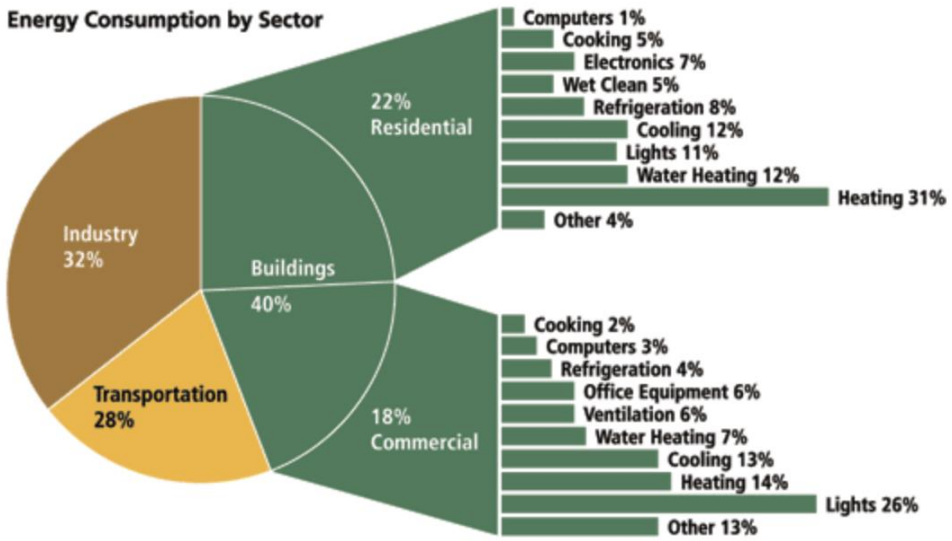
Toward Automated Building Energy Modeling with Large Language Models

Gang Jiang, PhD candidate

University of Utah

Background and motivation



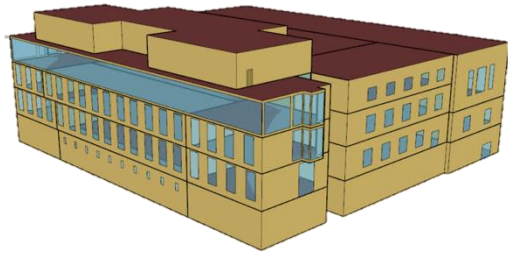
Buildings account for 40% of global energy consumption and CO_2 emission

Energy Efficiency, and decarbonization

Real-world building

Tool

Building Energy Modeling



Digital twin

Why LLM?

Accessible Modeling

EnergyPlus With Experts

🤔 Imagine

EnergyPlus With ???

Current: ~200,000 users

In the near Future: ???

60% of research based on building energy models

Scenario Scalability

Traditional - Eppy

SOTA – LLM Driven

Future

- Domain knowledge
- Technical Proficiency
- Config setting
- Programming skill
- Specific learning
- Limited functionality
- Poor scalability

V1 Param Settings

V2 Geo Diversities

V3 System Changings

V4 More complex scenarios

V5 More customized requirements

Repos

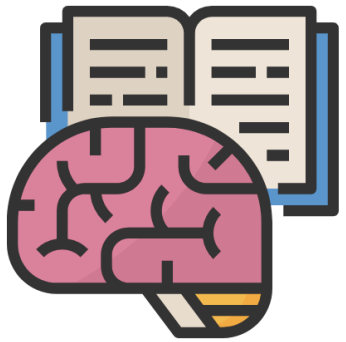
Power

Language

Automation

- Model Archi
- Transfer learning
- Task fitting
- Scaling Law
- Accessibility
- Interaction
- Customization

