2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AN ONTOLOGY-DRIVEN FRAMEWORK FOR INTELLIGENT EMBODIED CARBON DECISION-SUPPORT IN METRO STATION DESIGN

Linghui Xie, and Xueqing Zhang
Hong Kong University of Science and Technology, Hong Kong, China

Abstract

Conventional element-centric BIM-LCA integration omits temporary works (e.g., excavation support, formwork, dewatering) and faces semantic barriers (e.g., inconsistent IFC attribute naming), causing significant carbon underreporting. A dynamic ontology-driven framework was proposed to overcome this. Its semantic core is a multi-module ontology, shifting assessment from static elements to context-aware construction processes. An automated pipeline transforms OpenBIM data into a carbon knowledge graph, capturing permanent and temporary emissions. A federated neuro-symbolic multi-agent system, combining LLM-based probabilistic reasoning with deterministic ontology constraints, enables natural language querying and scenario simulation for interactive decision support.

Introduction

As urbanization accelerates, metro systems are pivotal for sustainable mobility. In China, as of March 2025, 304 metro lines span 6,530 stations, with 157 lines under construction (He et al. 2025). While operation offers environmental benefits, construction is energy-intensive: a kilometer of open-cut metro emits 39,000–84,000 t CO_2e (Wang et al. 2023; Gan et al. 2024; Zhang et al. 2024). This intensity is comparable to constructing approximately 46 kilometers of highway paradigms y(Chen et al. 2023a). Therefore, calculating and reducing greenhouse gas emissions during the materialization stage is a critical priority for industry (Geng et al. 2022; He et al. 2025).

Early-stage design decisions lock in >80% of emission reduction potential (Gao et al. 2024b), but comprehensive LCA requires detailed data only available later (Roberts et al. 2020), leaving a gap between decision needs and data availability (Hollberg et al. 2020). BIM-LCA integration automates quantity extraction and scenario simulation (Gao et al. 2024a; Kim et al. 2024), yet remains ill-equipped for underground metro complexities (Li et al. 2024).

Current methods face three technical limitations for metro stations. First, the element-centric paradigm omits temporary works (excavation support, dewatering, earthwork logistics) rarely modeled in IFC files, causing significant underreporting. Second, tools lack domain-aware reasoning; early-stage low-LOD models cannot infer implicit construction sequences or backfill requirements. Third, semantic disconnect between design models and environmental databases forces error-prone manual mapping (Cavalliere et al. 2019; Abbasi and Noorzai 2021; Cheng et al. 2022).

To overcome these, we propose an ontology-driven OpenBIM framework that shifts from data extraction to knowledge-enabled inference. It harmonizes structural, enclosure, and geological data via formal logic rules that deduce implicit construction processes from explicit design geometry (e.g., inferring excavation zones from foundation pit design). The framework is validated on six open-cut metro stations with three contributions: (1) the Metro Station Carbon Assessment Ontology (MSCA-Onto) decoupling geometry from construction logic; (2) an automated inference mechanism reconstructing implicit work quantities from sparse geometry; (3) a neuro-symbolic cognitive interface combining LLM reasoning with deterministic ontological constraints, turning the tool into an interactive engineering copilot.

Methodology

Overview

Conventional BIM-LCA methods are constrained by an element-centric paradigm, relying on static bills of quantities from explicit geometry. This causes three shortcomings: (1) data incompleteness – omission of temporary works (formwork, excavation support, dewatering); (2) semantic interoperability barriers – inconsistent IFC naming conflicts with LCA databases, requiring manual mapping; (3) static post-hoc reports – no interactive “what-if” simulation. To address these, we propose a dynamic ontology-driven framework with three integrated layers, each targeting one gap.

The Semantic Foundation Layer (Layer I) redefines the underlying assessment logic to resolve semantic ambiguities. Instantiated by the Metro Station Carbon Assessment Ontology (MSCA-Onto) and based on a System-Element-Process-Resource-Carbon (S-E-P-R-C) architecture, this layer shifts the focus from static elements to context-aware construction processes. By establishing a unified vocabulary, it bridges the semantic gap between heterogeneous design data and environmental databases. Building upon this foundation, The Automated Computational Layer (Layer II) operationalizes the ontology to quantify emissions comprehensively. It transforms OpenBIM data into a computable Carbon Knowledge Graph (CKG) by automating semantic alignment and geometric enrichment. Crucially, this layer utilizes ontological inference rules to deduce the consumption of unmodeled temporary works from permanent structures, thereby ensuring a complete system boundary and correcting the underreporting inherent in conventional methods.

While the first two layers establish a rigorous and complete data repository, a cognitive interaction layer (Layer III) is implemented to bridge the final cognitive gap between structured data and human decision-making. Designed as a federated neuro-symbolic Multi-Agent System (MAS), this layer synergizes the probabilistic reasoning of Large Language Models (LLMs) with the deterministic logic of the ontology. It transforms the CKG from a passive database into an active decision-support engine, enabling stakeholders to synthesize optimization strategies and perform various scenario analysis through natural language interaction. Collectively, this framework not only ensures accurate quantification during the design phase but also establishes a scalable, semantic continuum for lifecycle carbon management.

