



The 59th Hawaii International Conference on System Sciences (HICSS)

Large Language Models Based In-Context Learning for Early Stage Building Life Cycle Assessment

Jan 7, 2026

Hyeyun Eunice Jung, Muran Yu,
Jie Wang, Kincho H. Law, Michael D. Lepech



1

Introduction

2

Methodology: building LCA dataset and methods

3

Results: LLM method comparative analysis

4

Conclusion

Building Design Space is Complex



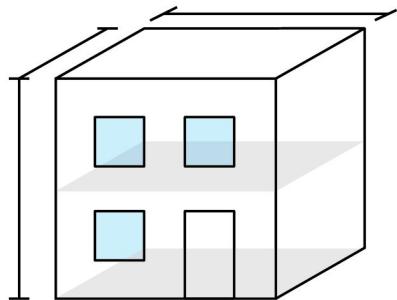
Design variables:

- 1) Number of buildings: 3 or 4
- 2) Orientation: 0-360°
- 3) Number of stories: 5-8
- 4) Footprint
- 5) Window-Wall-Ratio: 15-50%
- 6) Shading presence
- 7) Shading depth
- 8) Materials:
Cladding (7), Substructure (2), Partition (5), Floor finish (8),
Floor structure (12), Columns & beams (10), Windows (5), Wall
structure (6), Wall finishes (2), Roof (10)
- 9) Dimensions:
Cladding (4), Flooring surface (4), Ceiling (4), Wall finishes (4),
Substructure (8)

Design space size: 5.73×10^{23}

Calculating Embodied Carbon in Buildings

Material amount
From Building 3D Model



Global Warming Potential
(GWP) per material
from LCI Database

X

kg CO₂e / unit of material

=

kg CO₂e

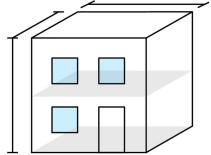
Assumptions

Buildings have thousands of parameters

Facilitating LCA Quantitation using LLMs - Fast and Recursive

Current Practice

Architect



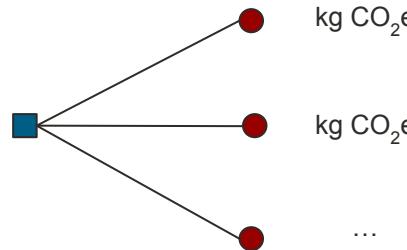
Few design options

Engineer

Several days

LCA tool

(One Click LCA, Tally,
LC3, etc.)



Linear data transmission

Research Goal

Parameters	
Area	m ²
Width	m
...	...

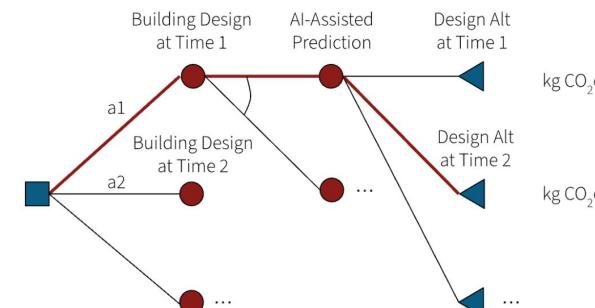
Set of parameters

Architect

Several hours

LLM tool

(ChatGPT, Claude, etc.)



Non-linear and recursive process



1

Introduction

2

Methodology: building LCA dataset and methods

3

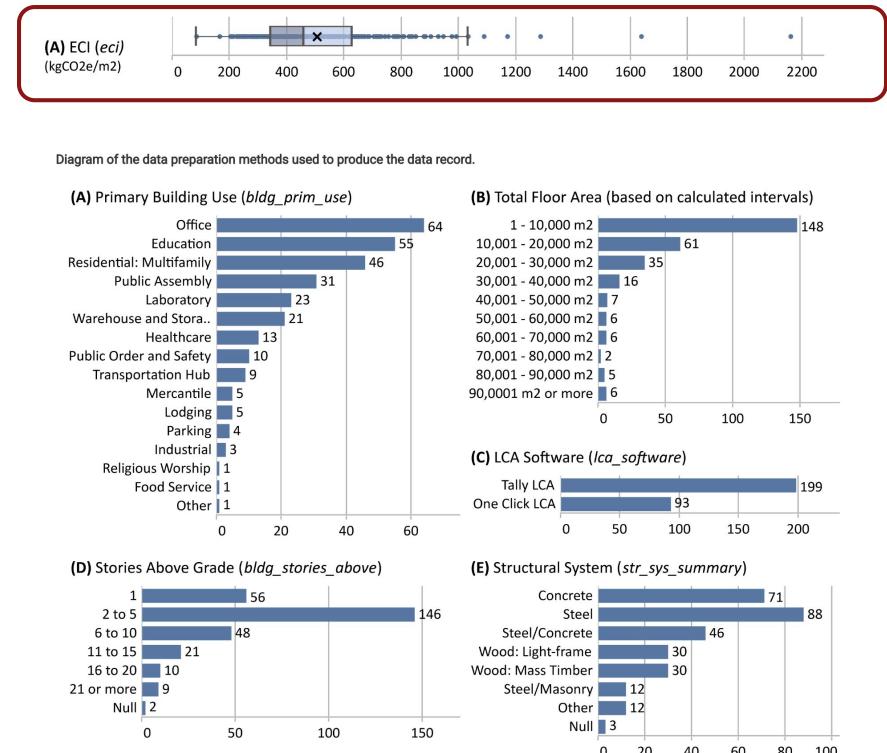
Results: LLM method comparative analysis

