarity. This allows the system to compare reduced duplicate or semantically equivalent nodes in Phase 1, as well as pre-processing for GraphSAGE embeddings.

GraphSAGE embeddings represent the structural position of each node in the EKG. In Phase 0, the GraphSAGE model is trained on the historical EKG using PyTorch Geometric (PyG). The semantic embeddings of node names and descriptions are used as node features, while the graph structure provides neighborhood information. Through this process, GraphSAGE learns how nodes are positioned in relation to actors, locations, materials, macro-trends, micro-events, parameter observations, and selected LCI parameters.

In Phase 1, the two embeddings are used for different purposes. Semantic embeddings support Global Node Resolution by identifying whether a newly extracted node already exists in the historical EKG. GraphSAGE embeddings support LP by comparing newly monitored events with historical parameter-related pathways. In short, semantic embeddings answer whether two nodes refer to the same concept, while GraphSAGE embeddings answer whether a new event is structurally similar to historical parameter-related events. Therefore, the Global Embedding agent prepares the historical EKG for Phase 1 by creating semantic representations for duplicate-node detection and structural representations for identifying candidate event-parameter links.

Appendix C: The design details of agents in Phase 1

Trend Monitoring Agents

This agent is designed to monitor external news for updates on the "Strategy Nodes" identified in Phase 0 and conduct a broader search on the core LCA parameters. This is because we must capture new real-world on targeted trend; while ensuring we do not miss wider parameter shifts by only focusing on specific targets.

Document Embedding Filter Agent

This agent is designed to filter out similar news from the Trend Monitoring agent to ensure the representative of the newly added events without repeating. This filter applied an embedding model again to convert captured news into semantic embedding, filtering out articles that were semantically similar to previously processed documents. This provided a fresh, diverse, and non-redundant stream of real-time data. The semantic embedding difference between Document Embedding Filter agent and Global Embedding agent is level of application, former is document-level, while later is node-level.

EKG construction from monitored news

The Node Extraction, Edge Extraction, and Node Resolution agents are designed to transform this filtered real-time news into an EKG as Phase 0 did. This is because the news should be presented as nodes and edges so that the GraphSAGE model can accurately recognize and compare them, finally discovering potential links. Thus, the same three-agent extraction pipeline from Phase 0 was reused but explicitly instructed the system to retain pre-existing nodes (Strategy Nodes) as architectural anchors. This helped the agent to obtain a structured daily mini-EKG that securely hooks into the foundational EKG, preventing the creation of disconnected orphan nodes.

Global Node Resolution Agent

Next, this agent is designed to merge the newly extracted event subgraph with the historical EKG. This is because failing to merge them will cause the graph to diverge. This agent uses Neo4j's high-performance vector search combined with LLM verification. Newly extracted nodes were converted into semantic embeddings and mapped against the existing graph's semantic embeddings from Global Embedding in Phase 0. Candidate nodes exceeding a similarity threshold were then retrieved for a final LLM-based verification. If they are highly similar, the candidate nodes will be removed since there are already nodes in the EKG that can represent them.

Noticeably, to prevent semantic collapse, defined as the erroneous merging of similar but temporally distinct events, the LLM is required to verify temporal attributes before resolution. This agent then produced a final integrated Neo4j database, where new nodes are safely assigned updated semantic embeddings and properly fused with historical data, keeping the EKG accurate and ready for LP.

Export to Neo4j Agent

This agent is designed to import the JSON file into Neo4j like phase 0. It tags newly added nodes with the "Updated" label to help subsequent agents identify them and extract updated nodes and their immediate neighbors for the next agent to perform GraphSAGE embedding.

Scenario Generation Agent

To ensure the logical plausibility and traceability of the generated scenarios, the Scenario Generation agent translates newly monitored events into bounded LCI parameter-change scenarios employing a Graph RAG approach.

Traditional document-level RAG retrieves isolated text chunks, which often fails to capture complex causal chains and supply chain dynamics (Gao et al., 2025). Graph RAG overcomes this limitation by grounding the LLM's reasoning in the structural connections of a knowledge graph. Instead of just reading text snippets, the AI can see exactly how different events and parameters are linked, enabling more logical and context-aware inferences (Edge et al., 2024).

In this framework, the Scenario Generation agent uses different tools for distinct tasks, with Link Prediction (LP) acting as the core retrieval mechanism of Graph RAG architecture. It first performs LP between newly added event nodes labeled "Updated" and parameter-related nodes in the EKG. The GraphSAGE model identifies these possible hidden event-parameter links and retrieves relevant historical parameter pathways. Subsequently, Neo4j is used to retrieve the related graph evidence and source information. The LLM then uses this graph-based contextual evidence to explain the possible impact mechanism and generate bounded scenario values.

The Scenario Generation agent then uses different tools for different tasks. GraphSAGE model is used to identify possible links between new events and LCI parameters. Neo4j is used to retrieve related graph evidence and source information. LLM then uses this evidence to explain the possible impact mechanism and generate bounded scenario values. Table 11 summarizes the role of each component.

The workflow consists of six steps. First, the GraphSAGE model trained in Phase 0 generates structural GraphSAGE embeddings for the newly updated event nodes and their neighbors. Second, a vector similarity search with GraphSAGE embeddings identifies historical parameter observations or selected LCI parameter nodes that share a similar network structure. Third, the original monitored news document is retrieved to serve as textual evidence.