Metro station carbon assessment ontology

The MSCA-Onto serves as the semantic foundation of the proposed framework, engineered to facilitate a paradigm shift from static, element-centric accounting to dynamic, process-aware carbon assessment. The development of this ontology follows a hybrid methodology that integrates the structural rigor of METHONTOLOGY with the lifecycle flexibility of the NeON framework. Spanning six stages from specification to evaluation, the process involved a comprehensive synthesis of construction standards, IFC schemas, and project documentation. To ensure high interoperability, ontology strategically reuses established schemas such as QUDT for units and PROV-O for provenance, ultimately formalizing domain knowledge into OWL 2 DL expressions supplemented by SWRL rules for advanced automated reasoning.

The structural core of MSCA-Onto is the S-E-P-R-C framework, which establishes a unified semantic layer to decouple static geometric data from dynamic resource consumption. Within this architecture, the System and Element modules decompose the metro station into functionally distinct engineering systems and specialized subtypes, such as diaphragm walls and bracing, effectively bridging the semantic gap between generic IFC schemas and underground engineering practice. The Process module represents the framework's primary innovation by utilizing WorkPackageTemplate entities as semantic adapters. These templates function as inference engines that automatically deduce implicit construction activities, such as formwork or excavation, from sparse BIM geometry. This process-based logic subsequently connects to the Resource and Carbon modules, which map activities to material descriptors and emission factors from external databases to enable multi-level carbon aggregation across lifecycle stages A1 to A5. The overall architecture is illustrated in Figure 1. Six competence questions validated MSCA-Onto's ability to quantify material consumption, evaluate low-carbon substitution, analyze temporary structure recycling, identify carbon hotspots, and assess logistics footprints, proving its utility for early-stage carbon management.

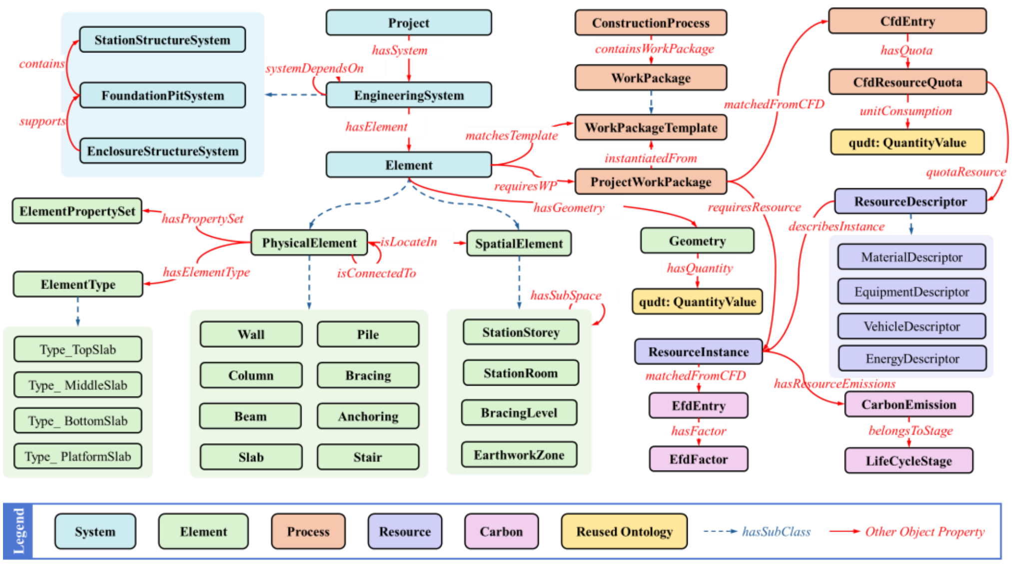

Figure 1: Overall architecture of MSCA-Onto

Table 1 summarizes the six competence questions used to validate MSCA-Onto, along with the SPARQL query patterns and expected outcomes. For instance, CQ5 identifies carbon hotspots via a grouping and ranking query; the results across six stations consistently ranked diaphragm walls and slabs as the top two contributors (42% and 28% of structural carbon, respectively). CQ6 quantifies logistics emissions by joining material supply distances with transport emission factors, yielding a contribution of 8–12% of A1–A3 emissions. The full executable SPARQL queries and result datasets are provided in the supplementary material.

