RetrievalAugmented LLM Framework for Robust Automated Generation of HVAC Diagnostic Models

Chujie Lu
Building Physics and Sustainable Design
KU Leuven
Leuven, Belgium
Faculty of Architecture and the Built Enviornment
Delft University of Technology
Delft, Netherlands
chujie.lu@kuleuven.be

Laure Itard
Delft University of Technology
Delft, Netherlands
L.C.M.Itard@tudelft.nl

Dirk Saelens
KU Leuven
Leuven, Belgium
EnergyVille
Genk, Belgium
dirk.saelens@kuleuven.be

Abstract

Diagnostic Bayesian networks (DBNs) offer effective and interpretable solutions for HVAC fault detection and diagnosis (FDD), yet their practical construction and deployment require substantial manual effort. The work proposes a retrieval-augmented LLM framework to achieve robust and automated DBN generation. Preliminary results demonstrate that the proposed framework effectively mitigates hallucinations and significantly improves the structural quality of the generated DBNs.

CCS Concepts

• **Computing methodologies** → *Knowledge representation and reasoning*; • **Applied computing** → *Engineering*.

Keywords

Fault Detection and Diagnosis, Large Language Model, Retrieval Augmented Generation, Bayesian Network, HVAC Systems

ACM Reference Format:
Chujie Lu, Laure Itard, and Dirk Saelens. 2026. RetrievalAugmented LLM Framework for Robust Automated Generation of HVAC Diagnostic Models. In *ACM Sustainability Week 2026 (ACM Sustainability Week Companion '26), June 22–25, 2026, Banff, AB, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3765611.3815346

1 Introduction

Fault detection and diagnosis (FDD) in HVAC systems plays a vital role in building management systems by reducing unnecessary energy consumption and ensuring that indoor environmental performance meets design expectations. While data-driven FDD methods have gained increasing attention, their adoption in the building industry remains limited due to the scarcity of high-quality labeled datasets, challenges in model generalization, and concerns regarding their "black-box" nature.

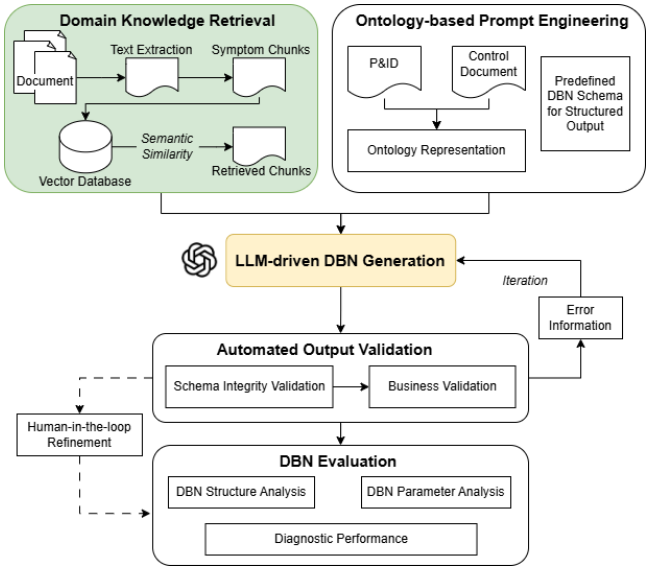

Figure 1: Retrieval-Augmented LLM Framework for Robust Automated Generation of HVAC DBNs

Diagnostic Bayesian networks (DBNs), as a probabilistic graphical model, offer effective, robust and interpretable symbolic FDD solutions for HVAC systems [4]. However, constructing and deploying DBNs in practice is a time-consuming process, as it typically requires system-specific analysis of HVAC configurations, manual design of rule-based fault symptoms, and careful specification of conditional probabilities [7, 8]. These efforts further rely on engineers with interdisciplinary expertise spanning HVAC systems and probabilistic modeling or programming, which creates substantial adoption barriers.

Recently, large language models (LLMs) have demonstrated remarkable progress across a wide range of domains and applications. In the building industry, LLMs have shown growing potential in supporting tasks such as engineering diagram digitalization, automated energy model development, indoor air distribution design, and energy retrofit assessment [3, 5]. In our previous work, we explored the feasibility of using LLMs as an assistant to automatically

ACM Sustainability Week Companion '26, Banff, AB, Canada

ACM ISBN 979-8-4007-2199-1/26/06
https://doi.org/10.1145/3765611.3815346

generate DBNs for HVAC systems [2]. Although the results demonstrated the strong generative and reasoning capabilities of LLMs, the generation process was highly stochastic and occasionally incomplete, resulting in the omission of key diagnostic symptoms. More critically, hallucination effects were observed, , where incorrect or non-existent symptoms were introduced. To address this challenge, this study proposes a retrieval-augmented LLM framework that systematically integrates domain knowledge and HVAC fault reports to mitigate hallucination and enhance the robustness of automated DBN model generation, thereby improving the HVAC FDD performance.

2 Method

The proposed retrieval-augemented LLM framework is illustrated in Figure 1.

First, *domain knowledge retrieval* aims to provide reliable and domain-specific information to ground LLM reasoning. HVAC-related documents, such as fault reports and technical manuals, are first processed through text extraction and segmented into symptom-level knowledge chunks (manageable units). These chunks are embedded and stored in a vector database. Given a target fault scenario, semantically relevant knowledge (potential symptoms) is retrieved via similarity search.