Fourth, because LLMs only process natural language, the retrieved graph evidence, such as historical events and their connection pathways, is first translated into plain text. The LLM then acts as a comprehensive reasoning engine by synthesizing three key inputs: (1) the text-based graph context from LP, including parameter and historical cases from parameter observation, (2) the original news document as evidence text, (3) scenario boundary matrix. By evaluating the new event through this structural historical lens and expert-defined boundaries, the scenario generator can understand the hidden relationships and generate bounded best-, average-, and worst-case percentage changes.

Fifth, the relevant updated events are dynamically connected to the corresponding parameters using a PREDICTED_IMPACT edge. Finally, Neo4j labels are strictly managed to distinguish these temporary predicted links from empirical historical relationships, thereby preventing graph pollution.

To reduce hallucinations and enforce consistency, the LLM follows a three-step reasoning protocol. First, it checks whether the evidence describes a real and relevant event with a plausible physical connection to the target LCI parameter. Second, it determines the direction and severity of the parameter-change signal, assigning a severity level from 1 to 4 based on the evidence, historical EKG cases, and trigger mechanism. Third, it converts the assessed direction and severity into best-, average-, and worst-case percentage changes within the expert-defined scenario boundary matrix.

Table 11. Functional division of components in the Scenario Generation agent

Component	Task in Scenario Generation	Output
GraphSAGE model	Inductive embedding and topological inference	Structural embeddings and candidate event-parameter links
Neo4j / Cypher	Graph querying, vector search, and evidence retrieval	Historical pathways, parameter observations, and source evidence

LLM Scenario Generator	Mechanism interpretation and bounded scenario generation	Best-, average-, and worst-case parameter changes
Scenario boundary matrix	Numerical constraint and severity mapping	Allowed percentage-change ranges
Human expert	Boundary definition and final interpretation	Plausible and LCA-relevant scenario assumptions

Together, these components allow the agent to separate graph-based candidate link detection from LLM-based scenario reasoning and expert-constrained numerical quantification.

Appendix D: The Metrics in the Validations

Stage 1: Grader Validation

According to ground truth in Table 12, a positive case refers to a document describing an external event that can be linked to at least one selected CRM or LCI parameter through a defined trigger mechanism. The document must provide evidence that the event could reasonably affect a parameter used in the scenario reasoning.

In contrast, a hard negative case refers to a document that appears semantically related to the selected parameters, materials, regions, or disruption triggers, but does not provide sufficient evidence of a direct or plausible parameter-level effect. In this study, hard negatives included four main types of cases: (1) articles related to similar parameter names but referring to different indicators, such as China's general carbon intensity rather than grid carbon intensity; (2) retrospective articles describing past trends or historical capacity additions without indicating a current or future parameter change; (3) early-stage plans, announcements, or policy ambitions that had not yet produced observable effects; and (4) articles mentioning relevant disruption triggers, such as the Strait of Hormuz crisis or El Niño, but focusing only on secondary economic or social impacts rather than physical route changes, generation changes, or other LCI-relevant mechanisms.

Table 12. Ground truth metrics for positive and negative case annotation

Criterion	Positive case: included	Negative case: excluded
Temporal relevance	Active, ongoing, or near-term events that may affect the selected assessment period	Retrospective news, historical summaries, long-term aspirations, or events without near-term relevance
Parameter match	Events directly linked to one of the predefined LCI parameters: electricity carbon intensity or graphite transport distance/routing	Events related only to broad market trends, demand, prices, investment, or company strategy without direct LCI parameter relevance
Geographic / system scope	Events occurring within the relevant physical system, such as China's electricity system or major graphite transport routes	Events outside the defined geographical, technological, or supply-chain boundary
Physical causal mechanism	Events that plausibly alter physical drivers, such as fuel switching, renewable curtailment, hydropower shortage, rerouting, port congestion, or sourcing shifts	Purely commercial or financial changes without physical alteration, such as price fluctuation, freight cost increase, market speculation, or stock movement
Evidence sufficiency	The source provides enough evidence to infer how and why the event may affect the parameter	The source is vague, promotional, unsupported, or lacks a clear causal explanation

Stage 2: Scenario plausibility validation

Table 13 shows three metrics for the scenario plausibility validation (meth rule and logical quality) and Table 14 shows the score sheet reasoning for LLM-as-a-judge.

Table 13. Scenario plausibility validation metrics

Metric	Purpose	Pass / fail rule
Numerical Consistency	Checks whether Best, Average, and Worst values follow the expected numerical order according to the scenario direction	Pass if the values are ordered consistently with the direction of impact; fail if the order is mathematically inconsistent

Boundary Compliance	Checks on whether generated values remain within the expert-defined scenario boundary matrix	Pass if all values fall within the allowed minimum and maximum range for the selected parameter and severity level; fail if any value exceeds the boundary
Reasoning Quality	Evaluates whether the explanation justifies the selected direction, severity, and numerical values	Scored from 1 to 5 by an independent LLM-as-a-Judge. Pass if the independent reasoning score is greater than 3; fail if the score is 3 or below.

Table 14. Score sheet for LLM-as-a-Judge reasoning quality validation

Score	Level	Criteria
5	Excellent	Flawless logic: Accurately extracts the mechanism (Step 1) and chooses the severity level (Step 2). Perfectly justifies the magnitude using evidence within the scenario boundary matrix limits (Step 3).
4	Good	Logical & Evidence-backed: Logic is sound, but the percentage magnitude in Step 3 is slightly conservative or aggressive given the severity of the evidence.
3	Fair	Generic reasoning: Acceptable logic but lacks depth. Fails to deeply integrate historical references or specific details from the evidence.
2	Poor	Logical disconnect: Contradictions between steps (e.g., deducing "Decrease" but generating positive numbers) or completely ignoring Parameter minimum and maximum limits within the severity level.
1	Fail / Hallucination	Fabrication: Fabricates facts do not present in the raw evidence, maps to the wrong mechanism, or the reasoning contradicts basic domain knowledge or physical plausibility.