Challenges



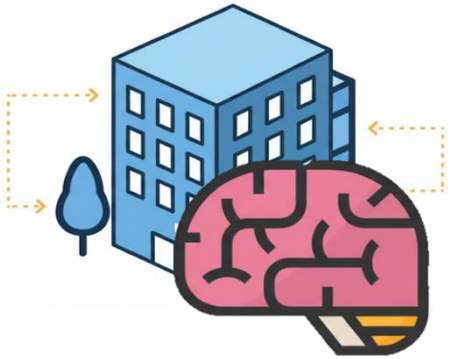
Domain Understanding

➢ Domian task: ChatGPT 5 >

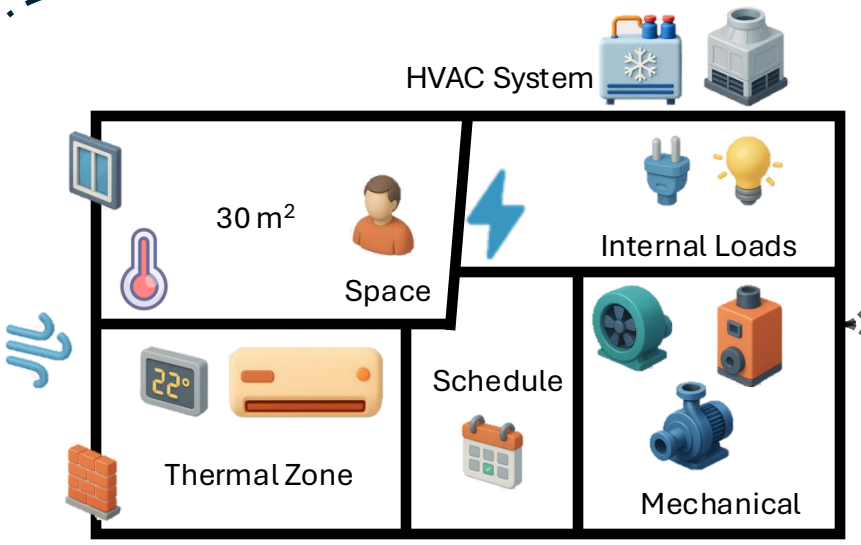
Input: Use EnergyPlus to simulate a building, that is 20 meters long, 10 meters wide, and 3 meters high. Provide me the runable error free IDF file.

Output: An **IDF-like** file, but with errors (hallucination).

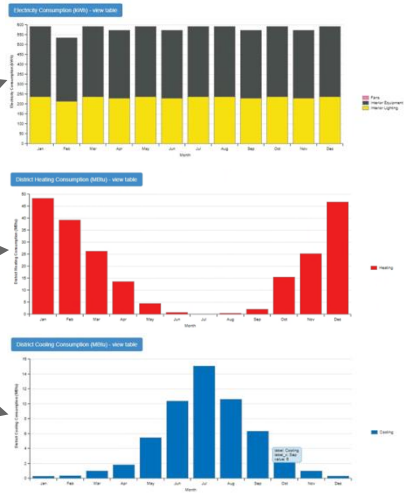
Domian LLM



Modeling Complexity



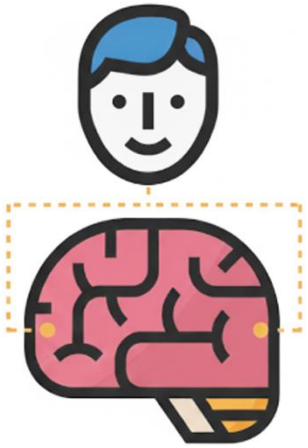
Building Energy Model



Energy Consumptions

- ❖ Long Context
- ❖ Coupling
- ❖ Scenario Cover

Challenges



Consistency

- General task:

Input:

wirte a shot song < 50 words

Same input, different output

Output 1:

🌙 **"Shine On"**

The night is long, the stars are near,
Hope keeps glowing, bright and clear.
Step by step, we find our way,
Through the dark, into the day.

Output 2:

🌙 **Little Song** 🌙

Wake with the sun,
Dreams just begun.
Hearts beat fast,
Moments won't last.

Prompting: Design specific instructions (or even examples) for LLMs

Fine-tuning: Prepare task specific training datasets

RAG: Retrieval augmented generation on extra knowledge base

MCP: Model context protocol, put your tools in context, let LLMs know how to apply tools (e.g., Python codes)

Agentic pipe: Hybrid tricks, e.g., applied multi-LLMs and multi-techs

- Building Modeling Task:
  - Users' diversity input description, but with same modeling intent (high consitency)
  - Generated result must be error free
  - 99% accuracy = 0% accuracy, means the pipeline still needs human involvement.

Solution 1: Prompting to inform

❖ **Prompt Engineering Framework**

<User>:
{Task Instruction}
+{Input-Output Demonstration #n+ ... +k_n with Explanations}
<Assistant>:
{Answer #n}

User Natural Language

<User>:
{Task Instruction}
+{Input-Output Demonstration #1}+...
+{Input-Output Demonstration #n}
+{Desired Task Input}

Reasoning

Logic of BEM

Self-learning

LLM

Foundation

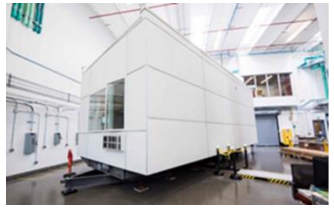
Domain Knowledge

Large Language Model

Open-source SOTA LLMs
e.g., Llama, Qwen, etc.
Or ChatGPT, etc.

