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

A MULTI-AGENT RAG-ENABLED KNOWLEDGE SYSTEM FOR NATIONAL-SCALE WHOLE-LIFE CARBON ASSESSMENT OF RESIDENTIAL RETROFITS

Sobia Bano[1], Usman Ali[1], Conor Shaw[2], Tobias Maile[3], Fatemeh Golpayegani[4] and James O'Donnell[5]
[1]School of Mechanical and Materials Engineering, University College Dublin, Ireland
[2]Department of Sustainability Science, Lappeenranta-Lahti University of Technology, Finland
[3]Technical University of Applied Sciences Augsburg, Germany
[4]School of Computer Science, University College Dublin, Ireland
[5]School of Engineering, University of Limerick, Limerick, Ireland

Abstract

National residential retrofit is essential for meeting climate targets, making Whole-Life Carbon Assessment (WLCA) critical for balancing operational and embodied emissions. However, fragmented datasets and incomplete material inventories limit its application at scale. This study proposes a multi-agent, Retrieval-Augmented Generation (RAG)–enabled Knowledge-Based System (KBS) for national-scale WLCA of residential retrofit scenarios. Domain-specific LLM agents retrieve, interpret, and explain carbon metrics with full data provenance, improving transparency and reducing manual effort. Case studies on Irish semi-detached dwellings achieve 88% feasibility, 81% nDCG, 70% Precision, and 79% Mean Reciprocal Rank, confirming the system's potential for evidence-based national decarbonisation planning.

Introduction

Global population growth and the associated demand for housing and infrastructure continue to increase the built environment's climate impact. Recent demographic projections estimate the world population will approach ∼9.7–9.8 billion by 2050 (INED, 2024). At the same time, international and regional policy frameworks are tightening requirements on building energy and carbon performance. In the European Union, the recast Energy Performance of Buildings Directive (EPBD) has strengthened the pathway toward a zero-emission building stock and explicitly advances whole-life thinking, including requirements to calculate and disclose life-cycle global warming potential for new buildings according to a phased timeline (European Commission, 2025). These developments increase the need for decision-support tools that quantify carbon impacts consistently and transparently across large residential portfolios.

Residential retrofit programmes are widely recognised as an effective near-term strategy for reducing operational emissions in existing housing stocks, yet they also introduce embodied emissions that can materially influence net carbon outcomes over the assessment horizon. National-scale and case-study analyses consistently identify this operational–embodied trade-off, showing that retrofits optimised for operational performance may increase whole-life impacts when embodied effects are excluded (McGinley et al., 2022). Whole-life carbon assessments across diverse residential contexts further indicate that an exclusive focus on operational reductions can shift emissions toward materials and construction activities as retrofit depth and deployment scale increase (Hegarty and Kinnane, 2023). Simulation-based LCA studies of residential retrofits similarly demonstrate that conclusions regarding optimal retrofit strategies are highly sensitive to methodological assumptions, including system boundaries, replacement cycles, grid decarbonisation pathways, and uncertainty treatment (Aghamolaei et al., 2025). The broader retrofit LCA literature highlights persistent challenges in translating existing guidance into routine practice, with analyst judgement and modelling choices continuing to drive substantial variability in reported outcomes (Hurst and O'Donovan, 2025).

National-scale complexity has been addressed through the use of building archetypes, building-stock models, and parametric simulation pipelines. Reviews indicate that archetype-based approaches enable scalable analysis but remain sensitive to data quality, representativeness, and calibration decisions (Shen et al., 2024). Bottom-up building-stock frameworks capture heterogeneity and support targeted strategies when detailed inputs are available, although fragmented and incomplete datasets remain a major limitation (Penaka et al., 2024). Parametric life-cycle approaches enable systematic exploration of retrofit options under uncertainty, but their application at scale depends on robust inventories, consistent life-cycle boundaries, and transparent assumptions (Arbulu et al., 2024).

Recent advances in large language models (LLMs) and Retrieval-Augmented Generation (RAG) frameworks present an alternative pathway for scaling whole-life carbon assessment while maintaining traceability. RAG approaches ground model outputs in retrieved evidence, reducing hallucination risk (Lewis et al., 2020), while multi-agent RAG architectures decompose complex analytical workflows into specialised agents, improving robustness and interpretability (Xie et al., 2025). These developments are increasingly discussed within sustainability assessment research as mechanisms for improving scalability and reducing expert workload, alongside continued emphasis on transparency, auditability, and governance.