4

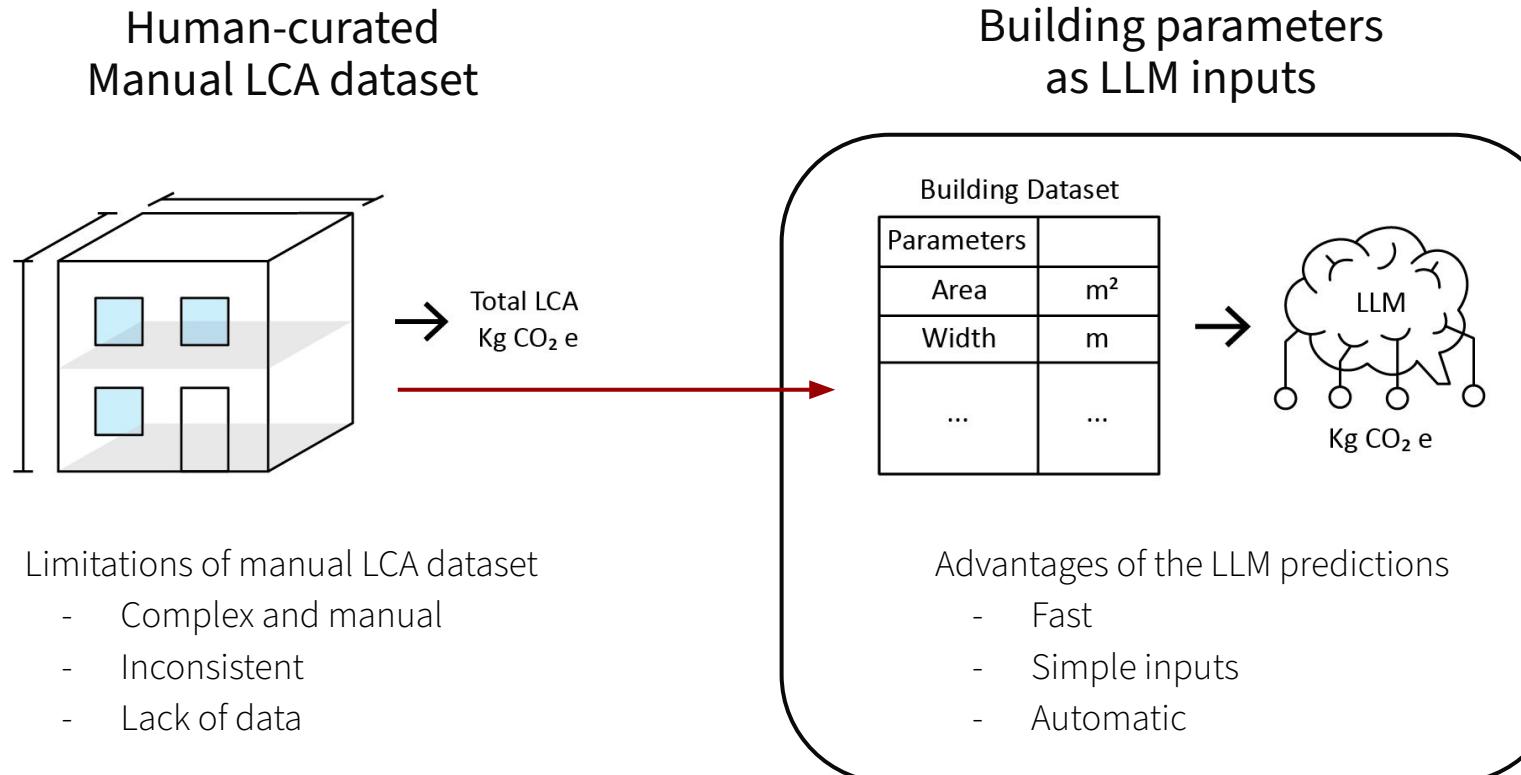
Conclusion

Building LCA Dataset from Carbon Leadership Forum (CLF)