Correspondi ng BEMs

Knowledge Informed



NREL, iUnit building

Few shots

BEM Examples

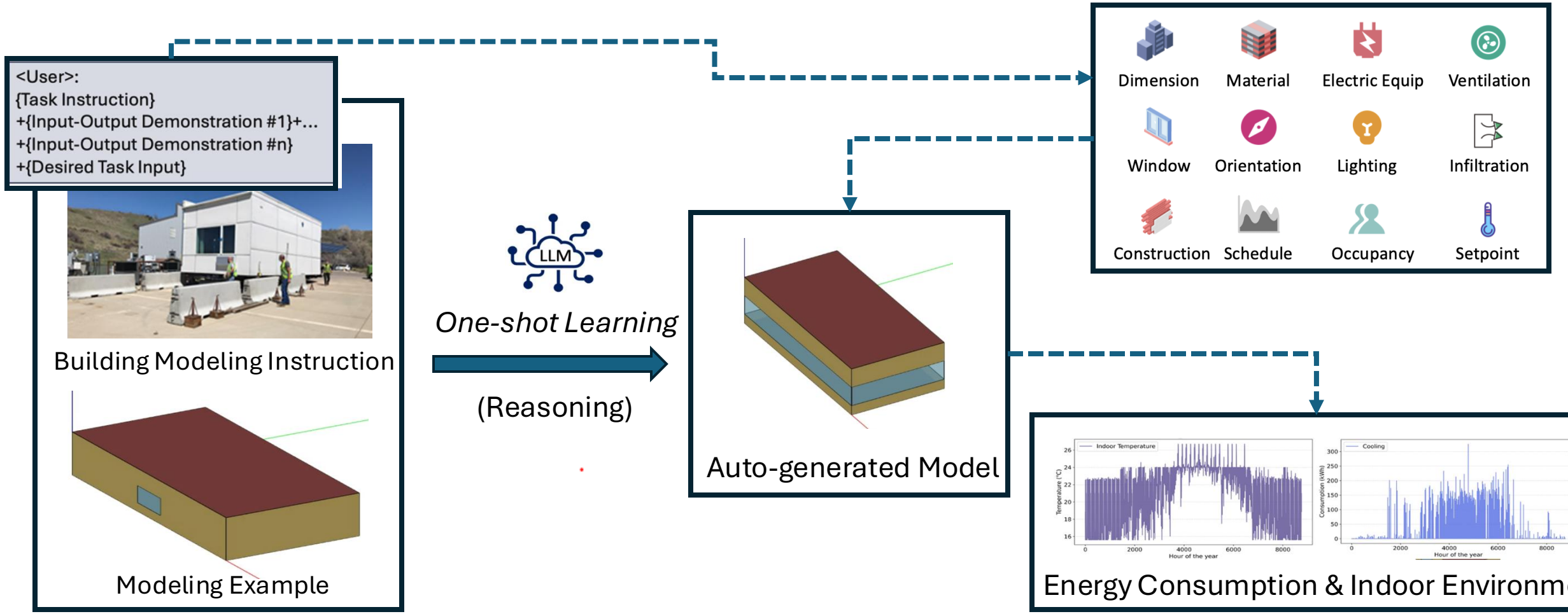
Prompting Strategies

Division Explanation

Jiang G, Ma Z, Zhang L, Chen J. (2025) Prompt engineering to inform large language model in automated building energy modeling. Energy