Appendix E: The Codes of Document Grader and Answer Grader

```
import os
from langchain_core.prompts import ChatPromptTemplate
from pydantic import BaseModel, Field
from langchain_openai import ChatOpenAI

api_key = os.getenv("AZURE_OPENAI_KEY")
endpoint = os.getenv("AZURE_OPENAI_ENDPOINT")


llm = ChatOpenAI(
    model = "gpt-5-nano",
```

```
    base_url=endpoint,
    api_key=api_key,
    temperature= 0,
)

class DocumentGrade(BaseModel):
    is_credible: bool = Field(
    description="True if the source is from a third-party news outlet, industry
tracker, official government domain, or corporate press release. False ONLY if it is
explicitly an anonymous rant, personal social media post, or unbacked personal blog."
    )
    is_relevant: bool = Field(
        description="True if the text contains physical supply chain events altering
LCA parameters (e.g., factory building, material shift). False if it's purely
financial (stock prices, earnings) or general industry commentary."
    )
    is_factual: bool = Field(
    description="True if REALIZED (physically occurred), an ONGOING macro-shift, or
an empirical scientific study revealing current structural limits. False ONLY if
purely speculative without physical action."
    )
    rationale: str = Field(
        description="One brief sentence explaining the grading decision. If either is
False, state which one failed and why."
    )

document_grader_prompt = ChatPromptTemplate.from_messages([
    ("system", """You are a strict "Intelligence Filter" for an automated supply
chain monitoring pipeline.
Your job is to evaluate incoming documents based on THREE strict criteria. You must
act as a ruthless filter to prevent noise and hallucinations from entering the
system.

[CRITERIA 1: CREDIBILITY (Is the source a valid organizational or journalistic
entity?)]
You MUST apply a highly LENIENT standard for credibility. The data pipeline has
already filtered out pure spam.
- PASS (Valid):
  1. Government or Official domains (e.g., .gov, .gc.ca).
  2. Corporate Press Releases (These are Primary Sources. Do NOT penalize them for
using promotional or PR language).
  3. Industry-specific News Portals, B2B platforms, or independent news aggregators
(Even if they are not mainstream media like Reuters).
```

```
- FAIL (Invalid): ONLY output FALSE if the text is blatantly an unbacked personal
opinion, an anonymous social media rant (e.g., a random Reddit post), or purely
subjective fiction. If a third-party organization is involved, it is CREDIBLE.

[CRITERIA 2: RELEVANCE & BOUNDARY MATCH (CRITICAL)]
- PASS: Contains events that can disrupt the target LCA parameter: '{parameter}'.
- FAIL (Thematic Mismatch): The document discusses logistics/shipping when the target
is electricity/carbon, OR vice versa.
- FAIL (Geographic Mismatch): The document discusses an event in a completely
unrelated region (e.g., US grid changes when the parameter is strictly China's grid)
UNLESS a direct cross-border physical domino effect is explicitly mentioned.
- FAIL (Purely Commercial Noise): Generic financial background or daily stock ticker
movements.

[CRITERIA 3: FACTUALITY & ACTION STATE (Is it a concrete physical shock?)]
- PASS (Realized or Authoritative Prediction): The event has physically occurred,
been officially enacted, OR it is a confirmed, high-probability authoritative
prediction (e.g., "Official drought forecast issued", "Groundbreaking for a new
plant").
- FAIL (Speculative or Purely Retrospective):

[WHITELIST FOR FACTUALITY (Must Output TRUE)]
1. Scientific & Empirical Studies: Peer-reviewed studies (e.g., Nature, academic
reports) mapping or exposing CURRENT, ONGOING physical constraints (e.g., grid
inflexibility, curtailment limits) are FACTUAL. They describe the actual state of the
world, not speculations.
2. Ongoing Geopolitical Shifts: Reports of a nation actively changing its physical
supply chain (e.g., "replacing Middle East oil with domestic coal") are FACTUAL
structural changes, not merely analyst forecasts.

Output true ONLY IF the document provides actionable intelligence that directly
answers or supports the provided Question/Intent.
"""),
    ("human", """
Content Snippet:
{context};
url:{url}
""")
])


document_grader = document_grader_prompt | llm.with_structured_output(DocumentGrade)
```

```
import os
from pydantic import BaseModel, Field
```

```python
from langchain_core.prompts import ChatPromptTemplate
from langchain_openai import ChatOpenAI


api_key = os.getenv("AZURE_OPENAI_KEY")
endpoint = os.getenv("AZURE_OPENAI_ENDPOINT")


llm = ChatOpenAI(
    model = "gpt-5-nano",
    base_url=endpoint,
    api_key=api_key,
    temperature= 0,
)


class AnswerGrade(BaseModel):
    is_answered: bool = Field(
        description="True IF the context can answer the user question"
    )
    reasoning: str = Field(
        description="A 1-sentence explanation of why this document can or can not
answer user question."
    )

# Prompt
answer_grader_prompt = ChatPromptTemplate.from_messages([
    ("system", """You are a Quality Assurance Evaluator for an Event Knowledge Graph
(EKG).
Your ONLY task is to determine if the text contains SUFFICIENT CAUSAL EVIDENCE to
alter the target LCA parameter: '{parameter}'.

[CORE DEFINITION: PHYSICAL vs. COMMERCIAL]
To output TRUE, the event MUST trigger or be HIGHLY EXPECTED to trigger a "Physical
Structure Alteration" that STRICTLY MATCHES the Geography and Domain of
'{parameter}'.
- EXCLUDE Purely Commercial Fluctuations: Do NOT output TRUE if the text ONLY
describes freight rates or price hikes WITHOUT physical grid/routing changes.

[TEMPORAL HORIZON & ACTION STATE]
1. TRUE: Active/Ongoing Shocks.
2. TRUE: Authoritative Predictive Warnings & Strategic Lock-ins backed by credible
bodies.
3. FALSE: Early-Stage/Aspirational goals without immediate physical lock-ins.
```

```
[VALID MECHANISMS BY PARAMETER]
Identify which category '{parameter}' belongs to, and ONLY apply that category's
logic:
If 'Grid Carbon/Energy': Focus on Fuel Switching, Extreme Weather affecting
renewables, Policy/Infrastructure surges, or Energy Market Spillovers.
If 'Transport/Logistics': Focus on Route Disruption, Port Congestion Spillover, or
Sourcing Shifts.

[LENIENCY RULES - ONLY IF GEOGRAPHY MATCHES] 🌟
You must ONLY apply these assumptions if the event occurs in the SAME GEOGRAPHIC
REGION or SUPPLY CHAIN NODE as the '{parameter}':
- If a major chokepoint is blocked, ASSUME geographical rerouting (TRUE).
- If severe weather halts renewables, ASSUME fossil fuel compensation (TRUE).
- The Grid Seesaw: If clean energy is curtailed/constrained, ASSUME fossil fuels
increase to fill the gap (TRUE).
- Fuel Switching: Mention of "gas-to-coal" or "reopening coal" is absolute proof of
mix change (TRUE).

[RED FLAGS FOR IMMEDIATE FALSE (STRICT EXCLUSION)] 🚨
Even with leniency, you MUST output FALSE if the text contains these red flags:
1. Geographic Mismatch: The disruption occurs in a completely unrelated region (e.g.,
US/Texas grid issues when evaluating China's Grid Carbon, or vice versa).
2. Thematic Mismatch: The text discusses shipping/transport issues (e.g., Red Sea,
Panama) when evaluating a domestic electricity grid parameter, or vice versa.
3. Analyst/Expert Speculation: General opinions from observers ("Analysts expect...")
rather than direct actions.
4. Downstream Commercial Consequences: Focuses entirely on secondary economic impacts
(inflation, retail prices) rather than the physical chokepoint."""),
    ("human", "Documents: {context}")
])

answer_grader = answer_grader_prompt | llm.with_structured_output(AnswerGrade)
```