Table 1 Ontology validation competency questions

Competence Question	Validation Objective	SPARQL Pattern
CQ1: What are the material consumptions and total carbon emissions generated by waterproofing engineering in the main structure?	Material Consumption and Carbon Quantification	Process→Resource→Carbon chain with material filtering
CQ2: How much carbon reduction can be achieved by substituting conventional concrete with GGBS-based low-carbon concrete in primary structural elements?	Low-carbon material decision support	Comparative analysis with emission factor lookup from EFD
CQ3: What is the carbon impact of temporary support system demolition activities, including recycling credits?	Temporary structure lifecycle assessment	Temporal process analysis with negative emission accounting
CQ4: What is the carbon emission comparison between vertical retaining systems and horizontal bracing systems in the enclosure structure?	Comparative System Analysis	Cross-system comparison with intensity metrics calculation
CQ5: What are the top 10 carbon hotspots among structural components, and what is their carbon intensity ranking?	Component-level carbon hotspot identification	Multi-criteria ranking with statistical aggregation
CQ6: What is the carbon contribution of material transportation in construction logistics?	Construction logistics carbon analysis	Distance-weight-emission calculation per transport mode

Automated carbon assessment pipeline based on OpenBIM

Based on the MSCA-Onto, this study establishes a systematic automated pipeline designed to transform heterogeneous OpenBIM data into a comprehensive Carbon Knowledge Graph (CKG) suitable for high-precision embodied carbon assessment. The workflow begins by anchoring the assessment in two hierarchically structured reference databases: a construction process database (CPD) for resource inventory and an emission factor database (EFD) for environmental impact, which are strategically decoupled to ensure data flexibility and fitness for purpose. To resolve the semantic ambiguity inherent in raw IFC data, the pipeline employs a hybrid topological reasoning engine. By utilizing spatial indexing and geometric algorithms, such as ray casting and collision detection, the system infers critical spatial relationships to accurately classify elements based on their functional roles, such as distinguishing a generic slab from a specific station platform. For instance, an unclassified IfcSlab instance is resolved as a "Platform Slab" via three topological checks: normal vector orientation (ray casting), vertical position (R-tree query), and adjacency to passenger spaces (collision detection). Simultaneously, to address the omission of temporary earthworks in standard BIM, the module executes a rigorous geometric reconstruction of the foundation pit. Using a "profile-sweep-trim" operation, the system synthesizes horizontal enclosure envelopes with vertical structural surfaces to generate precise volumetric domains for excavation and backfill, which are then intersected with geological layers for quantification.

The pipeline culminates in an automated element-process-resource mapping mechanism that operationalizes the ontology's inference logic. Through domain-specific SWRL rules, semantically enriched elements are matched with WorkPackageTemplate adapters. For example, the following SWRL rule triggers an excavation support work package when the depth of a foundation pit exceeds 5 meters:

```
StationPit(?pit) ∧ hasDepth(?pit, ?depth) ∧ swrlb:greaterThan(?depth, 5.0) → requiresWorkPackage(?pit, ExcavationSupportWorkPackage)
```

Figure 2: An example of the SWRL rule

These adapters dynamically retrieve construction methods from the CPD and associate emission factors from the EFD, instantiating ProjectWorkPackage and ResourceInstance entities. This enables bottom-up aggregation of carbon emissions across stages A1–A5 (material production, transport, on-site construction), ensuring full traceability from individual resources to the project footprint. The complete workflow is shown in Figure 3.

A Multi-agent system for interactive carbon intelligence

To bridge the critical cognitive gap between the static Carbon Knowledge Graph (CKG) and the dynamic, unstructured nature of engineering decision-making, this study proposes a federated neuro-symbolic Multi-Agent System (MAS). Designed around a rigorous divergence - convergence workflow, the MAS strictly decouples cognitive planning from domain-specific execution across three logical layers: cognitive planning, semantic execution, and decision synthesis. This hierarchical architecture ensures that the probabilistic reasoning of Large Language Models (LLMs) is constrained by the deterministic logic of the MSCA-Onto, enabling reliable support for information retrieval, knowledge correction, and strategic analysis entirely through natural language interaction.

The workflow is governed by the Cognitive Planning Layer, where the Orchestrator Agent (A_{orch}) functions as the central reasoning engine. Upon receiving a user query, A_{orch} utilizes a fine-tuned LLM for intent classification and employs Chain-of-Thought (CoT) reasoning to decompose the request into a machine-readable Task Dependency Graph (TDG). To constrain the LLM's probabilistic reasoning without exceeding context window limits, the system does not parse the entire OWL ontology as a raw prompt. Instead, a compressed schema dictionary—containing only the permitted S-E-P-R-C core classes, taxonomic relations, and operational properties—is dynamically injected into the LLM's system prompt to grant it structural awareness. Meanwhile, strict semantic verification (e.g., SWRL rule checking) is deferred to the downstream symbolic agents.