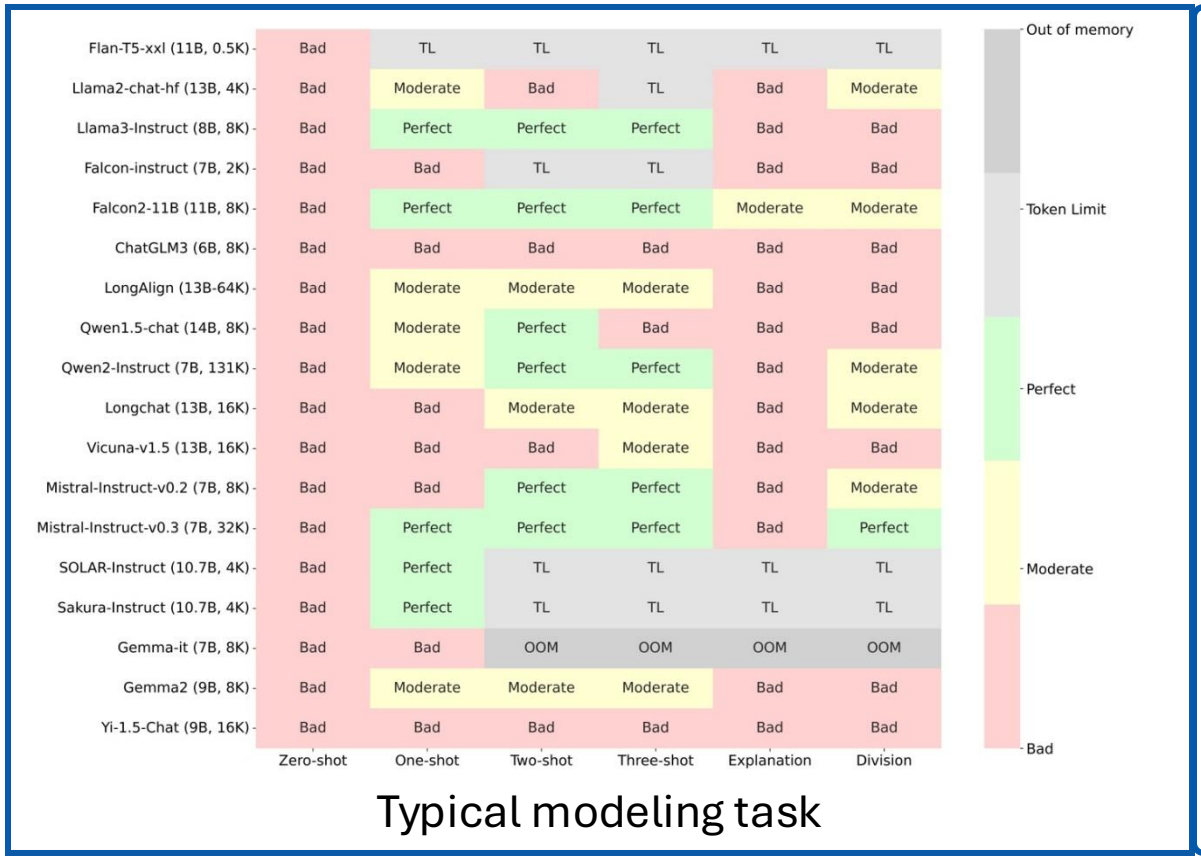
Case Study

❖ **Modeling Scenario: Building Energy Analysis & Retrofit**

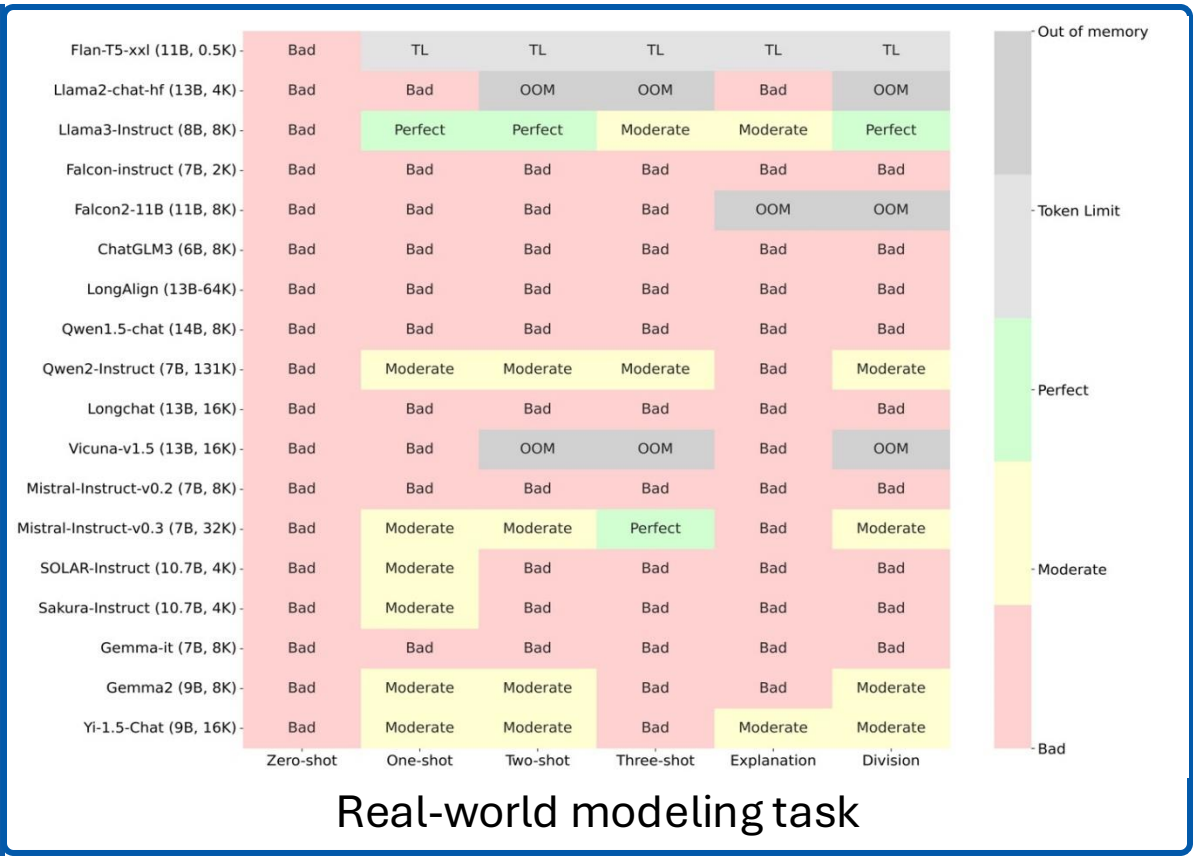


https://github.com/Gangjiang1/Prompting-for-Auto-building-Modeling/tree/main

Statistic Analysis



Typical modeling task



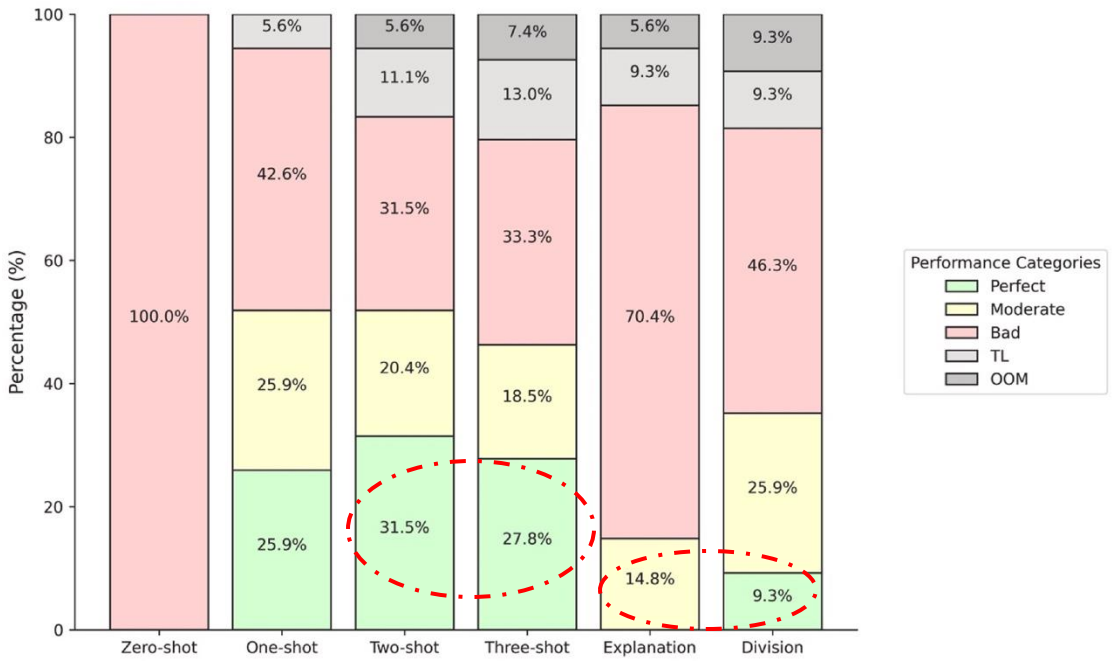