Appendix F: List of Retained Highly Relevant Articles (Sample)

Table 15. List of Retained Highly Relevant Articles with URL links

No.	Article Title	Source	Date	URL
1	More renewables, more coal: Where are China's emissions really headed?	Dialogue Earth	April 3, 2024	Link

2	Sichuan power crunch sparks calls for rethink of coal in China's energy mix	ThinkChina	Aug 26, 2022	Link
3	The Compound Heatwave and Drought Event in the Summer of 2022 and the Impacts on the Power System in Southwest China	Energies Journal	May 5, 2025	Link
4	Blocked Suez Forces Ships to Look at Long Trip Around Africa	SupplyChainBrain	March 25, 2021	Link
5	Drought behind Panama Canal's 2023 shipping disruption 'unlikely' without El Niño	Carbon Brief	May 01, 2024	Link
...	...	...	...	...
80	Auto firms race to secure non-Chinese graphite for EVs as shortages loom	MINING.COM	Jun 21, 2023	Link

Appendix G: The Calculation and Threshold of EKG

The main value of the EKG is that it can connect participants, events, and parameters into a structured network. Events that may appear unrelated to isolation can become linked through shared mechanisms or common effects. This allows the EKG to support early detection of emerging trends and possible domino effects in the supply chain. In other words, the graph does not only store individual event records; it helps reveal how small signals may be connected to wider system-level changes.

To evaluate whether the EKG formed a meaningful network rather than a set of isolated event chains, this study used Global Network Density (GND) for directed graph as a structural indicator as defined by Jilbert (2020) in Equation 6:

$$GND = \frac{E}{N(N-1)} \text{ (Equation 6)}$$

Where E is the number of directed edges and N is the number of nodes.

If the GND falls below the inverse of the number of nodes, implying that each node is connected to fewer than one edge on average, the graph fragments into isolated clusters, failing to form a giant component (Durrett, 2010). This study uses this threshold to evaluate the effectiveness of Node Resolution agent, proving it has transitioned from fragmented, independent events into a structurally connected knowledge network. Before Node Resolution agent, the graph contained 2,289 nodes and 2,099 edges, with a GND of 0.0004. After the Node Resolution Agent merged duplicate and semantically similar nodes, and after researcher review of content quality, graph density, and interpretability, the final graph was reduced to 760 nodes and 1,799 edges. As a result, GND increased to 0.003, which is about 7.5 times higher than the initial value and above 0.0013. However, this structural improvement should not be interpreted as proof of graph completeness or causal exhaustiveness. It only indicates that the final graph became more connected and analytically usable after node resolution.

Reference

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey. *IEEe Access*, *13*, 18912-18936.

Arias, A., Moreira, M. T., Heijungs, R., & Cucurachi, S. (2025). Advancing parametric life cycle assessment (pa-LCA): A systematic review and methodological roadmap for enhanced sustainability assessments. *Sustainable Production and Consumption*.

Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2024). Self-rag: Learning to retrieve, generate, and critique through self-reflection. International conference on learning representations,

Attílio, L. A. (2026). Geopolitical tensions between the US and China and renewable energy. *Energy Policy*, *208*, 114893.

Auffhammer, M., Wang, M., Xie, L., & Xu, J. (2021). Renewable electricity development in China: Policies, performance, and challenges. *Review of Environmental Economics and Policy*, *15*(2), 323-339.

Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic ai: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet*, *17*(9), 404.

Bednarski, L., Roscoe, S., Blome, C., & Schleper, M. C. (2025). Geopolitical disruptions in global supply chains: a state-of-the-art literature review. *Production planning & control*, *36*(4), 536-562.

Beloin-Saint-Pierre, D., Albers, A., Hélias, A., Tiruta-Barna, L., Fantke, P., Levasseur, A., Benetto, E., Benoist, A., & Collet, P. (2020). Addressing temporal considerations in life cycle assessment. *Science of The Total Environment*, *743*, 140700.

Bhattacharyya, S., Roy, S., Lin, X., Campagnol, N., Vlad, A., & Ajayan, P. M. (2026). Graphite: the new critical mineral. *Nature Reviews Materials*, *11*(1), 65-78.

Bobba, S., Carrara, S., Huisman, J., Mathieux, F., & Pavel, C. (2020). Critical raw materials for strategic technologies and sectors in the EU. A Foresight Study.

Britt, A., & Czarnota, K. (2024). A review of critical mineral resources in Australia. *Australian Journal of Earth Sciences*, *71*(8), 1016-1049.

Cheng, Z.-Q., Dong, Y., Shi, A., Liu, W., Hu, Y., O'Connor, J., Hauptmann, A. G., & Whitefoot, K. (2024). Shield: Llm-driven schema induction for predictive analytics in ev battery supply chain disruptions.

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track,

Cimprich, A., Bach, V., Helbig, C., Thorenz, A., Schrijvers, D., Sonnemann, G., Young, S. B., Sonderegger, T., & Berger, M. (2019). Raw material criticality assessment as a complement to environmental life cycle assessment: Examining methods for product-level supply risk assessment. *Journal of Industrial Ecology*, *23*(5), 1226-1236.

Cornago, S., Tan, Y. S., Brondi, C., Ramakrishna, S., & Low, J. S. C. (2022). Systematic literature review on dynamic life cycle inventory: towards industry 4.0 applications. *Sustainability*, *14*(11), 6464.

da Costa, T. P., da Costa, D. M. B., & Murphy, F. (2024). A systematic review of real-time data monitoring and its potential application to support dynamic life cycle inventories. *Environmental Impact Assessment Review*, *105*, 107416.

Deng, J., Chen, C., Huang, X., Chen, W., & Cheng, L. (2023). Research on the construction of event logic knowledge graph of supply chain management. *Advanced engineering informatics*, *56*, 101921.

Derouiche, H., Brahmi, Z., & Mazeni, H. (2025). Agentic ai frameworks: Architectures, protocols, and design challenges. *arXiv preprint arXiv:2508.10146*.

Dong, F., Yu, B., Hadachin, T., Dai, Y., Wang, Y., Zhang, S., & Long, R. (2018). Drivers of carbon emission intensity change in China. *Resources, Conservation and Recycling*, *129*, 187-201.

Dong, Y., He, J., Hou, Y., Du, D., Xu, Z., Yu, S., Xia, Y., & Chen, H. (2026). DeltaBox: Scaling Stateful AI Agents with Millisecond-Level Sandbox Checkpoint/Rollback. *arXiv preprint arXiv:2605.22781*.

Durrett, R. (2010). *Random graph dynamics* (Vol. 20). Cambridge university press.

Edge, D., Krzyzanowski, T., & Basisty, F. (2024). A Graph RAG Approach to Query-Focused Summarization. *arXiv preprint arXiv:2404.16130*.

Elias Mota, B. A., Cerqueira de Sousa Gouveia Carvalho, A. I., Azevedo Rodrigues Gomes, M. I., & Ferreira Dias Barbosa-Povoa, A. P. (2020). Business strategy for sustainable development: Impact of life cycle inventory and life cycle impact assessment steps in supply chain design and planning. *Business Strategy and the Environment*, *29*(1), 87-117.

Ericsson, M., & Löf, O. (2020). *Extractive dependency in lower-income countries: Evolving trends during the transition to a low-carbon future* (9292568779).

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., & Wahler, A. (2020). Introduction: what is a knowledge graph? In *Knowledge graphs: Methodology, tools and selected use cases* (pp. 1-10). Springer.

Ferrari, A. M., Volpi, L., Settembre-Blundo, D., & García-Muiña, F. E. (2021). Dynamic life cycle assessment (LCA) integrating life cycle inventory (LCI) and Enterprise resource planning (ERP) in an industry 4.0 environment. *Journal of Cleaner Production*, *286*, 125314.

Fnais, A., Rezgui, Y., Petri, I., Beach, T., Yeung, J., Ghoroghi, A., & Kubicki, S. (2022). The application of life cycle assessment in buildings: challenges, and directions for future research. *The International Journal of Life Cycle Assessment*, *27*(5), 627-654.

Francis, N., Taylor, A., Green, A., Guagliardo, P., Libkin, L., Lindaaker, T., Marsault, V., Plantikow, S., Rydberg, M., & Selmer, P. (2018). *Cypher: An Evolving Query Language for Property Graphs*. https://doi.org/10.1145/3183713.3190657

Gaddala, V. S. (2023). Unleashing the Power of Generative AI and RAG Agents inSupply Chain Management: A Futuristic Perspective. *IRE Journals*, *6*(12), 1411-1417.

