

Cite this as: Zhe SUN, Qiwei YAO, Ling SHI, Huaqiang JIN, Yingjie XU, Peng YANG, Han XIAO, Dongyu CHEN, Panpan ZHAO, Xi SHEN, 2025. Virtual sample diffusion generation method guided by large language model-generated knowledge for enhancing information completeness and zero-shot fault diagnosis in building thermal systems. *Journal of Zhejiang University-SCIENCE A*, 26(10):895-916. <https://doi.org/10.1631/jzus.A2400560>

# **Virtual sample diffusion generation method guided by large language model-generated knowledge for enhancing information completeness and zero-shot fault diagnosis in building thermal systems**

## **Key words:**

Information completeness, Large language models, Virtual sample generation, Knowledge-guided, Building air conditioning systems

# Exposure scenarios

## Problem Background

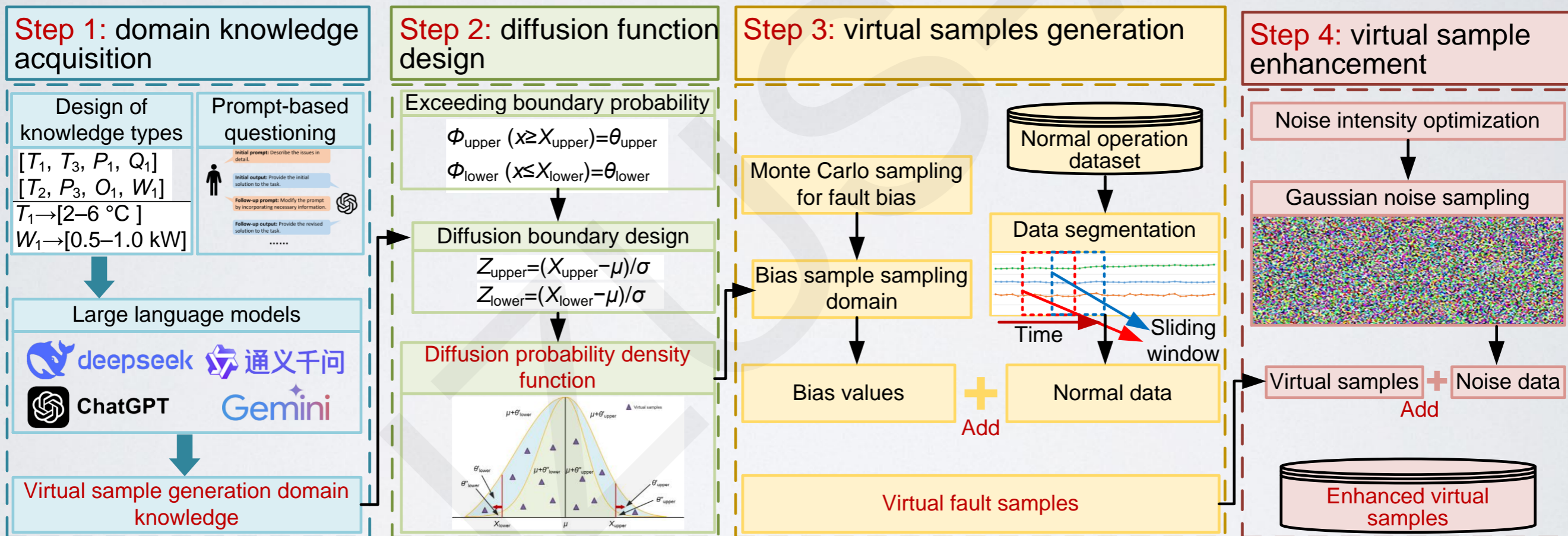
- HVAC systems have high data redundancy but low effective information.
- Very few fault samples, leading to poor model generalization.
- Traditional virtual samples rely on a small number of real samples.