- 292 buildings contains 72 parameters with carbon intensity
- 30 contributors across North America
- LLM Inputs: **49 parameters** excluding carbon data e.g. region, climate, building type, area, height, structural system, foundation, mass total
- LLM Outputs: **range of kg CO₂e/m²** for each building
 - Comparison with **eci_a_to_c_cfa**: Embodied carbon intensity (Sum of the project's total GWP for life cycle stages A-C normalized by conditioned floor area (CFA) for new construction projects)

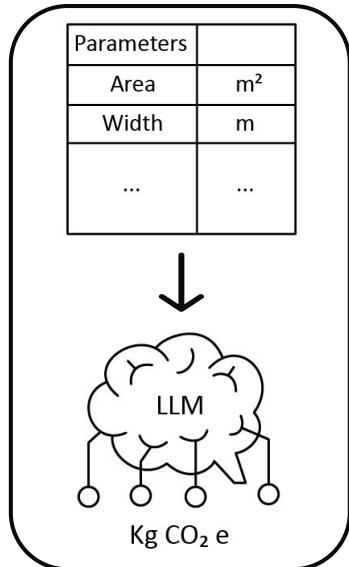


Transitioning Manual LCA dataset to LLM Input



In-context Learning Methods for LCA Prediction

LLM Innate Knowledge

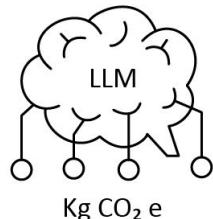


In-context Learning Methods for LCA Prediction

LLM Innate Knowledge

Domain Knowledge

Parameters	
Area	m ²
Width	m
...	...



5 journal papers and report
related to building LCA



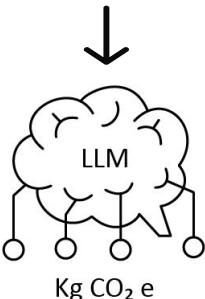
In-context Learning Methods for LCA Prediction

LLM Innate Knowledge

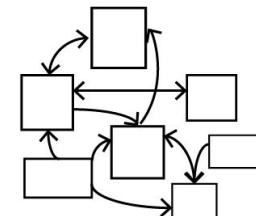
Domain Knowledge

Ontology-based Knowledge

Parameters	
Area	m ²
Width	m
...	...



5 journal papers and report related to building LCA



Building ontology:
Structured representation of building domain knowledge in each design step.
Interconnections between building parameters and complexity are systematically captured.

Building LCA process documentation

Detailed Claude 3-7-sonnet API Prompt Structure

LLM Innate Knowledge

(3)

Domain Knowledge

(1) + (3)

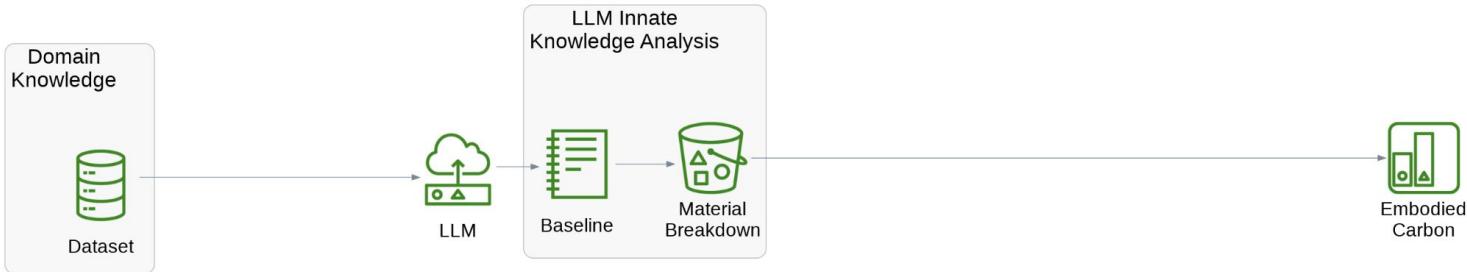
Ontology-based Knowledge

(1) + (2) + (3)

- (1) These five journal papers (.pdf) contain background information about early building LCA process.
- (2) This document (.txt) contains a detailed practice guide on the life cycle assessment of buildings focused on implementation. Use this guideline for LCA estimation.
- (3) Forget the information from previous sessions and calculate everything from the beginning. **data_glossary_LLM** (.xlsx) contains feature names and descriptions for the data. Reference the given feature information above. Estimate the embodied carbon value of (x) buildings **in the Excel file** (.xlsx). Each row represents different building data. Provide each range for each data in units of kgCO₂e/m². For floor area, you need to use bldg_cfa. Update the ranges. If cfa is not provided, you can use bldg_renovated_cfa + bldg_renovated_gfa for renovated buildings. Include life cycle stages A-C. Show results in a table with the range with cfa. If it does not converge: Can you provide a plausible range for each building given the information in units of kg CO₂e/m²?