Gao, Y., Xiong, Y., Zhong, Y., Bi, Y., Xue, M., & Wang, H. (2025). Synergizing rag and reasoning: A systematic review. *arXiv preprint arXiv:2504.15909*.

Ghosh, A., & Sourav, S. (2026). When Shocks Are Permanent: Port Congestion-Induced Hysteresis in Maritime Disruptions-Evidence from COVID-19, Suez, and Red Sea Crises.

Granvik, P., Hanski, J., Lähdesmäki, S., Jokilaakso, A., & Huttunen-Saarivirta, E. (2025). Critical raw materials for green transition: Key parameters and feasibility index for sufficiency. *Resources,*

Conservation and Recycling, *218*, 108197. https://doi.org/https://doi.org/10.1016/j.resconrec.2025.108197
Gridin, I. (2025). Building Agent Systems. In *The Practical Guide to Large Language Models: Hands-On AI Applications with Hugging Face Transformers* (pp. 175-223). Springer.
Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., & Liu, H. (2026). A survey on llm-as-a-judge. *The Innovation*, *7*(6).
Guan, S., Cheng, X., Bai, L., Zhang, F., Li, Z., Zeng, Y., Jin, X., & Guo, J. (2022). What is event knowledge graph: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *35*(7), 7569-7589.
Gupta, S., Ranjan, R., & Singh, S. N. (2024). A comprehensive survey of retrieval-augmented generation (rag): Evolution, current landscape and future directions. *arXiv preprint arXiv:2410.12837*.
Hosamo, H., Coelho, G. B., Buvik, E., Drissi, S., & Kraniotis, D. (2024). Building sustainability through a novel exploration of dynamic LCA uncertainty: Overview and state of the art. *Building and Environment*, *264*, 111922.
Hosseini, S., & Seilani, H. (2025). The role of agentic ai in shaping a smart future: A systematic review. *Array*, *26*, 100399.
Hu, Z., Li, Z., Jin, X., Bai, L., Guo, J., & Cheng, X. (2025). Large language model-based event relation extraction with rationales. Proceedings of the 31st International Conference on Computational Linguistics,
Huang, K., & Hughes, C. (2025). Agentic AI Communication Protocols and Security. In *Securing AI Agents: Foundations, Frameworks, and Real-World Deployment* (pp. 81-110). Springer.
Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., & Qin, B. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, *43*(2), 1-55.
Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I., Dutot, V., Appanderanda, M., Crick, T., & De', R. (2025). AI agents and agentic systems: A multi-expert analysis. *Journal of Computer Information Systems*, *65*(4), 489-517.
Hülagü, S., Dullaert, W., Eruguz, A. S., Heijungs, R., & Inghels, D. (2025). Integrating life cycle assessment into supply chain optimization. *PloS one*, *20*(1), e0316710.
Ibn-Mohammed, T., Mustapha, K., Abdulkareem, M., Fuensanta, A. U., Pecunia, V., & Dancer, C. (2023). Toward artificial intelligence and machine learning-enabled frameworks for improved predictions of lifecycle environmental impacts of functional materials and devices. *MRS Communications*, *13*(5), 795-811.
International Energy Agency. (2026). *Emissions – Electricity 2026 – analysis*. IEA. https://www.iea.org/reports/electricity-2026/emissions
Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, *33*(2), 494-514.
Jilbert, O. L. a. I. (2020). *2.9 density | Social Networks: An Introduction*. https://bookdown.org/omarlizardo/_main/2-9-density.html
Joshi, S. (2025). A comprehensive survey of ai agent frameworks and their applications in financial services. *Available at SSRN 5252182*.
Keith, B., Horning, M., & Mitra, T. (2020). Evaluating the inverted pyramid structure through automatic 5w1h extraction and summarization. *Computational Journalism C+ J*.
Khalid, O., Farooqi, A. U. H., & Bilal, M. (2026). Agentic AI: a review, applications and open research challenges. *Computing*, *108*(6), 80.
Knez, T., & Žitnik, S. (2023). Event-centric temporal knowledge graph construction: A survey. *Mathematics*, *11*(23), 4852.
Krishnan, N. (2025). Advancing multi-agent systems through model context protocol: Architecture, implementation, and applications. *arXiv preprint arXiv:2504.21030*.

Lei, Z. (2024). LangGraph vs LangChain: Which Framework Is Best for AI Agents? *MetaVision Journal of Multidisciplinary Studies (MVJMS)*, *1*(4), 16-26.

Lewicka, E., Guzik, K., & Galos, K. (2021). On the Possibilities of Critical Raw Materials Production from the EU's Primary Sources. *Resources*, *10*(5), 50. https://www.mdpi.com/2079-9276/10/5/50

Li, Q., Jiang, P., Wang, J., Yang, M., & Yang, Y. (2025). A kind of intelligent dynamic industrial event knowledge graph and its application in process stability evaluation. *Journal of Intelligent Manufacturing*, *36*(3), 1801-1818.

Li, W., Yang, W., Zhang, F., Wu, S., & Li, Z. (2024). Extreme weather impact on carbon-neutral power system operation schemes: A case study of 2060 Sichuan Province. *Energy*, *313*, 133677.