Real-world modeling task

Benchmarking LLMs on BEM:

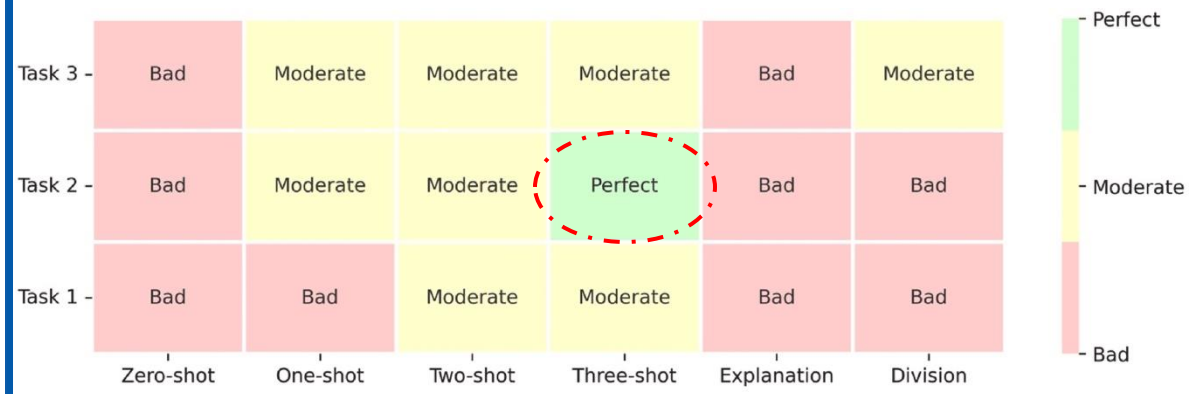
- For those scenarios, just one example (one-shot learning) is workable!!
- Over-design may yield deficient performance.
- Chain-of-thought (Division or Explanation) can improve modeling accuracy and success rate.