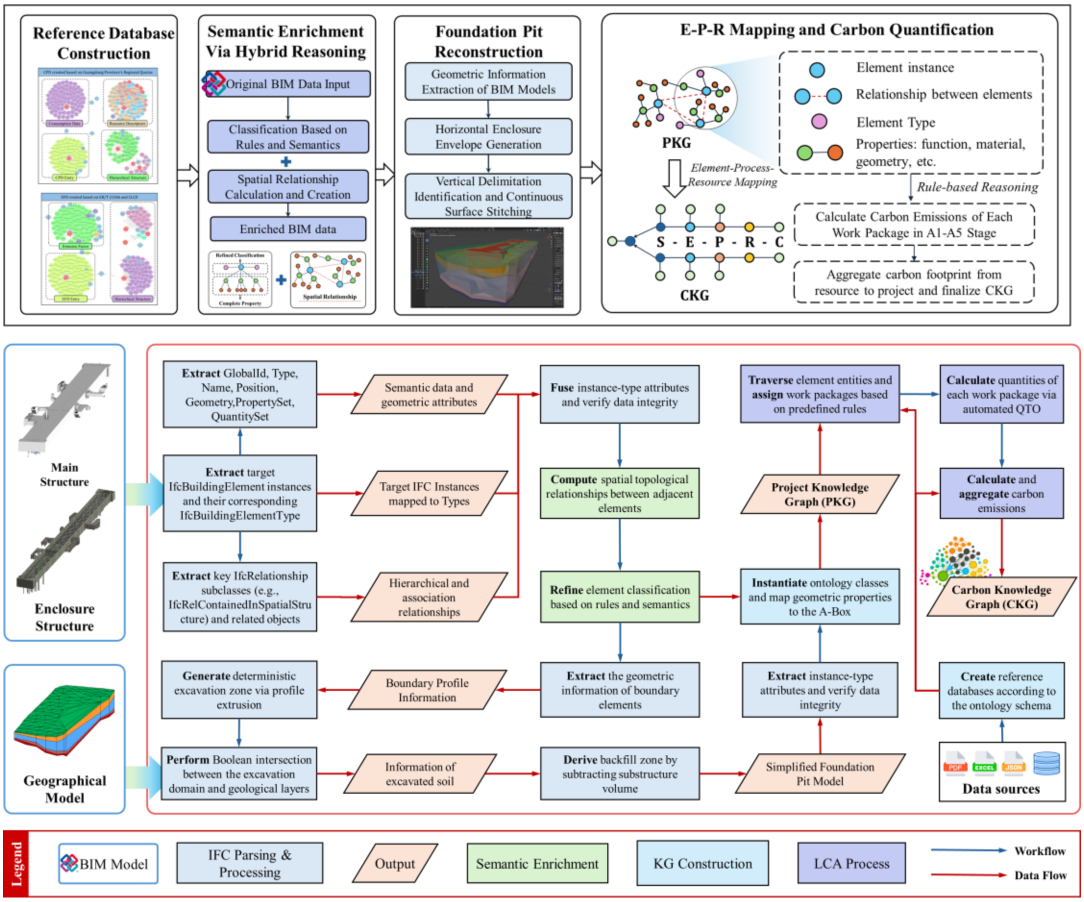

Figure 3: Workflow of automated carbon assessment pipeline

The LLM performs orchestration only: it maps natural language to a predefined set of task templates and produces a TDG. It does not generate SPARQL or numerical parameters directly. For example, given "Replace C30 with C40 concrete in the bottom slab", CoT produces: (1) Intent: Strategic Analysis; (2) Entity: bottom slab → Type_BottomSlab; (3) Parameter: C30→ C40; (4) TDG: [Extraction→Reader→Calculation]. All formal queries and calculations are executed by symbolic agents using deterministic templates. The LLM's reasoning is confined to intent classification, ambiguity detection, and workflow planning. To optimize efficiency, the system implements a hybrid shared-state architecture: control signals are dispatched asynchronously to trigger downstream agents, while data artifacts are managed via a global "blackboard" pattern. This ensures that agents read from and write to a validated, synchronized context, minimizing communication overhead while maintaining strict data consistency.

The Semantic Execution Layer operationalizes these tasks through a suite of functionally atomized agents. The Metadata Agent ($\boldsymbol{A}_{\text{meta}}$) serves as the semantic bridge, parsing complex queries into atomic propositions and mapping ambiguous terms to precise ontology entities via a "decomposition-retrieval-verification" pipeline. For hypothetical scenarios, the Extraction Agent ($\boldsymbol{A}_{\text{extr}}$) translates user intents into a "delta graph specification" ($\boldsymbol{G}_{\text{spec}}$), a provenance annotated artifact that prioritizes user-defined parameters over system defaults while filtering out invalid structural alterations. Simultaneously, the Unified Reader Agent ($\boldsymbol{A}_{\text{read}}$) executes a dual-path retrieval strategy-using deterministic SPARQL for project data and hierarchical semantic search for external reference databases-to construct a composite data context ($\boldsymbol{D}_{\text{comp}}$). These inputs converge at the Calculation Agent ($\boldsymbol{A}_{\text{calc}}$), which performs a sandboxed semantic overlay operation $R = f_{\text{template}}\left(D_{\text{comp}} \oplus G_{\text{spec}}\right)$. By dynamically superimposing user constraints onto the baseline context, this agent ensures that all quantitative derivations are mathematically rigorous and fully traceable.