Interactive LLM Sessions Reveal LCA Prediction Analysis

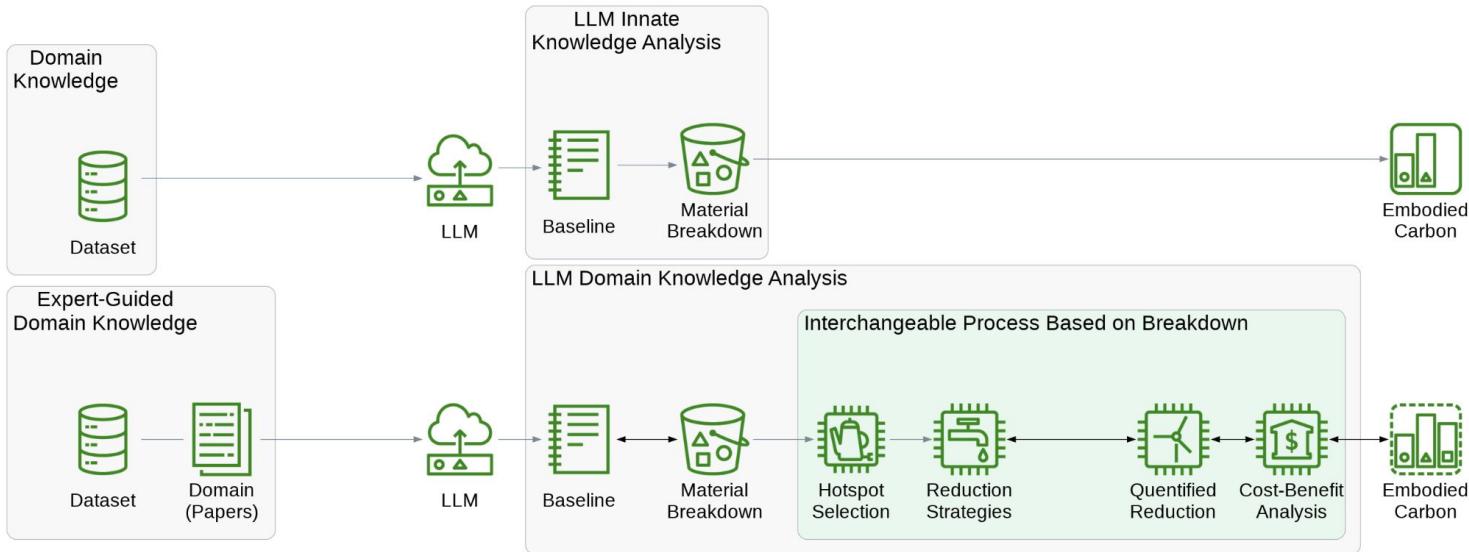
Innate
Knowledge



Interactive LLM Sessions Reveal LCA Prediction Analysis

Innate
Knowledge

Domain
Knowledge

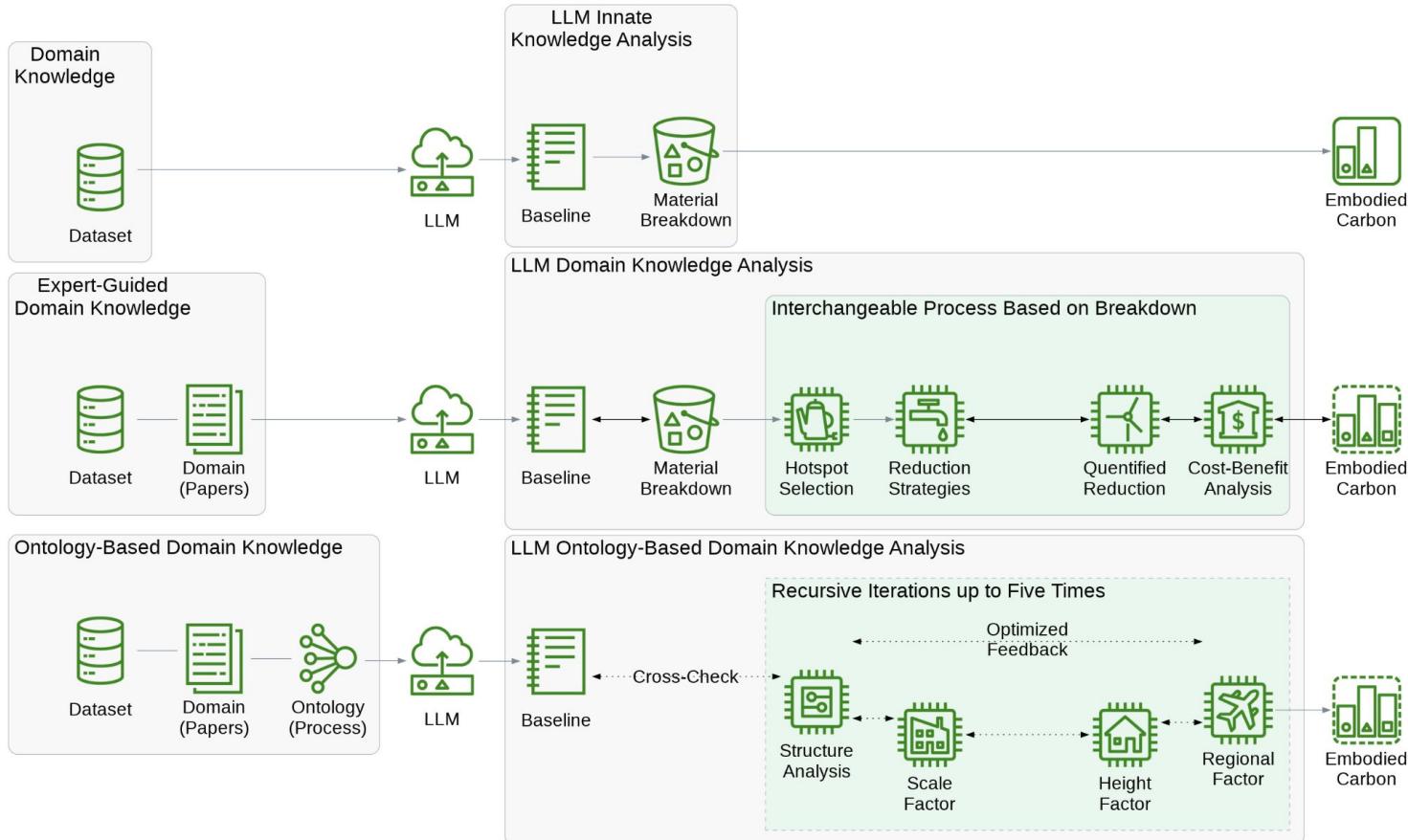