Mancini, L., Sala, S., Recchioni, M., Benini, L., Goralczyk, M., & Pennington, D. (2015). Potential of life cycle assessment for supporting the management of critical raw materials. *The International Journal of Life Cycle Assessment*, *20*(1), 100-116. https://doi.org/10.1007/s11367-014-0808-0

Masudin, I., Alfarisi, S., Suhartin, E., Restuputri, D. P., & Shariff, S. S. R. (2025). Hotspot identification for improving Cane-Sugar sustainable supply chain through the integration of life cycle assessment and value stream mapping. *Process Integration and Optimization for Sustainability*, *9*(2), 605-624.

Mendoza Beltran, A., Cox, B., Mutel, C., Van Vuuren, D. P., Font Vivanco, D., Deetman, S., Edelenbosch, O. Y., Guinée, J., & Tukker, A. (2020). When the background matters: using scenarios from integrated assessment models in prospective life cycle assessment. *Journal of Industrial Ecology*, *24*(1), 64-79.

Mertens, J., Dewulf, J., Breyer, C., Belmans, R., Gendron, C., Geoffron, P., Goossens, L., Fischer, C., Du Fornel, E., & Hayhoe, K. (2024). From emissions to resources: mitigating the critical raw material supply chain vulnerability of renewable energy technologies: J. Mertens et al. *Mineral Economics*, *37*(3), 669-676.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, J. J. (2013). Graph database applications and concepts with Neo4j. Proceedings of the southern association for information systems conference, Atlanta, GA, USA,

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. Computer science on-line conference,

Narayanan, P. K. (2024). Getting Started with Data Validation using Pydantic and Pandera. In *Data Engineering for Machine Learning Pipelines: From Python Libraries to ML Pipelines and Cloud Platforms* (pp. 163-196). Springer.

Notteboom, T., Haralambides, H., & Cullinane, K. (2024). The Red Sea Crisis: ramifications for vessel operations, shipping networks, and maritime supply chains. *Maritime Economics & Logistics*, *26*(1), 1-20.

Obraczka, D., Schuchart, J., & Rahm, E. (2021). Embedding-assisted entity resolution for knowledge graphs. Second International Workshop on Knowledge Graph Construction,

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, *36*(7), 3580-3599.

Pandey, A. K., Ghiassian, S. D., & Loscalzo, J. (2026). Network-based precision medicine and systems pharmacology. *British Journal of Pharmacology*, *183*(8), 1691-1708.

Park, J., Cho, S.-J., Shin, S., Kim, R., Shin, D., & Shin, Y. (2025). Overview of graphite supply chain and its challenges. *Geosciences Journal*, *29*(3), 329-341.

Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge graphs: Opportunities and challenges. *Artificial intelligence review*, *56*(11), 13071-13102.

Peng, T., Gao, L., Agbozo, R. S. K., Xu, Y., Svynarenko, K., Wu, Q., Li, C., & Tang, R. (2024). Knowledge graph-based mapping and recommendation to automate life cycle assessment. *Advanced engineering informatics*, *62*, 102752. https://doi.org/https://doi.org/10.1016/j.aei.2024.102752

Petri, I., Amin, A., Ghoroghi, A., Hodorog, A., & Rezgui, Y. (2025). Digital twins for dynamic life cycle assessment in the built environment. *Science of The Total Environment*, *993*, 179930.
Pigné, Y., Gutiérrez, T. N., Gibon, T., Schaubroeck, T., Popovici, E., Shimako, A. H., Benetto, E., & Tiruta-Barna, L. (2020). A tool to operationalize dynamic LCA, including time differentiation on the complete background database. *The International Journal of Life Cycle Assessment*, *25*(2), 267-279.
Pommeret, A., Ricci, F., & Schubert, K. (2022). Critical raw materials for the energy transition. *European Economic Review*, *141*, 103991. https://doi.org/https://doi.org/10.1016/j.euroecorev.2021.103991
Ponnock, J., Kenneally, G., Briggs, M. R., Yeo, E., Patterson III, T., Kinberg, N., Kalinowski, M., & Hechtman, D. (2025). Real-Time RAG for the Identification of Supply Chain Vulnerabilities. *arXiv preprint arXiv:2509.10469*.
Qin, M., Su, C.-W., Umar, M., Lobonţ, O.-R., & Manta, A. G. (2023). Are climate and geopolitics the challenges to sustainable development? Novel evidence from the global supply chain. *Economic Analysis and Policy*, *77*, 748-763.
Raheem, T., & Hossain, G. (2025). Agentic ai systems: Opportunities, challenges, and trustworthiness. 2025 IEEE International Conference on Electro Information Technology (eIT),
Regulation (EU) 2023/956 of the European Parliament and of the Council of 10 May 2023 establishing a carbon border adjustment mechanism. (2023). Official Journal of the European Union
Regulation (EU) 2023/1542 of the European Parliament and of the Council of 12 July 2023 concerning batteries and waste batteries. (2023). Official Journal of the European Union Retrieved from https://eur-lex.europa.eu/eli/reg/2023/1542/oj
Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 establishing a framework for the setting of ecodesign requirements for sustainable products. (2024). Official Journal of the European Union Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1781/oj
Reisinger, J., Kugler, S., Kovacic, I., & Knoll, M. (2022). Parametric optimization and decision support model framework for life cycle cost analysis and life cycle assessment of flexible industrial building structures integrating production planning. *Buildings*, *12*(2), 162.
Ritoe, A., Patrahau, I., & Rademaker, M. (2022). Supply chain challenges & recommendations for a critical mineral.
Romeiko, X. X., Zhang, X., Pang, Y., Gao, F., Xu, M., Lin, S., & Babbitt, C. (2024). A review of machine learning applications in life cycle assessment studies. *Science of The Total Environment*, *912*, 168969.
Rospocher, M., Van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., Soroa, A., Ploeger, T., & Bogaard, T. (2016). Building event-centric knowledge graphs from news. *Journal of Web Semantics*, *37*, 132-151.
Rossi, A., Barbosa, D., Firmani, D., Matinata, A., & Merialdo, P. (2021). Knowledge graph embedding for link prediction: A comparative analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *15*(2), 1-49.
Rovelli, D., Brondi, C., Andreotti, M., Abbate, E., Zanforlin, M., & Ballarino, A. (2022). A modular tool to support data management for LCA in industry: Methodology, application and potentialities. *Sustainability*, *14*(7), 3746.
Saavedra-Rubio, K., Thonemann, N., Crenna, E., Lemoine, B., Caliandro, P., & Laurent, A. (2022). Stepwise guidance for data collection in the life cycle inventory (LCI) phase: Building technology-related LCI blocks. *Journal of Cleaner Production*, *366*, 132903.
Salla, J. V. E., de Almeida, T. A., & Silva, D. A. L. (2025). Integrating machine learning with life cycle assessment: a comprehensive review and guide for predicting environmental impacts. *The International Journal of Life Cycle Assessment*, 1-23.
Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. *Information Fusion*, 103599.

Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges. *Information Fusion*, *126*, 103599. https://doi.org/https://doi.org/10.1016/j.inffus.2025.103599

Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, *27*(4S), 4023-4031.

Shirai, S., Bhattacharjya, D., & Hassanzadeh, O. (2023). Event prediction using case-based reasoning over knowledge graphs. Proceedings of the ACM Web Conference 2023,

Shu, D., Chen, T., Jin, M., Zhang, C., Du, M., & Zhang, Y. (2024). *Knowledge Graph Large Language Model (KG-LLM) for Link Prediction*.

Sohn, J., Kalbar, P., Goldstein, B., & Birkved, M. (2020). Defining temporally dynamic life cycle assessment: a review. *Integrated environmental assessment and management*, *16*(3), 314-323.

Sprecher, B., Daigo, I., Spekkink, W., Vos, M., Kleijn, R., Murakami, S., & Kramer, G. J. (2017). Novel indicators for the quantification of resilience in critical material supply chains, with a 2010 rare earth crisis case study. *Environmental science & technology*, *51*(7), 3860-3870.

Su, S., Ju, J., Ding, Y., Yuan, J., & Cui, P. (2022). A comprehensive dynamic life cycle assessment model: considering temporally and spatially dependent variations. *International Journal of Environmental Research and Public Health*, *19*(21), 14000.

Surovtseva, D., Crossin, E., Pell, R., & Stamford, L. (2022). Toward a life cycle inventory for graphite production. *Journal of Industrial Ecology*, *26*(3), 964-979.

Taleb, N. N. (2007). Black swans and the domains of statistics. *The american statistician*, *61*(3), 198-200.

Topsakal, O., & Akinci, T. C. (2023). Creating large language model applications utilizing langchain: A primer on developing llm apps fast. International conference on applied engineering and natural sciences,

Trajanoska, M., Stojanov, R., & Trajanov, D. (2023). Enhancing knowledge graph construction using large language models. *arXiv preprint arXiv:2305.04676*.

United Nations Conference on Trade and Development. (2024a). *Disruptions in key global shipping route - Suez Canal, Panama Canal, and Black Sea - signal unprecedented challenges for global trade affecting millions of people in every region*. https://unctad.org/press-material/disruptions-key-global-shipping-route-suez-canal-panama-canal-and-black-sea-signal

United Nations Conference on Trade and Development. (2024b). *Red Sea Crisis and implications for trade facilitation in Africa. UN Trade and Development*. https://unctad.org/news/red-sea-crisis-and-implications-trade-facilitation-africa

van Gaalen, J. M., & Chris Slootweg, J. (2025). From critical raw materials to circular raw materials. *ChemSusChem*, *18*(2), e202401170.

Wang, C., Liu, X., Yue, Y., Tang, X., Zhang, T., Jiayang, C., Yao, Y., Gao, W., Hu, X., & Qi, Z. (2023). Survey on factuality in large language models: Knowledge, retrieval and domain-specificity. *arXiv preprint arXiv:2310.07521*.

Xavier, M., Laikh, T., Patil, S., & Vyatkin, V. (2024). LLM-powered multi-actor system for intelligent analysis and visualization of IEC 61499 control systems. IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society,

Yan, S.-Q., Gu, J.-C., Zhu, Y., & Ling, Z.-H. (2024). Corrective retrieval augmented generation.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Yue, Z., Mustakim, A., Mangan, J., & Yalcin, E. (2024). Carbon footprint impacts arising from disruptions to container shipping networks. *Transportation Research Part D: Transport and Environment*, *134*, 104335.

Zhang, Z., Li, Y., Lai, D., Zhou, N., Zhan, Q., & Wang, W. (2025). Quantifying Weather's Share in Dynamic Grid Emission Factors via SHAP: A Multi-Timescale Attribution Framework. *Processes*, *13*(11), 3393.

Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J., & Cui, B. (2026). Retrieval-augmented generation for ai-generated content: A survey. *Data Science and Engineering*, 1-29.

Zhu, Y., Wang, X., Chen, J., Qiao, S., Ou, Y., Yao, Y., Deng, S., Chen, H., & Zhang, N. (2024). Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *World Wide Web*, *27*(5), 58.

Zong, C., Banihashemi, F., & Lang, W. (2025). Dynamic life cycle impact assessment (DLCIA) in a sustainable building planning process. *Scientific Reports*, *15*(1), 32680. https://doi.org/10.1038/s41598-025-20599-1