This study introduces a novel national-scale WLCA framework for residential retrofits that integrates parametric simulation with a retrieval-augmented, multi-agent knowledge-based system to address challenges posed by fragmented data, limited transparency, and poor scalability. The core contribution lies in the tight coupling of ISO-consistent operational and embodied carbon accounting with a domain-specific multi-agent RAG architecture, where dedicated Operational, Embodied Carbon, and Whole-Life Carbon Agents retrieve, validate, and synthesize results directly from an authoritative knowledge base, ensuring full numeric grounding and traceable provenance. The framework further advances current practice through an orchestrated interface and interactive dashboard that present complex WLCA results in a clear, explainable form suitable for stakeholders and policy decision-making. This design enables scalable exploration of retrofit pathways, transparent evaluation of operational–embodied trade-offs, and reproducible decision support for residential decarbonization planning at regional and national scales.

The paper follows a structured format, with Section 2 presenting the research methodology. Section 3 examines a comprehensive Irish case study, providing real-world context for the study's findings. Finally, Section 4 concludes the paper and outlines potential avenues for future research.

Methodology

This study proposes a scalable framework for whole-life carbon assessment of national residential building stocks by integrating data-driven archetype modeling, parametric simulation, and standardized life-cycle assessment. Representative dwelling archetypes are constructed from geometric, material, and operational data, with key parameters systematically varied to generate realistic retrofit scenarios and evaluate both operational and embodied carbon impacts. All simulation outputs and life-cycle results are stored in a structured, version-controlled knowledge base that preserves numerical consistency, provenance, and traceability, while a domain-specific, multi-agent retrieval-augmented generation layer enables auditable, constraint-aware querying and transparent trade-off analysis. Together, the framework translates high-dimensional WLCA evidence into explainable, policy-relevant insights and supports robust, reproducible evaluation of residential retrofit strategies at scale.

Data acquisition

The framework establishes an integrated data infrastructure that can be populated from nationally available statistics, open databases, or project-specific surveys. Geometric inputs describe spatial and envelope characteristics such as envelope U-values (W/m^2K), floor area (m^2), glazing ratio (WWR in %), and orientation (θ in $^\circ$), which govern heat transfer and solar gains. Non-geometric data capture operational conditions including occupancy schedules, internal gains, ventilation behaviour, and system efficiencies (Ali et al., 2019).

Embodied-carbon assessment relies on material and product-level information drawn from recognized life-cycle inventories and environmental product declarations. Typical sources include international or national databases such as ICE, ecoinvent, CIBSE TM65, or equivalent regional inventories; however, the framework allows substitution with any dataset providing transparent emission factors and system boundaries (Chartered Institution of Building Services Engineers (CIBSE), 2021; Circular Ecology, 2025; Ecoinvent Association, 2025). Climate data consisting of hourly temperature, solar irradiance, and wind profiles are incorporated to enable climate-sensitive operational modelling.

Archetype development

The methodology employs building archetypes representing dominant residential typologies to enable national-scale analysis without modelling individual dwellings. Each archetype encapsulates typical geometry, construction, and operational characteristics derived from national housing statistics and empirical studies. Baseline specifications include envelope assemblies, material composition, thermal performance (U-values in W/m^2K), building services, and occupant behaviour assumptions across age bands and dwelling forms. These attributes are formalized within the *ArchetypeData* table of the WLCA Knowledge Base System (KBS), where stable identifiers enable consistent linkage to operational simulations, embodied material records, and scenario states across the pipeline (Ali et al., 2019).

Archetypes are implemented in validated dynamic building simulation environments, such as EnergyPlus or IES VE, or equivalent tools capable of hourly energy modelling (Erdogan, 2019). Simulation outputs are benchmarked against national or measured datasets to ensure realistic performance. Once validated, archetypes function as flexible templates for systematic variation of materials, systems, and operating conditions in retrofit and whole-life carbon analyses.

Operational and embodied parametric feature selection

Parametric feature selection defines the multidimensional design space explored within the WLCA framework. Parameters are identified through literature review, data availability, and preliminary sensitivity testing, rather than expert judgement alone. Selected variables demonstrably influence energy demand, thermal behaviour, or material-related impacts and are registered in the WLCA Parametric KBS with metadata specifying domain relevance, allowable ranges, default values, and dataset versions.