Interactive LLM Sessions Reveal LCA Prediction Analysis

Innate
Knowledge

Domain
Knowledge

Ontology-Based
Knowledge





1

Introduction

2

Methodology: building LCA dataset and LLM methods

3

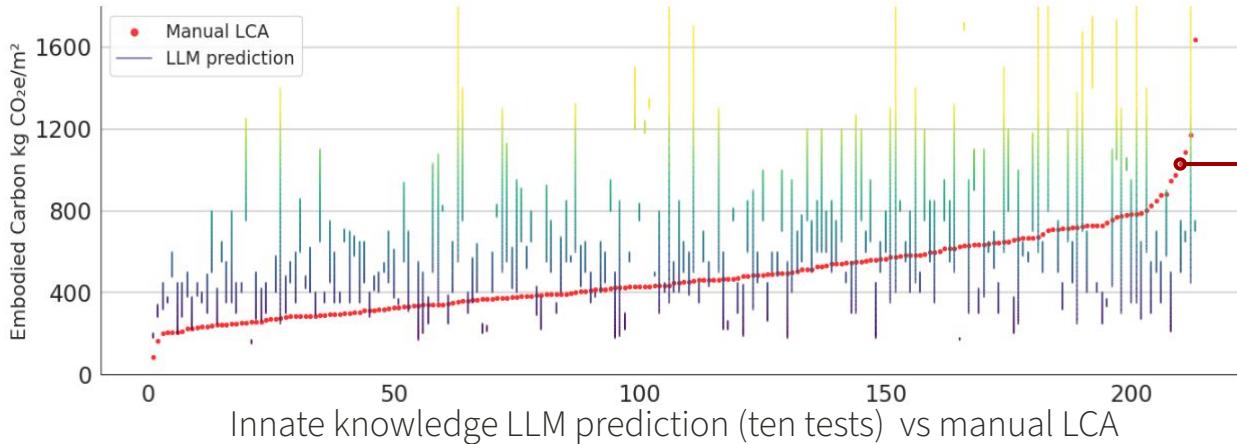
Results: LLM method comparative analysis