Finally, the Decision Synthesis Layer ($\boldsymbol{A}_{\text{syn}}$) acts as the system's executive interface and safety valve. It adaptively formats heterogeneous outputs-ranging from visualization of carbon hotspots to comparative scenario matrices-into coherent, human-readable responses enriched with provenance annotations. Crucially, for tasks involving persistent modifications to the CKG, $\boldsymbol{A}_{\text{syn}}$ enforces a strict "human-in-the-loop" protocol. Rather than executing updates immediately, it generates a comprehensive impact assessment forecasting the cascading effects of the proposed changes. The system then pauses to solicit explicit user adjudication, executing the final SPARQL UPDATE statements only after receiving authorization, thereby guaranteeing that stakeholders retain ultimate sovereignty over data integrity. The structure of MAS is shown in Figure 4.

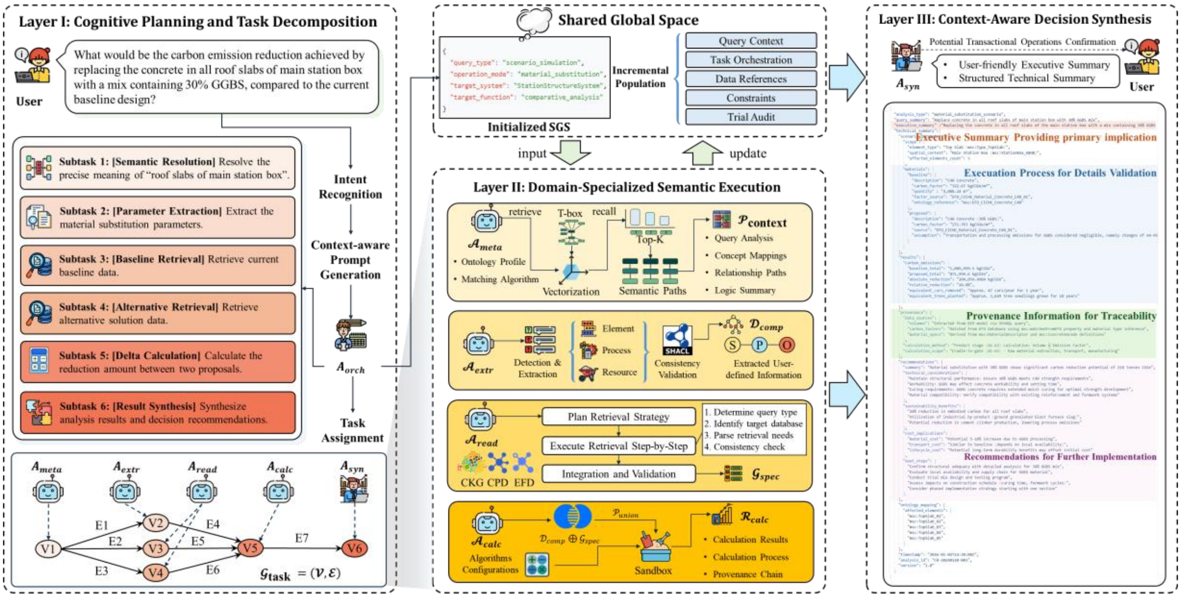

Figure 4: The workflow of MAS proposed

Implementation and results

Evaluation of the automated assessment pipeline

The empirical validity of the proposed framework was established using a diverse dataset comprising six open-cut metro station projects located in Guangdong Province, China. These projects represent a comprehensive range of structural typologies. The excavation depths vary significantly from 16.5 meters to 22.3 meters. This diversity ensures that the validation encompasses different geological conditions and construction complexities.

The input data for each project consisted of the original Industry Foundation Classes design models and the corresponding structural drawings. To rigorously test the scalability of the automated pipeline, we maintained the BIM models at a high level of fidelity. The dataset contains a total of over 44,000 individual geometric components. All three methods, including Manual Calculation (Baseline), Typical BIM-LCA Software, and the Proposed Framework, operated on these identical IFC datasets. They also referenced the exact same data infrastructure for construction processes and emission factors. This rigorous control ensures that any observed performance variance is attributable solely to methodological differences rather than data discrepancies. The CPD was populated using Guangdong Province's municipal construction quotas, covering 124 construction activities (e.g., diaphragm wall excavation, concrete pumping, formwork erection). The EFD (Emission Factor Database) combines Ecoinvent 3.8 for background processes (e.g., cement production, diesel combustion) and region-specific factors from the China Life Cycle Database (CLCD) for materials like steel and concrete. Transportation distances were set to default regional averages (50 km for concrete, 200 km for steel) unless specified otherwise in the BIM properties.