Parameters are categorized into geometric, operational, and material groups. Geometric variables influence envelope exposure and solar gains; operational parameters capture use patterns and system performance; material parameters affect heat transfer and embodied carbon intensity. Together, these parameters bound the parametric space feeding simulation, LCA calculation, and retrieval. Centralized storage enables the *Interface Agent*

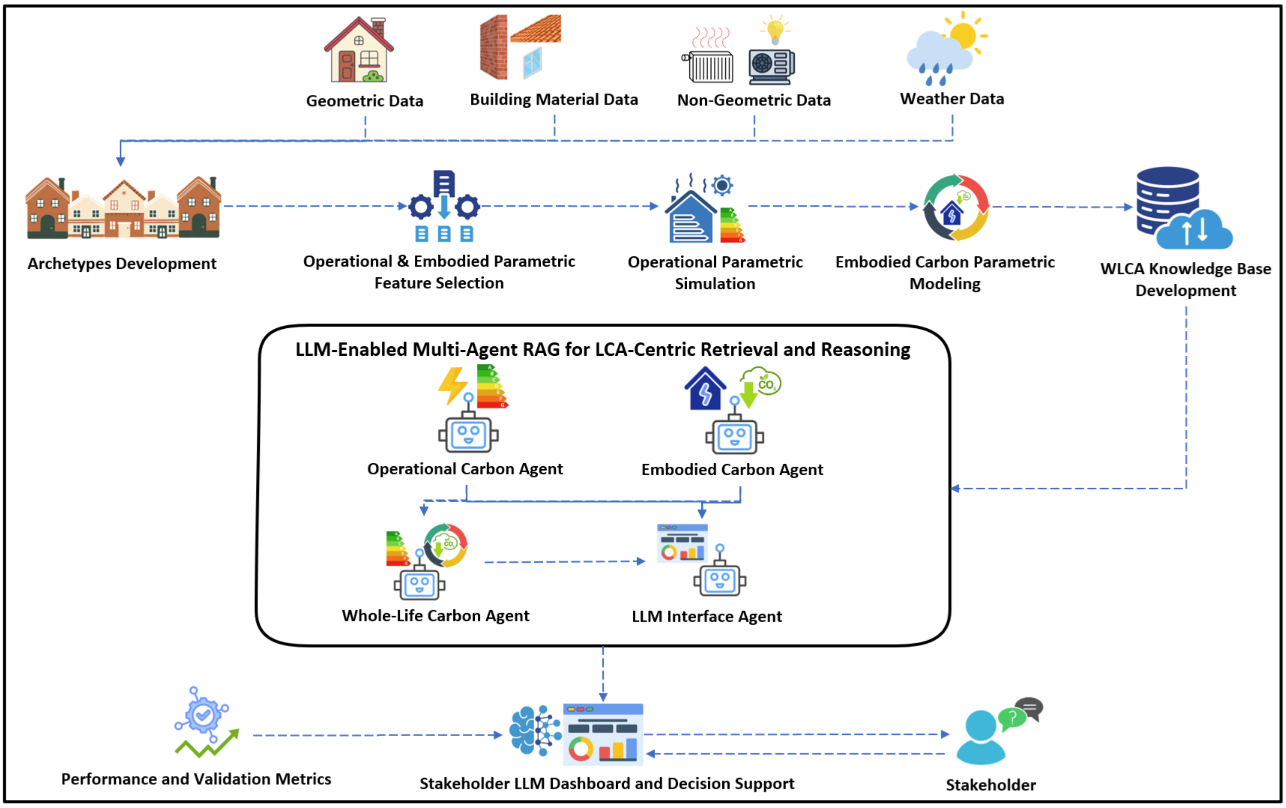

Figure 1: Methodology for Multi-Agent RAG-Enabled Knowledge System for national-Scale Whole-Life Carbon Assessment of Residential Retrofits.

and *Query Interpretation* process to consistently translate natural-language queries into structured retrieval constraints.

Operational parametric simulation

Operational simulation evaluates the effects of variations in form, systems, and building type on energy demand through an ensemble of scenarios generated by perturbing selected parameters. High-throughput execution is achieved using parametric automation tools such as jEPlus, Python-based workflows, or equivalent batch-processing environments (Zhang and Korolija, 2010). Each simulation is associated with a specific archetype identifier, parameter set, climate file, and scenario state, with results stored in the *OperationalData* table of the KBS.

To manage computational cost in high-dimensional spaces, statistical sampling methods such as Simple Random Sampling, Latin Hypercube Sampling, or Sobol sequences are employed. Simulation outputs include annual energy use intensity, heating and cooling loads, system consumption, and comfort metrics. These results form the primary evidence used by the *Operational Agent* within the RAG architecture. Post-processing normalizes outputs and structures them for integration with WLCA calculations, producing a reusable, queryable synthetic dataset that supports consistent reuse without re-running simulations.

Embodied carbon parametric modeling

For each parametric configuration, a cradle-to-grave lifecycle assessment is conducted in accordance with ISO 14040/14044 standards (ISO, 2006). The assessment follows the A–C life-cycle structure, covering product and construction stages (A1–A5), operational energy use (B6), and end-of-life processes (C1–C4). Lifecycle boundaries and assessment horizons are aligned within the WLCA Parametric KBS to ensure consistent integration of embodied and operational impacts by the *Whole-Life Carbon Agent*.