4

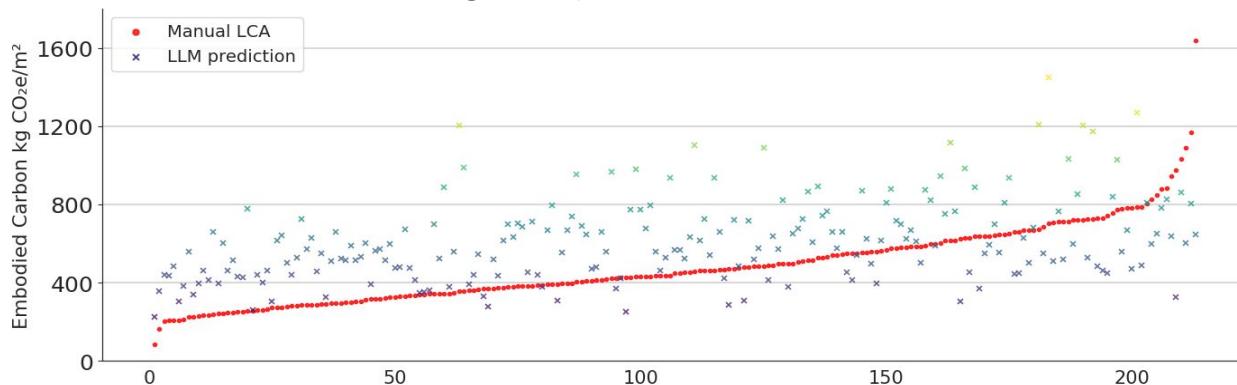
Conclusion

Comparing Manual LCA vs Innate Knowledge

Innate knowledge LLM prediction (one test) vs manual LCA

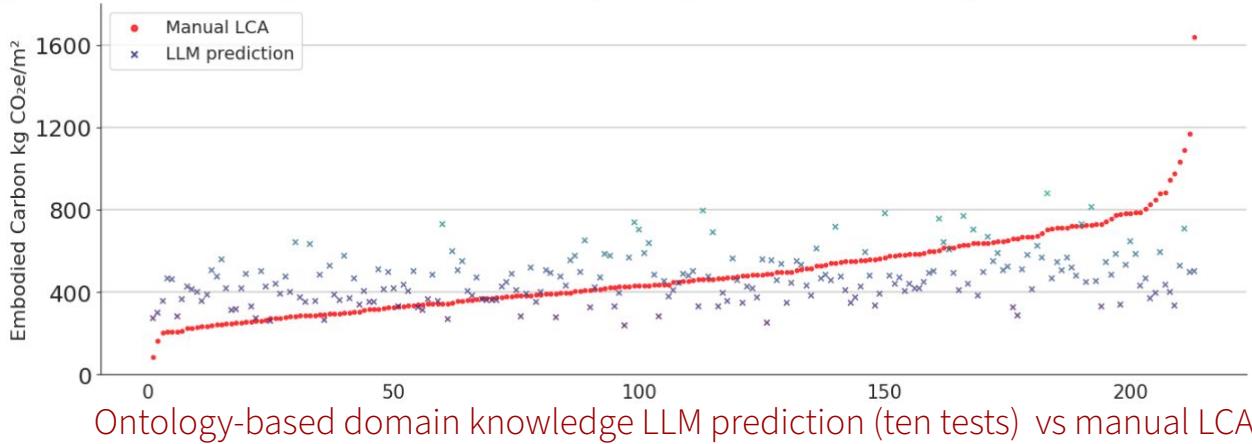


Innate knowledge LLM prediction (ten tests) vs manual LCA

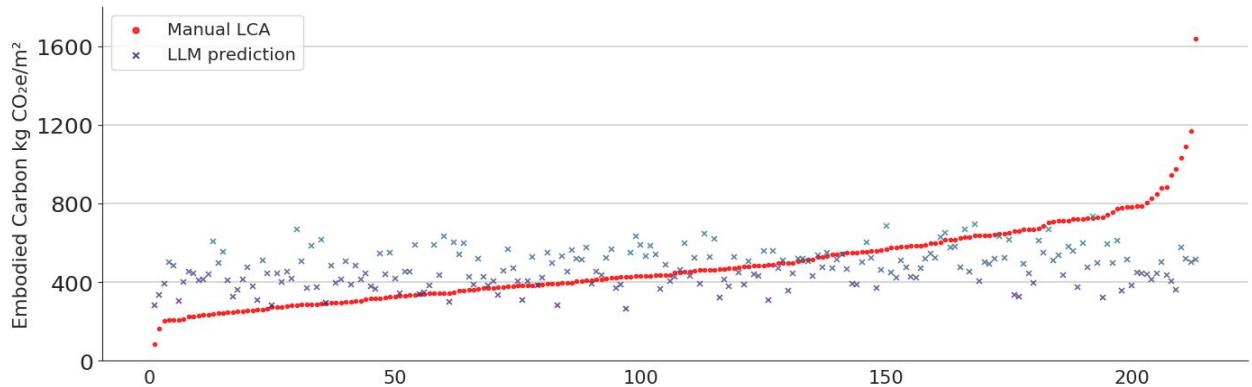


Comparing Domain vs Ontology-Based Knowledge

Domain knowledge LLM prediction (ten tests) vs manual LCA