To rigorously quantify the performance of the ontology-driven pipeline, the evaluation was structured around three key performance indicators designed to assess mathematical precision, statistical correlation, and operational efficiency. The first metric, mean absolute percentage error (MAPE), measures the average magnitude of error between the automated results C_{auto} and the manual ground-truth C_{man} across the n test projects, calculated as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{C_{man,i} - C_{auto,i}}{C_{man,i}}\right| \quad (1)$$

This provides a normalized assessment of accuracy where lower values signify higher alignment with expert standards. The second metric is the coefficient of determination (R^2), which evaluates the proportion of variance in the manual baseline that is predictable from the automated pipeline. Finally, temporal efficiency is assessed through the total end-to-end processing time required for a complete assessment.

To rigorously quantify the reliability of these generated profiles, we compared the automated results against the expert-verified manual baseline. We employed Mean Absolute Percentage Error as the primary accuracy metric. This measures the average magnitude of deviation between the automated results and the ground truth. Additionally, we used the coefficient of determination to evaluate linear consistency. The automated pipeline demonstrated exceptional alignment with the manual baseline. It achieved a project-level MAPE of 2.8% and an R^2 value of 0.983. This confirms that the rule-based, ontology-driven workflow can reliably replicate complex engineering logic without the significant deviations observed in standard BIM-LCA plugins. Standard tools often suffer from errors exceeding 23% due to their inability to account for the unmodeled auxiliary processes described above. The detailed accuracy comparison is summarized in Table 2.

Table 2: Accuracy comparison of carbon assessment methods

Metric	Manual Calculation (Baseline)	Typical BIM-LCA method	Ours
Project-Level Accuracy (MAPE)	N/A	23%	2.8%
Component-Level Accuracy (MAPE)	N/A	12%	5.5%
Correlation (R^2)	1.00	0.81	0.983

In terms of operational efficiency, the framework demonstrated a transformative reduction in assessment latency. The manual calculation process required an average of 7.5 person-hours per project. This duration was dominated by tedious quantity take-off and Excel-based mapping. In contrast, the automated pipeline completed the entire end-to-end assessment in an average of 22 minutes.

To understand the computational cost, we decomposed this 22-minute duration into its constituent phases. The semantic enrichment and reasoning phase consumed most of the time, taking approximately 12 minutes. This reflects the computational intensity of the topological algorithms used to resolve element relationships such as containment and adjacency checks. The geometric reconstruction of the braced foundation pit required an average of 8 minutes. This represents a necessary trade-off to capture the implicit earthwork quantities accurately.

The final mapping and calculation phase was highly efficient and completed in under 2 minutes. Although the Typical BIM-LCA method was faster at approximately 5 minutes, this speed resulted in the significant accuracy penalty noted in the previous section. Consequently, the proposed framework offers an optimal balance. It delivers professional-grade accuracy with industrial-scale efficiency that is approximately 20 times faster than manual workflows.

Evaluation of the Neuro-Symbolic MAS

To evaluate the efficacy and safety of the proposed Neuro-Symbolic MAS, a comprehensive performance assessment focusing on its interaction reliability across diverse engineering tasks was conducted. The evaluation utilized a benchmark suite of 120 standardized engineering queries distinctively categorized into three operational modes: information retrieval (IR), knowledge correction (KC), and strategic analysis (SA). To assess the framework's adaptability and model-agnostic robustness, the experiments were replicated using three distinct LLMs as the reasoning backbone: DeepSeek-V3, Gemini 3 Pro, and Qwen 2.5-14B.

To strictly quantify the system's engineering capabilities, we introduced a "trap mechanism" within the benchmark suite. Approximately 15% of the queries were intentionally designed with missing parameters or ambiguous intents, such as requesting a material update without specifying the target component. For these queries, a correct outcome is defined not as executing an operation, but as halting execution and requesting clarification. This tests the system's meta-cognitive ability to detect ambiguity, which is a critical safety feature in engineering.

Accordingly, the system's performance was measured using a quantitative net reliability score (R). Unlike subjective fluency metrics, this score is strictly grounded in the quantitative accuracy of the output. The metric rewards correct answers and penalize unsafe errors while neutrally treating safe failures where the system correctly refuses to answer due to ambiguity. The metric is formalized as:

$$R = \frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{correct}}^{(i)} - \lambda \cdot I_{\text{unsafe}}^{(i)}\right) \quad (2)$$

where $I_{\text{correct}}^{(i)}$ is a binary indicator set to 1 if the system's output strictly matches the ground truth or correctly identifies a trap. $I_{\text{unsafe}}^{(i)}$ is a penalty indicator set to 1 if the output contains a constraint violation, such as returning a physically impossible parameter or hallucinating a value for an ambiguous query. The safety penalty coefficient λ is set to 1.0, reflecting the engineering principle that providing erroneous data is as detrimental as the value of a correct insight. To ensure statistical robustness, the benchmark suite was executed three times for each configuration resulting in a total of 360 independent trials.

The experimental results, summarized in Table 3, demonstrate that the proposed neuro-symbolic architecture effectively standardizes performance output enabling even smaller models to achieve engineering-safe reliability levels.