Material emission factors are sourced from established inventories such as ICE, ecoinvent, and CIBSE TM65, or equivalent national datasets. Where project-specific data are unavailable, transparent proxy selection is permitted based on material function and composition, with uncertainty and provenance explicitly recorded. All material quantities, emission factors, and uncertainty annotations are persisted in the *EmbodiedData* table, enabling full auditability during retrieval.

An automated calculation pipeline links simulation outputs, material quantities, and emission factors to compute embodied and operational impacts across large scenario sets. Results are expressed as total life-cycle emissions ($kgCO_2e$) and normalized by floor area to enable cross-scenario comparison. This integrated dataset underpins the knowledge base and is consumed by the *Whole-Life Carbon Agent* to compute total WLCA values, savings, and baseline–post-retrofit comparisons within the auditable RAG pipeline.

WLCA knowledge base development

The Whole-Life Carbon Assessment Knowledge Base System (KBS) functions as the authoritative, version-controlled data backbone of the proposed multi-agent, retrieval-augmented framework. All user queries, agent reasoning, and dashboard outputs resolve directly to records stored in the KBS, which persists not only final whole-life carbon indicators but also the intermediate parameters, scenario states, and modeling artifacts required to reconstruct any reported value. This design ensures that WLCA results remain traceable, auditable, and reproducible across software environments and jurisdictions, while preserving numerical consistency and provenance throughout the assessment workflow.

At the data layer, the KBS is organized into normalized tables that store archetype descriptors in *ArchetypeData*, operational simulation outputs in *OperationalData*, embodied material inventories in *EmbodiedData*, and higher-level assessment parameters in the WLCA Parametric KBS, all indexed by explicit scenario and state identifiers. Baseline and post-retrofit conditions are encoded as distinct scenario states, enabling consistent before–after comparisons and trade-off analysis without duplicating archetype definitions. To support retrieval-augmented generation, the KBS exposes both structured access for deterministic constraint filtering and semantic representations for contextual retrieval, allowing the system to combine semantic relevance with strict numeric grounding and ensuring that all returned WLCA outputs are anchored in authoritative, auditable records.

LLM-enabled multi-agent RAG for LCA-centric retrieval and reasoning

The proposed system implements a domain-specific, multi-agent retrieval-augmented generation (RAG) layer that mediates between a stakeholder-facing dashboard and an underlying WLCA Knowledge Base System (KBS) (Table 1). The architecture is explicitly designed for research-grade life cycle assessment, where numerical correctness, scenario consistency, and full provenance traceability are mandatory. Rather than allowing a large language model to generate estimates directly, the system decomposes query handling, data retrieval, validation, domain reasoning, and output persistence into explicit agents and deterministic process steps.

To support transparency and interpretability, the operation of each architectural component is explained using the same representative retrofit query that is traced end-to-end in Table 1. The example considers a dwelling constructed in 1965, originally heated by a boiler, with the user seeking to upgrade to a heat pump, achieving a BER A2 rating while minimizing embodied carbon impacts.

User Query and Interface Agent The process begins when the user submits a natural-language query via the dashboard, such as: *"I have a house built in 1965 with a boiler and would like to upgrade to A2 rating with a heat pump and lower embodied carbon."* As shown in the first row of Table 1, this raw query is persisted verbatim in the *UserQuery* table, ensuring that the original user intent is preserved for audit and traceability.

The Interface Agent then performs query normalization. Informal expressions such as "house built in 1965," "boiler," and "A2 rating" are mapped to standardized domain concepts, including the 1950–1966 age band, a baseline boiler heating system, and a post-retrofit BER A2 performance target. The resulting normalized query string (Table 1, second row) removes linguistic variability while preserving semantic intent and provides a stable input for deterministic interpretation and retrieval.

Query Interpretation and Constraint Encoding. The normalized query is translated into a structured representation of intent and constraints. As illustrated in Table 1, this step extracts the retrofit action (boiler to heat pump), the target post-retrofit energy rating (BER A2 or better), the optimization objective (minimize embodied carbon), and the building context (constructed in 1965, corresponding to the 1950–1966 age band). These elements are encoded in a *RetrievalConstraints* record, which governs all subsequent retrieval and filtering operations and ensures that the user's intent is expressed in a fully specified and reproducible form.