Ontology-based domain knowledge LLM prediction (ten tests) vs manual LCA



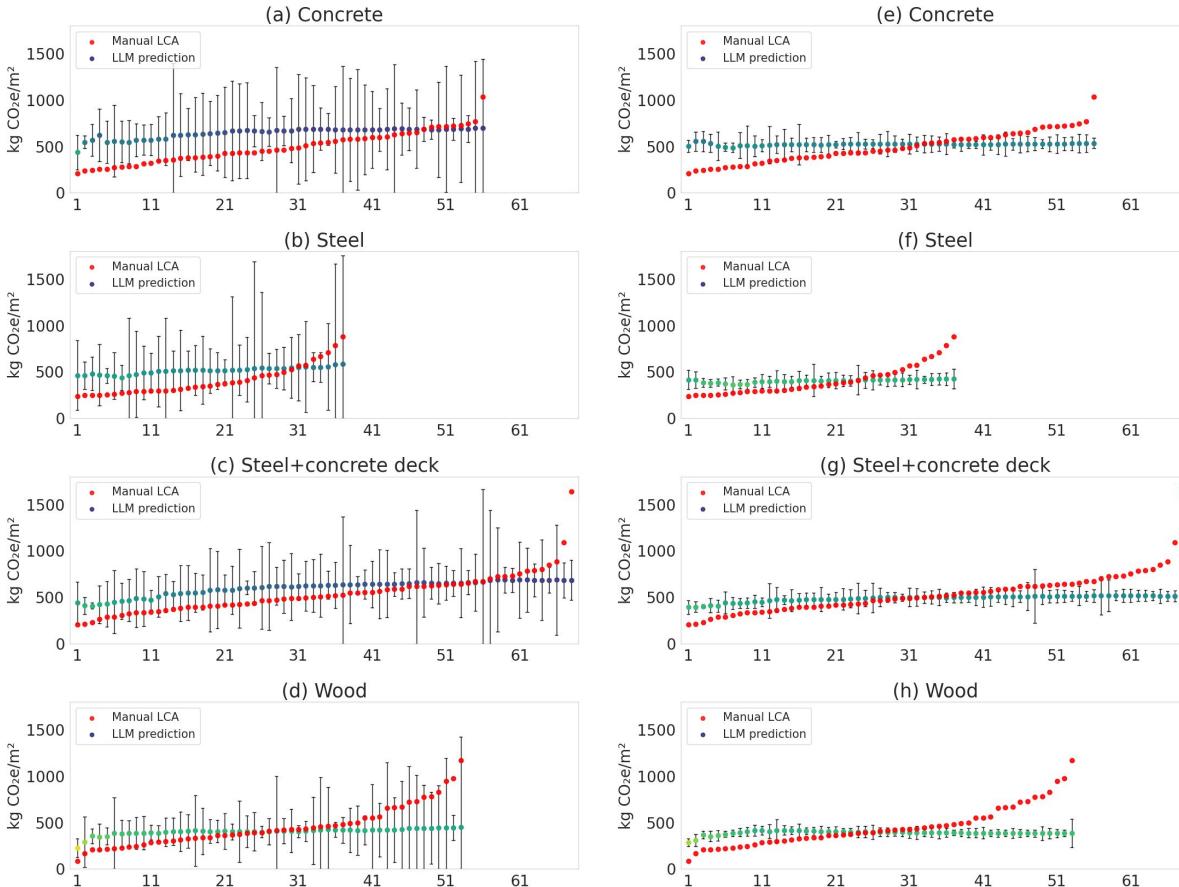
Error Reduction: Innate vs Ontology-Based

	Baseline manual	1. Innate	2. Domain	3. Ontology	Reduction (Innate vs Ontology)
RMSD*		392.4	244.3	230.9	41%
Distance		219.0	152.0	150.7	31%
Median	446.9	574.5	452.0	466.0	
Mean	482.5	613.3	464.4	471.8	
SD**		218.5	78.6	51.6	76%