Table 3: Comparative Performance of LLM Backbones within the MAS Framework

Operational Mode	Metric	Qwen 2.5-14B	DeepSeek-V3.2	Gemini 3-Pro
Information Retrieval	I_{correct}	68.2%	85.3%	92.1%
	I_{unsafe}	3.0%	2.0%	0.0%
	R	0.65	0.83	0.92
Knowledge Correction	I_{correct}	58.0%	82.5%	91.3%
	I_{unsafe}	6.0%	3.5%	0.0%
	R	0.52	0.79	0.91
Strategic Analysis	I_{correct}	50.0%	66.6%	86.4%
	I_{unsafe}	9.5%	4.0%	3.0%
	R	0.41	0.63	0.83
Global Performance	Avg. R Score	0.53	0.67	0.86
	Avg. Latency	18.5s	28.8s	34.2s

Using DeepSeek-V3 as the backbone, the system achieved a robust global R score of 0.91. It exhibited exceptional performance in Strategic Analysis tasks with a success rate of 92.0%, successfully decomposing complex "what-if" scenarios into valid delta specifications. The constraint violation rate was kept extremely low at 1.8%, proving that the orchestrator's planning logic effectively restrains the model's generation within ontological bounds.

The integration of Gemini 3 Pro yielded the highest reliability with a global R score of 0.86. Its superior reasoning capabilities minimized the errors in complex multi-hop queries. Notably, in KC tasks, it achieved a 100% success rate demonstrating perfect alignment with the feasibility protocol when parsing complex user directives. The absence of unsafe errors across all trials highlights the model's rigorous adherence to the provided prompts and its ability to correctly identify all trap questions.

Crucially, the lightweight Qwen 2.5-14B model demonstrated the structural value of the framework. While its raw reasoning capability is lower than the flagships resulting in a lower raw success rate, the MAS architecture successfully mitigated the risks. Instead of generating plausible hallucinations for ambiguous queries, the model frequently defaulted to "safe failures" when uncertain. Although its global R score (0.67) is lower, the constraint violation rate was contained at 3.5%. This is significantly better than the typical instability of unconstrained small models confirming that the framework functions as a cognitive guardrail that makes lightweight models viable for non-critical engineering tasks.

Discussion

The empirical results show that the element-centric BIM-LCA paradigm is insufficient for underground infrastructure. About 12.4% of total embodied carbon comes from unmodeled temporary work, revealing systematic scope leakage and confirming that geometric explicitness is a poor proxy for construction reality. Our ontology-driven pipeline bridges this gap using the S-E-P-R-C model as an active inference engine. The 2.8% error rate proves that semantic reconstruction is necessary for professional-grade accuracy, suggesting that future standards should mandate procedural logic beyond geometric LoD.

A key architectural insight from the MAS evaluation is that the reliability of generative AI in engineering depends more on system architecture than on model size. Lightweight local models achieved operational safety when constrained by the framework, implying that highly reliable, privacy-compliant tools do not require expensive cloud models. The neuro-symbolic architecture decouples probabilistic intent understanding from deterministic execution and defaults to safe failures on ambiguous queries, defining trustworthiness as the ability to recognize knowledge boundaries.

The framework reduced assessment latency twenty-fold (from 7.5 hours to 22 minutes), shifting engineers' cognitive load from tedious quantity take-off to high-value strategic analysis. This enables real-time carbon budgeting alongside cost and schedule during design. Accuracy depends on database granularity; missing methods fall back to generic averages. Geometric reconstruction is optimized for cut-and-cover stations and may need adaptation for tunnels or complex nodes. The human-in-the-loop step ensures safety but adds latency. The S-E-P-R-C architecture is domain-agnostic: extending to tunnels or bridges only requires domain-specific work package templates and geometric threshold rules. The manual baseline may contain expert bias; future work will validate it with real-world construction data.

Conclusion

This research addressed the systemic limitations of conventional BIM-LCA for underground infrastructure—specifically the omission of auxiliary processes and the lack of interactive decision support. A dynamic assessment framework based on a neuro-symbolic MAS was proposed and validated, marking a paradigm shift from static element-based carbon reporting to interactive, process-aware engineering intelligence.

The primary contribution is a robust computational methodology that harmonizes sparse design data with deep engineering knowledge. The MSCA-Onto, built on the S-E-P-R-C model, decouples geometry from resource consumption, enabling dynamic inference of typically invisible work packages (e.g., excavation support). The automated pipeline transforms OpenBIM data into a computable CKG, capturing the full life-cycle inventory with near-expert precision. Most significantly, integrating

LLMs with deterministic ontologies provides a viable pathway for reliable engineering AI. The federated MAS acts as a cognitive copilot, interpreting complex intents, executing rigorous simulations, and enforcing safety via safe-failure mechanisms. Empirical results show a 12.4% completeness improvement and high reliability across diverse backbones, validating the neuro-symbolic approach as a scalable industry solution.

Future work will focus on agent learning from user corrections and integrating real-time schedule data for dynamic carbon forecasting. This framework lays the foundation for intelligent, holistic carbon management, moving sustainability assessment into an integral, effortless part of infrastructure design.