Semantic Retrieval and Candidate Set Generation. Guided by the encoded constraints, the RAG layer initiates semantic retrieval over the WLCA KBS. This step identifies archetypes, retrofit scenarios, and related records that are contextually similar to the interpreted query. The results are stored as a ranked *CandidateSet* (Table 1), in which multiple mid-century dwelling archetypes with low-carbon retrofit characteristics are initially considered. This stage prioritizes recall and ensures that potentially relevant records are not prematurely excluded.

Structured Filtering and Compatibility Enforcement. The candidate records are then subjected to structured filtering using the explicitly encoded constraints. The candidates are retained only if they support a heat pump retrofit and achieve a post-retrofit BER A2 or better (Table 1). This step enforces methodological and scenario compatibility, eliminating records that are semantically similar but technically or contextually invalid for the query.

Join, Re-ranking, and Retrieved Record Bundling. Filtered records are joined across the *ArchetypeData*, *OperationalData*, *EmbodiedData*, and WLCA parametric tables to align operational simulations, material inventories, and assessment assumptions under a common archetype and scenario definition. The joined records are then re-ranked according to the optimization objective in this example, lowest embodied carbon per square metre, yielding

Table 1: End-to-end processing of a natural language retrofit query using the proposed retrieval-augmented, agent-based whole life carbon assessment framework.

Process Step	Input / Context	Example Output for the Query
1. User Query Capture	Raw user input	*"I have a house built in 1965 with a boiler and would like to upgrade to A2 rating with a heat pump and lower embodied carbon."*
2. Interface Agent (Normalization)	Raw query text	Normalized query string used for interpretation and retrieval: *"house built 1965 boiler upgrade heatpump a2 rating lower embodied carbon"*
3. Query Interpretation	Normalized query	Extracted intent and constraints: – Retrofit from boiler to heat pump – Target post-retrofit energy rating: BER A2 or better – Optimization objective: minimize embodied carbon – Building context: constructed in 1965 (Age Band D: 1950–1966)
4. Semantic Retrieval	Interpreted query and parametric knowledge base	Candidate archetypes ranked by semantic similarity to the query, e.g.: 1) 1950–1966 dwelling with heat pump retrofit 2) Post-war detached dwelling with insulation upgrade 3) Retrofit archetype with low-carbon heating
5. Structured Filtering	Candidate set and extracted constraints	Records filtered to retain only those that: – Achieve BER A2 or better after retrofit – Are compatible with heat pump systems
6. Join and Re-ranking	Filtered candidate set and optimization objective	Remaining records re-ranked by lowest embodied carbon per square metre, yielding a top-K shortlist
7. Grounding and Validation	Top-ranked records	Selection of the best-matching archetype with full provenance: – Document identifier – Dataset version – Fields used for energy and carbon calculations
8. Operational Carbon Agent	Grounded archetype	Before/after operational performance estimates, e.g.: – BER before retrofit: C2 – BER after retrofit: A2 – Primary energy intensity reduced from 210 to 75 $kWh/m^2/yr$
9. Embodied Carbon Agent	Grounded archetype	Embodied carbon breakdown by components (walls, windows, roof, floor, heat pump) and total embodied carbon per m^2
10. Whole Life Carbon Agent	Grounded archetype	Combined 60-year whole life carbon result including operational and embodied emissions, e.g.: – Operational: 540 $kgCO_2e/m^2$ – Embodied: 390 $kgCO_2e/m^2$ – Total WLCA: 930 $kgCO_2e/m^2$
11. Auditable Output Storage	Agent outputs and provenance	Persisted auditable record containing retrieval preview, operational results, embodied results, WLCA results, and dataset version

a bounded Top-K shortlist. This shortlist is materialized as a *RetrievedRecordBundle*, representing the best-supported interpretation of the user query.

Grounding and Validation. Before numerical reasoning occurs, the retrieved records undergo grounding and validation. This step verifies that the selected archetype is fully parameterized, that all required operational and embodied data are present, and that dataset versions and source documents are explicitly identified. Only records that pass these checks are exposed to downstream agents, ensuring that all reported results are fully grounded and reproducible.

Operational Carbon Agent. The Operational Carbon Agent consumes the grounded records to report on pre- and post-operational performance. In the example, this includes changes in BER rating and primary energy intensity, with reductions from a pre-retrofit C2 level to a post-retrofit A2 level, and a corresponding reduction in energy use intensity. All values are linked to underlying simulation outputs stored in the *OperationalData* table.

Embodied Carbon Agent. In parallel, the Embodied Carbon Agent computes embodied emissions associated with the retrofit measures. For the running example, this includes a component-level breakdown of embodied carbon contributions from walls, windows, roof, floor, and the heat pump system itself, aggregated into a total embodied carbon intensity per square metre. Factor sources and dataset versions are explicitly documented.