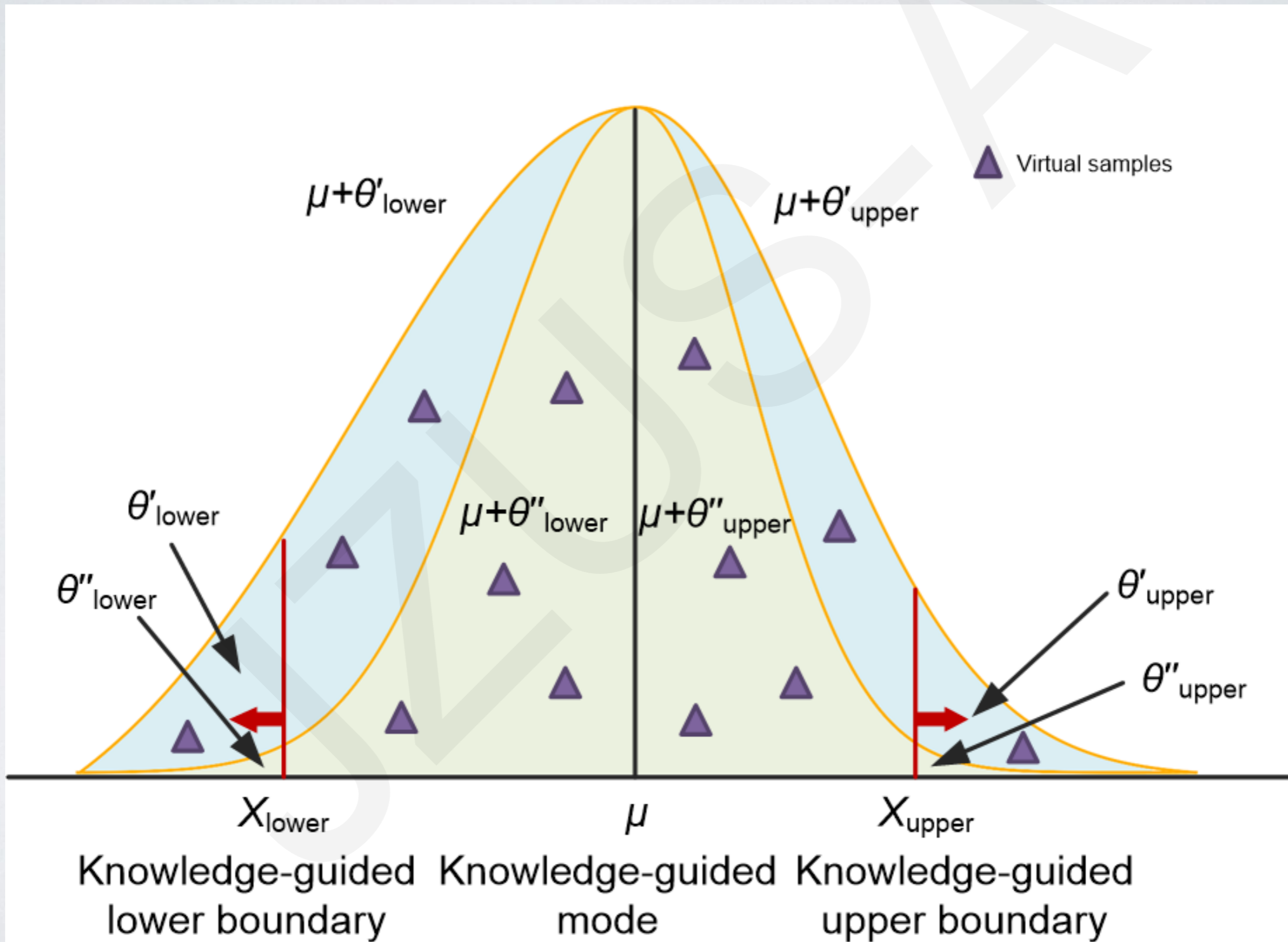
## Pain Points

- How to generate high-quality data without real samples?
- How to ensure the integrity of information in virtual samples?
- How to acquire expertise at low cost using LLM (Local Level Modeling)?