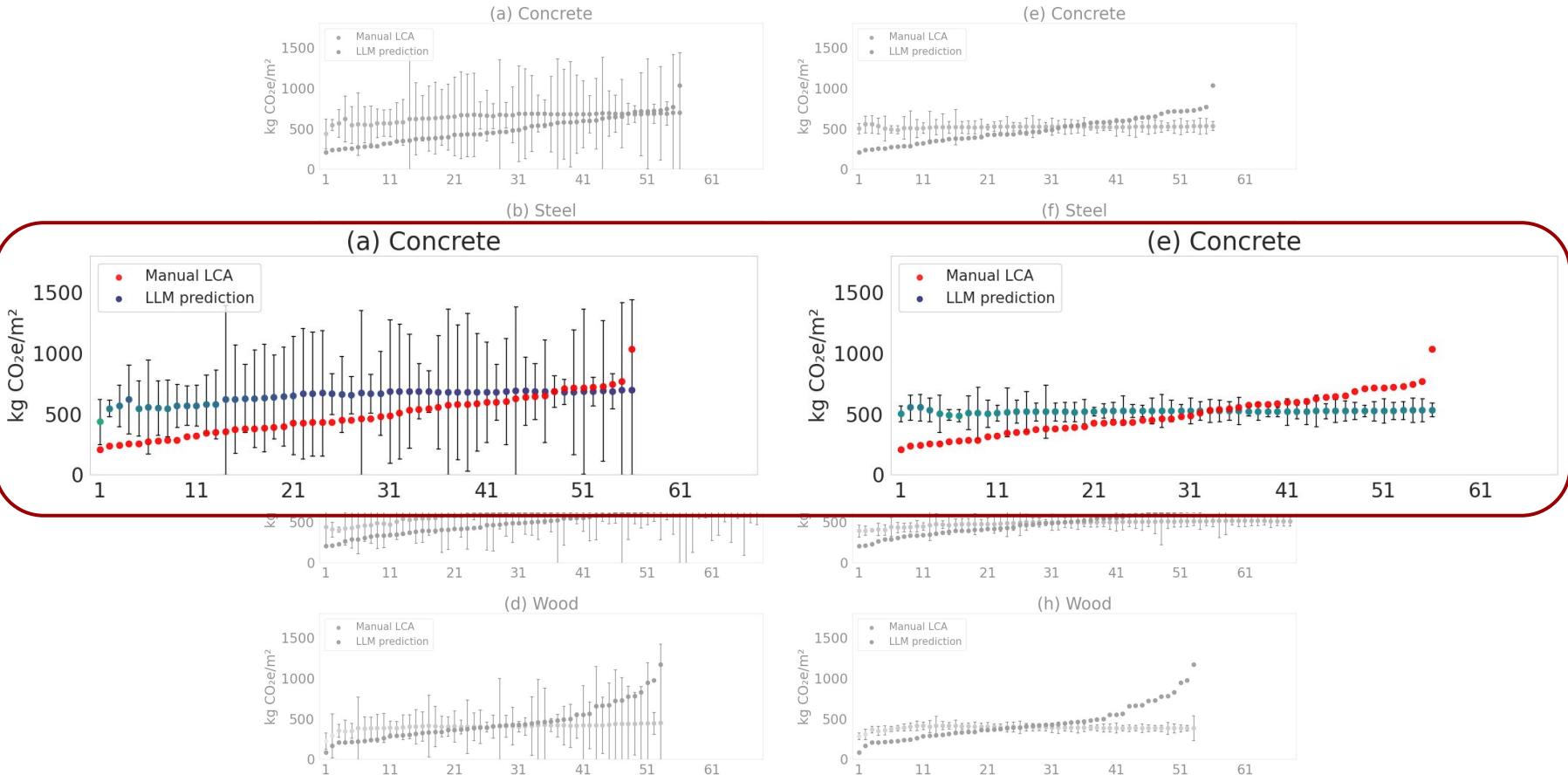
*Root Mean Square Error (RMSD)

**Standard deviation (SD)

Structural System Comparison: Innate vs Ontology-Based



Structural System Comparison: Innate vs Ontology-Based





- 1 Introduction
- 2 Methodology: building LCA dataset and methods
- 3 Results: LLM method comparative analysis
- 4 Conclusion

Conclusion

- The study evaluated three LLM methods for early building LCA, finding the **ontology-based approach** achieved **41% reduction in RMSE and 76% reduction in standard deviation** compared to using only innate knowledge.
- LLMs demonstrate **self-retrospective learning capabilities** where predictions gravitate toward mean values, resulting in underestimation of high carbon values and overestimation of low carbon values.
- The ontology-based approach transforms LLMs from simple pattern-matching to **iterative refinement through multiple feedback cycles**, enabling more accurate predictions when proper domain knowledge is provided.
- Future work will focus on developing **systematic human-LLM interaction frameworks** that integrate supervised learning, non-monotonic reasoning, and prompt engineering for practical industrial LCA applications, with potential extension to similar engineering problems like **clusters of buildings with similar features**.

Eunice Jung



LinkedIn

ehjung@stanford.edu



Paper

Thank you!

Questions?

Acknowledgement: Prof. Michael D. Lepech, Prof. Jie Wang, and Prof. Kincho H. Law

Stanford
ENGINEERING | Lepech Research Group

Stanford
ENGINEERING | Center for Sustainable Development &
Global Competitiveness

 Stanford University
Human-Centered
Artificial Intelligence

Stanford | Stanford Center at the
Incheon Global Campus

Yonghua Foundation