Whole-Life Carbon Agent. The Whole-Life Carbon Agent integrates operational and embodied results over a 60-year assessment horizon to compute total whole-life carbon impacts. The combined operational and embodied emissions enable direct comparison between baseline and retrofit scenarios and the quantification of net carbon savings (Table 1).

Auditable Output Storage. Finally, all agent outputs and associated provenance metadata are consolidated into a persistent *AuditableOutputs* record. This record includes a retrieval preview, operational results, embodied results, whole-life carbon results, and dataset version identifiers. Persisting these outputs ensures that all reported values can be reproduced by re-running the same pipeline against the same WLCA KBS snapshot.

Stakeholder LLM dashboard and decision support

The stakeholder-facing LLM dashboard is the primary interface to the multi-agent RAG system and is tightly integrated with its architecture. It performs no independent computation, acting instead as a thin interaction layer that submits user queries, presents grounded outputs, and exposes audit information generated by the backend. Each interaction corresponds to a *UserQuery* record and follows the same retrieval, validation, and agent-reasoning workflow.

Using natural language, users can explore archetypes, compare retrofit strategies, and analyze embodied–operational trade-offs across assessment horizons. Queries are routed through the *Interface Agent*, which coordinates retrieval and agent execution and returns structured results with narrative explanations. Quantitative outputs are displayed as tables and charts, while an embedded audit view provides dataset versions, document identifiers, factor sources, and scenario states. When evidence is incomplete, uncertainty flags and provenance gaps are shown instead of unsupported values.

The dashboard supports exporting results in CSV or JSON formats and generating auto-formatted reports with embedded audit metadata. Because outputs are derived from persisted *AuditableOutputs* records, all exported artifacts are reproducible outside the system. Portability across regions and jurisdictions is achieved by substituting local archetypes, material inventories, and climate datasets within the KBS while preserving schemas, retrieval logic, and audit guarantees, allowing the dashboard and methodology to be reimplemented in new contexts without modifying the core pipeline.

Performance and validation metrics

The evaluation framework assesses retrieval quality, answer faithfulness, and numerical correctness within the LCA-centric multi-agent RAG system to ensure transparency and methodological compliance. Retrieval performance is measured using standard information-retrieval metrics, including Recall@k, Precision@k, Mean Reciprocal Rank (MRR), and normalized Discounted Cumulative Gain (nDCG), computed on a labeled query set linking stakeholder questions to authoritative KBS records. High Recall@k and nDCG indicate that relevant WLCA evidence is retrieved and correctly ranked, while MRR captures how quickly usable evidence appears, which is particularly important in multi-agent workflows. In addition, feasibility-oriented checks evaluate whether at least one constraint-satisfying retrofit option is retrieved within top-k results, reflecting the decision-search nature of stakeholder queries (Sharma, 2025).

Faithfulness and numerical correctness are enforced through strict grounding of all reported values to versioned KBS records, with citation coverage, provenance completeness, and contradiction checks used to audit responses. WLCA outputs are further validated against ISO 14040/14044 system boundaries and accounting conventions, with spot checks ensuring that reported whole-life totals are consistent with the sum of operational and embodied components within defined tolerances (ISO, 2006). All evaluations are tied to dataset version tags and audit metadata, enabling reproducibility across regions by substituting local archetypes, material inventories, and climate data while preserving retrieval logic and guardrails.

Case study and results

This case study demonstrates how a proposed whole-life carbon (WLC) methodology can be applied to the Irish residential sector using nationally available datasets, regulatory conventions, and realistic retrofit pathways. Focusing on semi-detached dwellings as a representative housing typology, it shows an end-to-end workflow from stock characterization and archetype definition to dynamic operational energy simulation and embodied-carbon accounting (Table1). The national Energy Performance Certificate (EPC) database underpins the analysis by providing standardized dwelling attributes, which are translated into physically interpretable archetypes aligned with Ireland's Dwelling Energy Assessment Procedure (DEAP). These

Table 2: Summary of dashboard results for retrofit query on Irish semi-detached dwellings (roof built 2000).

Attribute	Result
Query constraints	I have a house built in 1965 with a boiler and would like to upgrade to A2 rating with a heat pump and lower embodied carbon.
BER (Pre → Post)	C2 → A2
Primary EUI	177.1 → 49.7 (kWh/m^2/yr)
WLCA 60y total	1620.8 → 211.6 (kgCO_2e/m^2)
WLCA savings	1409.2 (kgCO_2e/m^2)
Operational carbon 60y	1554.2 → 108.1 (kgCO_2e/m^2)
Embodied carbon 60y	66.6 → 103.5 (kgCO_2e/m^2)
Post-retrofit carbon split	Operational 51% / Embodied 49%

archetypes capture age-dependent construction practices, baseline thermal performance, and window characteristics, forming a consistent basis for simulating pre- and post-retrofit operational energy demand under controlled climate and occupancy assumptions.