# Knowledge-Guided Virtual Sample Generation Framework



# Knowledge-guided mega-trend-diffusion technique



# Large language model prompt

## Prompt

You are a senior HVAC engineer, and you are working on creating a universally applicable HVAC fault operation and maintenance manual.

Consider the following scenario: a chiller with a rated cooling capacity of **316 kW** uses **R134A** as the refrigerant, and both the condenser and evaporator are configured as **water-to-water shell-and-tube heat exchangers**. Under summer cooling conditions, with outdoor temperatures ranging from **25 to 35 °C**, the chilled water outlet temperature ranges from **5 to 10 °C**, and the cooling water return temperature ranges from **20 to 30 °C**.

When the chiller experiences the following fault: **non-condensable gases in refrigerant (1%–5% by volume nitrogen)**. Please analyze how the following 2 parameters will change: **condenser pressure**; **discharge superheat**.

Please note that we need to list the deviation range and average deviation of the two parameters relative to normal operating conditions when the fault occurs. For example, for fault A, parameter X will increase, with an increase range of approximately 3–5 °C and an average deviation of 3.8 °C. Please provide your reasoning.

## Output of GPT o3-mini-high

### 1. Condenser pressure:

- **Change:** The presence of non-condensable gases reduce the effective condensation area in the condenser, leading to an increase in condenser pressure.
- **Deviation range:** The condenser pressure typically **increases by about 3–5 bar** compared to normal operating conditions.
- **Average deviation:** Approximately **4 bar**, depending on the volume of non-condensable gas.

### 2. Discharge superheat:

- **Change:** Non-condensable gases affect condensation efficiency, causing an increase in discharge temperature, which in turn raises discharge superheat.
- **Deviation range:** Discharge superheat typically **increases by about 5–10 °C**.
- **Average deviation:** Approximately **7 °C**, reflecting the impact of reduced condenser efficiency.

### Reasoning:

Non-condensable gases hinder the heat transfer process in the condenser, leading to higher condenser pressure and more refrigerant remaining superheated, which results in an increase in discharge superheat.

## Output of DeepSeek R1

### 1. Condenser pressure:

- **Change:** Non-condensable gases occupy space in the condenser, reducing effective heat transfer and increasing total system pressure.
- **Deviation range:** Condenser pressure **increases by approximately 1.5–3.0 bar** compared to normal conditions.
- **Average deviation:** **About 2 bar**, proportional to the non-condensable volume (1%–5% nitrogen).

### 2. Discharge superheat:

- **Change:** Elevated condenser pressure raises refrigerant saturation temperature. However, the compressor discharge temperature increases further due to reduced condenser efficiency, amplifying superheat.
- **Deviation range:** Discharge superheat **increases by 8–15 °C**.
- **Average deviation:** **About 11 °C**, reflecting impaired heat rejection.

### Reasoning:

Non-condensable gases (e.g., nitrogen) increase the total pressure in the condenser by occupying volume without condensing. This reduces the effective partial pressure of the refrigerant, forcing the compressor to work harder to achieve the same cooling effect. The higher total pressure raises the refrigerant's saturation temperature, but the actual condensing efficiency drops, leading to a disproportionate rise in compressor discharge temperature.

# Main Contributions and Future Directions

## Main Contributions:

- Proposed KG-MTD, enabling diffusion without relying on real samples
- Leveraged LLMs to provide domain knowledge at low cost
- Combined Monte Carlo sampling with noise injection to increase information entropy
- Used virtual samples to enhance model generalization and diagnostic stability

## Future Directions:

- Defining knowledge structures for higher-dimensional HVAC models
- Integrating the framework with physics-based models (PINNs)
- Optimizing real-time deployment in practical scenarios
- Generating virtual samples for multi-fault combinations