...

Statistic Analysis



Performance distribution across various prompting strategies.



Test results of prompt engineering with ChatGPT-4o.

More thoughts:

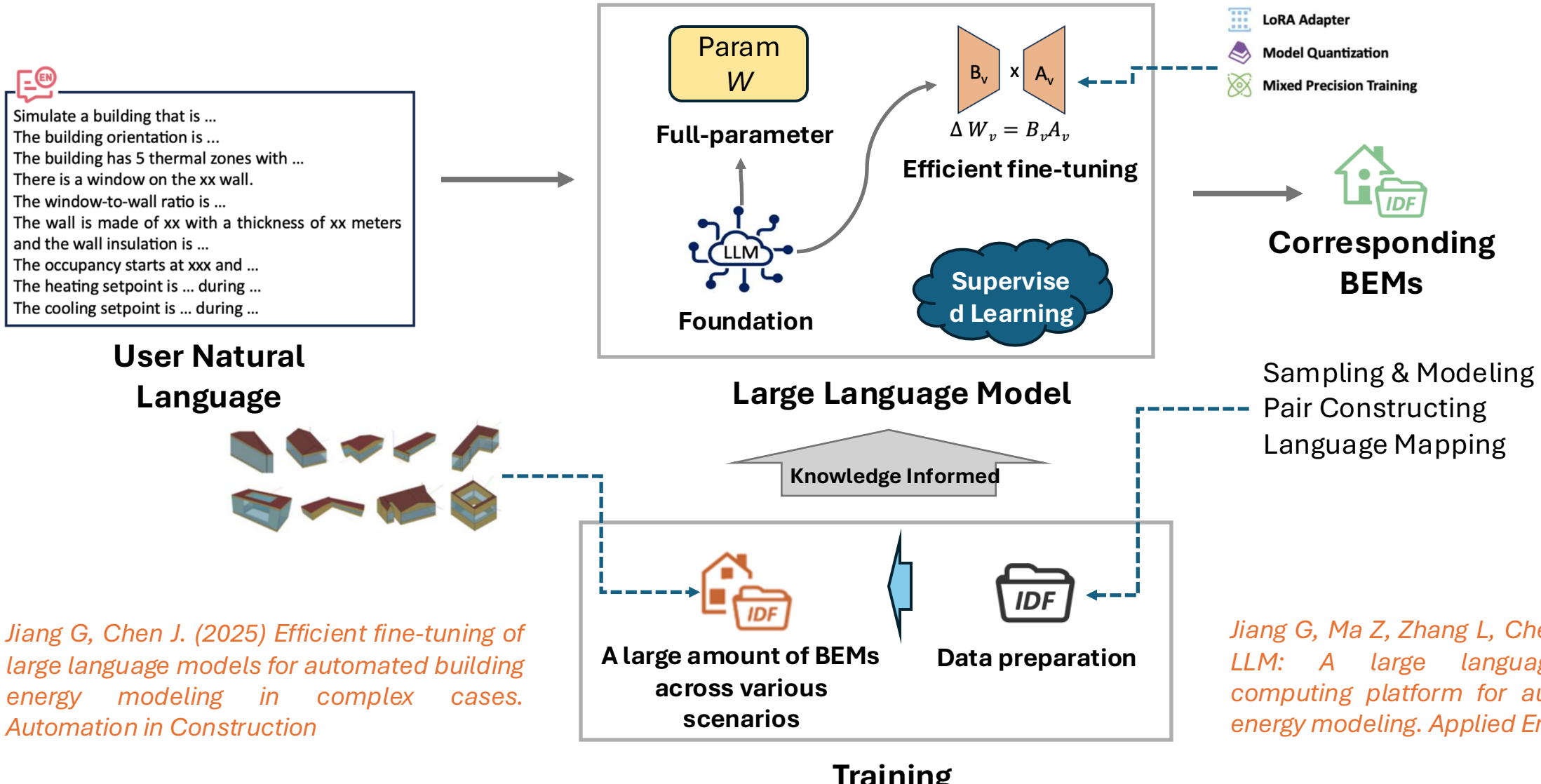
- One-shot vs Two-shot vs Three-shot
- Explanation vs Division
- Small open source LLM vs Big close source LLM

...