Operational and embodied carbon impacts are evaluated through a parametric simulation and whole-life carbon assessment pipeline that systematically explores retrofit scenarios, including envelope upgrades and low-carbon heating systems. Dynamic simulation outputs are combined with Irish and international embodied-carbon factors to account for initial construction, retrofit measures, component replacements, and end-of-life processes over the assessment horizon. Results are stored in a structured knowledge base that supports auditable, reproducible analysis and retrieval-augmented generation, enabling stakeholders to query and compare retrofit options by operational, embodied, or total life-cycle carbon. Together, the case study illustrates how an integrated, data-driven framework can support transparent whole-life carbon assessment and decision-making for residential decarbonization in the Irish context.

The framework enables stakeholders to enter constraint-based queries through the dashboard to analyse existing buildings and evaluate retrofit outcomes. The system identifies a transition from a current BER C2 dwelling to a post-retrofit A2 option with heatpump upgrade, reducing $WLCA_{60y}$ from 1620.8 to 211.6 kgCO_2e/m^2, corresponding to a saving of 1409.2 kgCO_2e/m^2 (approximately 86%). In parallel, Primary EUI decreases from 177.1 to 49.7 kWh/m^2/yr (a reduction of 128 kWh/m^2/yr, or 72%), while the post-retrofit carbon contribution shifts to 51% operational and 49% embodied, demonstrating transparent quantification of operational–embodied trade-offs (Table 2).

The study is evaluated using the multi-axis protocol described in the methodology, combining retrieval effectiveness, decision relevance, numerical faithfulness, and LCA correctness. Assessment is conducted on a stratified 20% sample of the 1.6 million parametric configurations (approximately 320,000 cases) using 30 stakeholder-style queries representing typical retrofit constraints. Each query yields on average 118 feasible designs after structured filtering, confirming that the design space remains highly multi-modal even under BER and envelope requirements, while the ten lowest-WLCA options form the reference set for decision-oriented metrics.

Table 3: Summary of retrieval performance across feasibility, ranking quality, and decision relevance metrics.

Method	Metrics at $K = \{1,3,5,10\}$
Full multi-agent system	Hit: 0.78/0.82/0.85/0.88 nDCG: 0.81/0.76/0.74/0.71 Precision: 0.74/0.72/0.71/0.70 Recall: 0.70/0.71/0.70/0.69 MRR: 0.74/0.76/0.77/0.79
Hybrid candidates	Hit: 0.61/0.69/0.73/0.79 nDCG: 0.69/0.67/0.65/0.63 Precision: 0.58/0.56/0.55/0.53 Recall: 0.57/0.55/0.54/0.52 MRR: 0.58/0.61/0.63/0.66
Semantic-only retrieval	Hit: 0.21/0.33/0.38/0.46 nDCG: 0.41/0.38/0.35/0.32 Precision: 0.24/0.21/0.20/0.19 Recall: 0.23/0.20/0.19/0.18 MRR: 0.24/0.27/0.29/0.31
Random (20% pool)	Hit: 0.10/0.11/0.18/0.32 nDCG: 0.16/0.17/0.17/0.18 Precision: 0.11/0.11/0.10/0.13 Recall: 0.11/0.10/0.10/0.12 MRR: 0.11/0.12/0.14/0.17

Feasibility discovery shows that the multi-agent framework reliably identifies valid retrofit options under strict constraints, achieving HitFeasible@1 of 0.78 and HitFeasible@10 of 0.88, far exceeding semantic-only retrieval (0.46 at $K = 10$) and random sampling (0.32). Hybrid candidate filtering improves feasibility (0.79 at $K = 10$) but remains inferior to the full WLCA-aware reranking, underscoring the value of objective-driven ordering in sparse solution spaces (Table 3).

Decision relevance and ranking quality further demonstrate the system's effectiveness. The framework attains PrecisionTop10@10 and RecallTop10@10 of 0.70, with strong early-rank performance (Precision@1 and MRR@1 of 0.74), indicating that optimal low-carbon designs typically appear at the top of the list. Ranking alignment with ideal WLCA ordering is high, with nDCG@1 of 0.81 and nDCG@10 of 0.71, while hybrid and semantic-only methods show substantially weaker performance.