Reference

Abbasi, Saman, and Esmatullah Noorzai. 2021. “The BIM-Based Multi-Optimization Approach in Order to Determine the Trade-off between Embodied and Operation Energy Focused on Renewable Energy Use.” Journal of Cleaner Production 281:125359. doi:10.1016/j.jclepro.2020.125359.

Cavalliere, Carmine, Guillaume Habert, Guido Raffaele Dell’Osso, and Alexander Hollberg. 2019. “Continuous BIM-Based Assessment of Embodied Environmental Impacts throughout the Design Process.” Journal of Cleaner Production 211:941–52. doi:10.1016/j.jclepro.2018.11.247.

Chen, Kunyang, Xiangsheng Chen, Lei Wang, Wensheng Yang, Tong Qiu, Dong Su, and Huanyu Wu. 2023. “Low-Carbon Effects of Constructing a Prefabricated Subway Station with Temporary Internal Supports: An Innovative Case of Shenzhen, China.” Journal of Cleaner Production 426:139023. doi:10.1016/j.jclepro.2023.139023.

Cheng, Baoquan, Kun Lu, Jianchang Li, Huihua Chen, Xiaowei Luo, and Muhammad Shafique. 2022. “Comprehensive Assessment of Embodied Environmental Impacts of Buildings Using Normalized Environmental Impact Factors.” Journal of Cleaner Production 334:130083. doi:10.1016/j.jclepro.2021.130083.

Gan, Mi, Dandan Li, Zhu Yao, Hao Yu, and Qichen Ou. 2024. “Intelligent Decision Modeling for Optimizing Railway Cold Chain Service Networks under Uncertainty.” Information Sciences 679:121112. doi:10.1016/j.ins.2024.121112.

Gao, Hui, Donglin Wang, Xiaoxu Du, and Zhongwei Zhao. 2024. “An LCA-BIM Integrated Model for Carbon-Emission Calculation of Prefabricated Buildings.” Renewable and Sustainable Energy Reviews 203:114775. doi:10.1016/j.rser.2024.114775.

Gao, Yu, Jiayuan Wang, and Tak Wing Yiu. 2024. “Multi-Information Integration-Based Life Cycle Analysis of Greenhouse Gas Emissions for Prefabricated Construction: A Case Study of Shenzhen.” Environmental Impact Assessment Review 104:107330. doi:10.1016/j.eiar.2023.107330.

Geng, Jingjing, Jiajia Wang, Jianguang Huang, Ding Zhou, Jing Bai, Jiayuan Wang, Hui Zhang, Huabo Duan, and Wenbo Zhang. 2022. “Quantification of the Carbon Emission of Urban Residential Buildings: The Case of the Greater Bay Area Cities in China.” Environmental Impact Assessment Review 95:106775. doi:10.1016/j.eiar.2022.106775.

He, Qiufeng, Xiangsheng Chen, Tong Qiu, Kunyang Chen, and Aidong Li. 2025. “GHG Emission Efficiency of Prefabricated Composite Subway Stations: A Novel Case Study of Shenzhen, China.” Environmental Impact Assessment Review 110:107664. doi:10.1016/j.eiar.2024.107664.

Hollberg, Alexander, Gianluca Genova, and Guillaume Habert. 2020. “Evaluation of BIM-Based LCA Results for Building Design.” Automation in Construction 109:102972. doi:10.1016/j.autcon.2019.102972.

Kim, Hyunsik, Seungjun Roh, and Jeonghwan Kim. 2024. “Embodied Carbon of BIM Bridge Models According to the Application of Off-Site Prefabrication: Precast Concrete Applied to Superstructure and Substructure.” Developments in the Built Environment 20:100550. doi:10.1016/j.dibe.2024.100550.

Li, Tao, Yi Rui, Hehua Zhu, Linhai Lu, and Xiaojun Li. 2024. “Comprehensive Digital Twin for Infrastructure: A Novel Ontology and Graph-Based Modelling Paradigm.” Advanced Engineering Informatics 62:102747. doi:10.1016/j.aei.2024.102747.

Roberts, Matthew, Stephen Allen, and David Coley. 2020. “Life Cycle Assessment in the Building Design Process – A Systematic Literature Review.” Building and Environment 185:107274. doi:10.1016/j.buildenv.2020.107274.

Wang, Qian, Yun Bai, Yao Chen, Qian Fu, and Tin Kin Ho. 2023. “Optimizing Vertical Alignment of Underground Metro for Energy Saving of Train Operation.” Energy 273:127295. doi:10.1016/j.energy.2023.127295.

Zhang, Huan, Zhuoran Li, Xianwang Fan, Deyi Duan, and Wandong Zheng. 2024. “Air-Conditioning Load Characteristics and Grey Box Predicting Model in Subway Stations.” Journal of Building Engineering 91:109656. doi:10.1016/j.jobe.2024.109656