Future exploration ... prompting with new approaches and new LLMs

Solution 2: Fine-tuning to learn

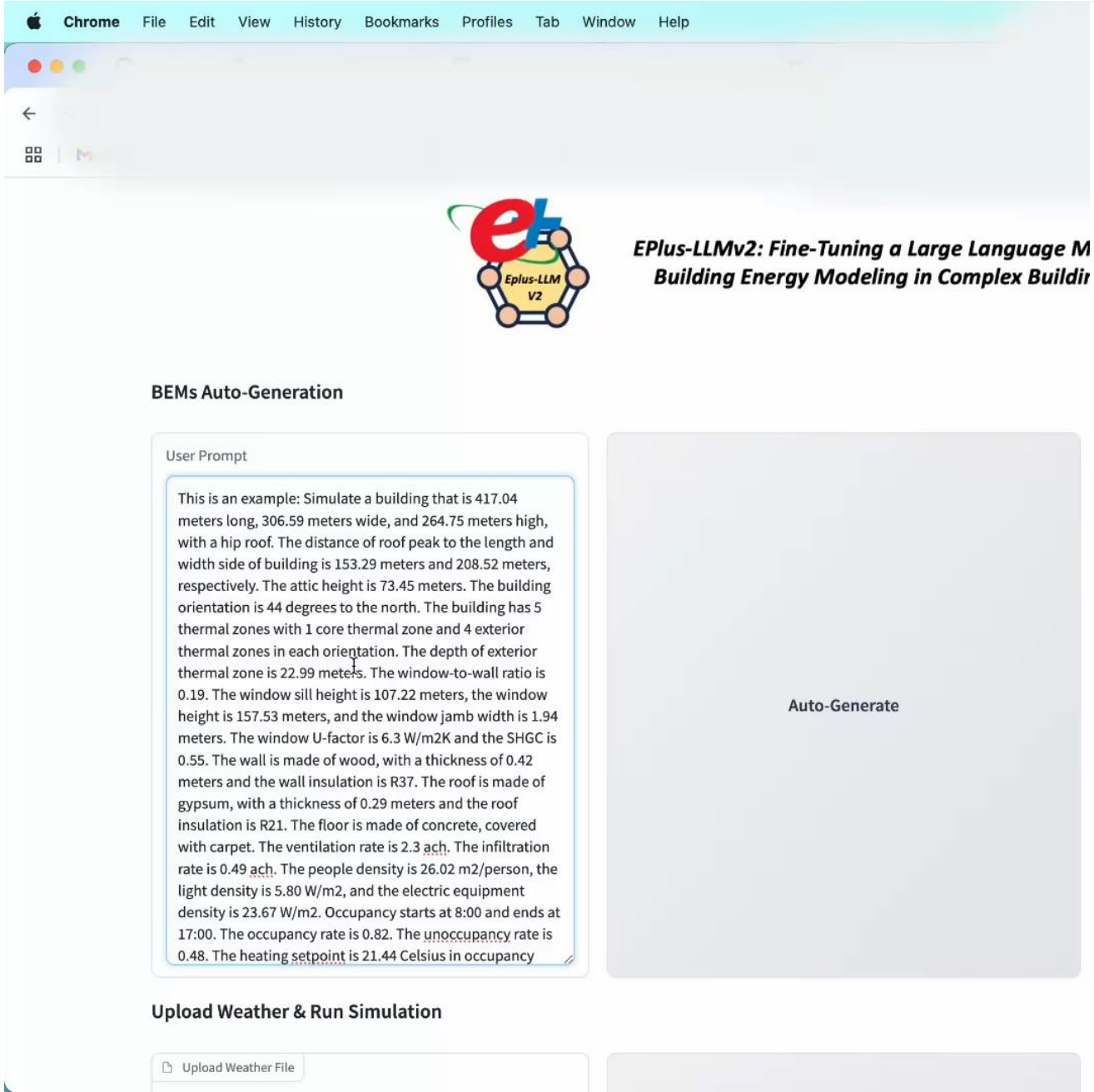
❖ **Fine-tuning Framework**



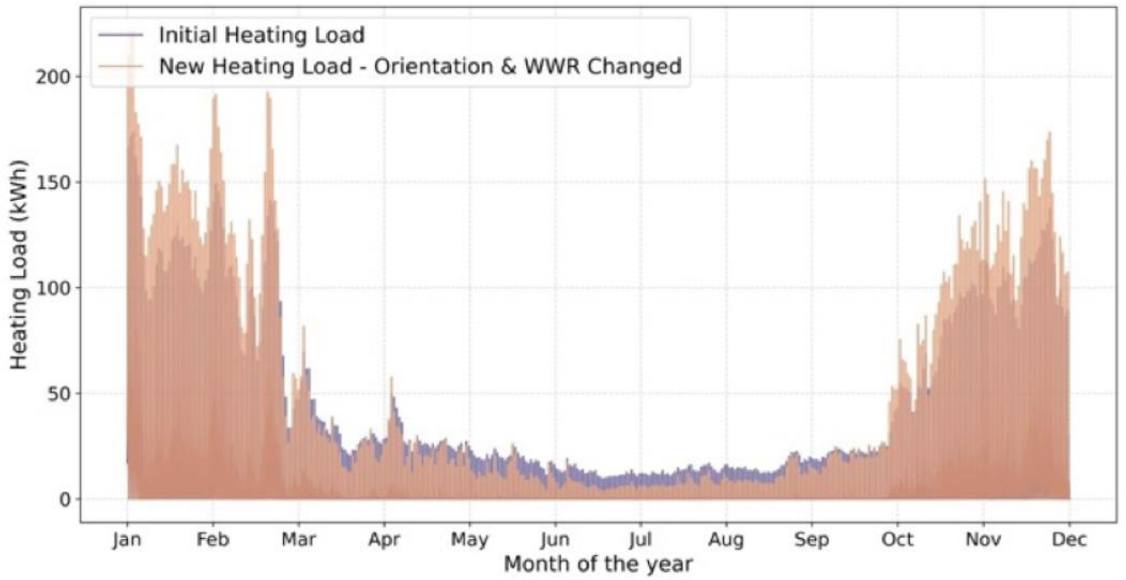
Jiang G, Chen J. (2025) Efficient fine-tuning of large language models for automated building energy modeling in complex cases. Automation in Construction

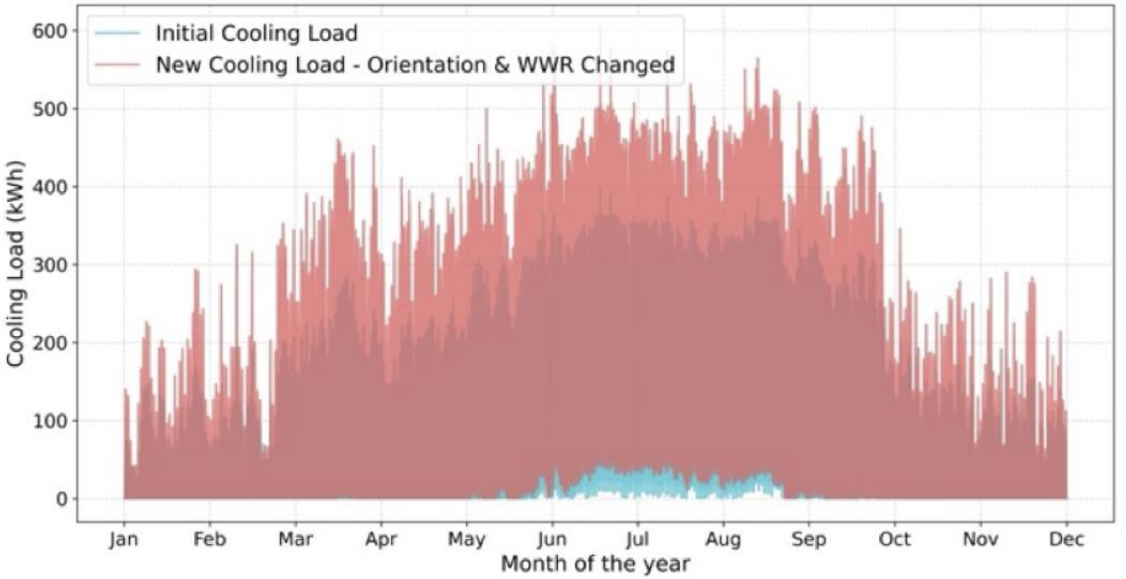
Jiang G, Ma Z, Zhang L, Chen J. (2024) EPlus-LLM: A large language model-based computing platform for automated building energy modeling. Applied Energy

❖ Video Demo



❖ Building Design & Retrofit





https://huggingface.co/EPlus-LLM

Summary

- ❖ Which LLM approach is BETTER?
  - ➢ Which TASK are you focused on?
  - ➢ Do you have computing source?
  - ➢ Do you have complex scenarios?
  - ➢ What are your requirements for robustness and scalability?
- ❖ Opportunities
  - ➢ Based on new tasks
  - ➢ Develop new pipelines
  - ➢ Integrated with domain knowledge
  - ➢ More automation, scalability, and robustness.