Faithfulness checks confirm that all reported results remain numerically grounded and auditable, with citation coverage of 0.98, a numeric mismatch rate of 0.01, and universal compliance with WLCA boundary arithmetic. Together, these results demonstrate that the framework can navigate large parametric retrofit spaces while preserving LCA rigor, transparency, and interactive usability, making it suitable for stakeholder-facing decision support and transferable to other national building stocks.

Conclusion and future work

This study presents a scalable framework for estimating whole-life carbon emissions across national residential building stocks by integrating parametric simulation, standardized WLCA, and a multi-agent retrieval layer. Heterogeneous inputs, including archetype definitions, operational simulation outputs, material inventories, and life-cycle calculations are consolidated within a structured knowledge base to support transparent, query-driven analysis. Application to the Irish residential sector demonstrates that the framework can explore more than one million retrofit configurations while preserving numeric grounding, data provenance, and interactive performance, with structured filtering and WLCA-aware re-ranking significantly improving the identification of feasible low-carbon options over semantic retrieval alone.

The framework delivers three key contributions: it unifies operational and embodied carbon assessment within a single parametric pipeline to enable consistent before and after retrofit evaluation; it provides versioned, auditable records that support reproducible national-scale WLCA without embedding assumptions in the language layer; and it enables natural-language access to complex WLCA evidence through multi-agent orchestration with enforced numeric guardrails. Future work will extend the approach to additional building typologies and integrate dynamic occupancy, monitored energy data, and spatial datasets to enhance realism and support neighbourhood- and stock-scale decarbonization planning.

Acknowledgments

This publication has emanated from research supported partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) under the Research, Development and Demonstration Funding Programme Grant No. 22/RDD/822. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI. This research is also part of the IEA EBC - Annex 91 – Open BIM for Energy Efficient Buildings.

References

(ISO 14044:2006). Environmental management: Life cycle assessment framework requirements and guidelines.

Aghamolaei, R. et al. (2025). Integrated life cycle assessment of residential retrofit strategies: Balancing operational and embodied carbon, lessons from an irish housing case study. Sustainability, 17(18):8173. Accessed: 2026-01-08.

Ali, U., Shamsi, M. H., Hoare, C., Mangina, E., and O'Donnell, J. (2019). A data-driven approach for multiscale building archetypes development. Energy and buildings, 202:109364.

Arbulu, M., Oregi, X., Etxepare, L., et al. (2024). Parametric simulation tool for the enviro-economic evaluation of energy renovation strategies in residential buildings with life cycle thinking: Pararenovate-lct. Energy and Buildings, 312:114182.

Chartered Institution of Building Services Engineers (CIBSE) (2021). Embodied carbon in building services: A calculation methodology (tm65). Accessed: 2025-06-26.

Circular Ecology (2025). Embodied carbon footprint database – the ice database. `https://circularecology.com/embodied-carbon-footprint-database.html`. Accessed: 2025-10-23.

Ecoinvent Association (2025). Ecoinvent database. `https://ecoinvent.org/database/`. Accessed: 2025-10-23.

Erdogan, E. (2019). Comparative studies of building performance simulation tools (energyplus, ies-ve, ida-ice).

European Commission (2025). Global warming potential of buildings. Accessed: 2026-01-08.

Hegarty, R. O. and Kinnane, O. (2023). A whole life carbon analysis of the irish residential sector - past, present and future. Energy and Climate Change, 4:100101.

Hurst, L. J. and O'Donovan, T. S. (2025). Defining new approaches for retrofit life cycle analysis to improve design outcomes. Energy and Buildings, page 116844.

INED (2024). 2024: the united nations publishes new world population projections. Accessed: 2026-01-08.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. In Advances in Neural Information Processing Systems (NeurIPS).

McGinley, O., Moran, P., and Goggins, J. (2022). An assessment of the key performance indicators (kpis) of energy efficient retrofits to existing residential buildings. Energies, 15(1):334.

Penaka, S. R., Feng, K., Olofsson, T., Rebbling, A., and Lu, W. (2024). Improved energy retrofit decision making through enhanced bottom-up building stock modelling. Energy and Buildings, 318:114492.

Sharma, C. (2025). Retrieval-augmented generation: A comprehensive survey of architectures, enhancements, and robustness frontiers. arXiv preprint arXiv:2506.00054.

Shen, P. et al. (2024). Archetype building energy modeling approaches and applications: A review. Renewable and Sustainable Energy Reviews, 199:114478.

Xie, Y., Jiang, B., Mallick, T., et al. (2025). Marsha: multi-agent rag system for hazard adaptation. npj Climate Action, 4(1).

Zhang, Y. and Korolija, I. (2010). Performing complex parametric simulations with jeplus. Shanghai Jiao Tong University.