Then, *ontology-based prompt engineering* is designed to structure LLM inputs and outputs in accordance with domain semantics and DBN requirements. HVAC piping and instrumentation diagrams (P&IDs) and control documents are mapped to an ontology representation that captures HVAC components, connections, and signal variables.

Given the retrieved potential symptoms and ontology-based prompts, *LLM-driven DBN generation* module automatically generate specific DBN structures and parameters. To ensure reliability, an *automated output validation* module is introduced, including schema integrity validation to check schema compliance and business validation to verify the basic plausibility, including consistent node definitions and normalized conditional probability tables. If validation fails, error information will be generated and fed back to LLM, enabling iterative refinement of the DBN generation process.

Finally, *DBN evaluation* is conducted through structure analysis, parameter analysis, and HVAC FDD performance assessment. When the generated DBN fails to satisfy the requirements, human-in-the-loop refinement can be incorporated to facilitate expert validation and adjustment.

3 Preliminary Results

In this work, an air handling unit (AHU) of an office building in the Netherlands is selected as the test case. Domain knowledge from Brains4Buildings project reports, and relevant literature [6–9] is integrated using LangChain [1]. Knowledge chunks are embedded with OpenAI embeddings and stored in a Chroma vector database, enabling semantic retrieval to support retrieval-augmented DBN generation. The ontology representation and predefined DBN schema are heuristically designed and encoded in a JSON-based format. GPT-5.2 is the selected LLM to perform DBN generation to diagnose four common AHU faults, including heating coil valve stuck at higher position (HCV Stuck-P), heat recovery wheel stuck at low speed (HRW Stuck-N), supply fan stuck at low speed (SFS Stuck-N), and supply air temperature sensor positive bias (SAT Bias-P).

Table 1: Structural comparison of generated DBNs

Approach	N_s	α_s	ρ_s	N_c	α_c	ρ_c
LLM-only	11	54.5%	60%	21	28.6%	42.9%
Retrieval-augmented LLM	9	100%	90%	13	84.6%	78.6%

Table 1 presents a structural comparison of the DBNs generated by the LLM-only approach and the proposed retrieval-augmented LLM framework, using an expert-predefined DBN structure as the reference. The LLM-only approach generates a larger number of symptoms (N_s) and fault–symptom connections (N_c). However, symptom precision (α_s) and connection precision (α_c) are only 54.5% and 28.6%, respectively. These results suggest that a significant portion of the generated symptoms and fault–symptom connections are incorrect or hallucinated in the LLM-only approach.

In contrast, the retrieval-augmented LLM framework produces fewer but more reliable diagnostic structures. Although the number of generated symptoms and connections is less, the proposed framework achieves significantly higher precision, reaching 100% symptom precision and 84.6% connection precision. Moreover, symptom recall (ρ_s) increases from 60% to 90% and connection recall (ρ_c) increases from 42.9% to 78.6%, thereby reducing the risk of missing diagnostically relevant symptoms.

4 Conclusion

This work demonstrates that the proposed retrieval-augmented LLM framework can effectively mitigate hallucination and improve the quality of generated DBN structure for HVAC systems. Further analysis on generated DBN parameters and FDD performance will be conducted in the future.

Acknowledgments

This work is funded by the Junior Postdoctoral Fellowship (1255626N) of Research Foundation Flanders (FWO). The authors also appreciate the support from Kropman B.V. and Dr. Arie Taal.

References

[1] LangChain Contributors. 2026. LangChain. https://www.langchain.com/ Accessed: April 7th, 2026.

[2] Chujie Lu and Laure Itard. 2026. Leveraging LLM for P&ID-based Automated Code Generation in HVAC Fault Detection and Diagnosis. In *Proceedings of the 15th REHVA HVAC World Congress - CLIMA 2025* (Milan, Italy). Springer Nature.

[3] Chujie Lu, Shalika Walker, Christian Struck, Laure Itard, and Dirk Saelens. 2025. Poster Abstract: P&ID-to-Graph: LLM-Assisted Digitalization of HVAC Diagrams. In *Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (Colorado School of Mines, Golden, CO, USA) *(BuildSys '25)*. Association for Computing Machinery, New York, NY, USA, 296–297. doi:10.1145/3736425.3772104

[4] Chujie Lu, Ziao Wang, Martín Mosteiro-Romero, and Laure Itard. 2025. Diagnostic Bayesian network in building energy systems: Current insights, practical challenges, and future trends. *Energ Build* 341 (2025), 115845. doi:10.1016/j.enbuild.2025.115845

[5] Nan Ma, Rania Labib, Robert Amor, Adrian Chong, Cheng Fan, Kasimir Forth, Xiaoqin Fu, Stefan Fuchs, Tianzhen Hong, Nina Klimenkova, Jabeom Koo, Shundong Li, Steven Tanner McCullough, June Young Park, Roee Shraga, Sungmin Yoon, Liang Zhang, and Yiting Zhang. 2026. Ten questions concerning Large Language Models (LLMs) for building applications. *Build Environ* 291 (2026), 114260.

doi:10.1016/j.buildenv.2026.114260

[6] Jeffrey Schein, Steven T. Bushby, Natascha S. Castro, and John M. House. 2006. A rule-based fault detection method for air handling units. *Energy and Buildings* 38, 12 (2006), 1485–1492. doi:10.1016/j.enbuild.2006.04.014

[7] Lars van Koetsveld van Ankeren, Chujie Lu, and Laure Itard. 2025. Implementing diagnostic Bayesian networks for heat recovery ventilation in real-world scenarios: A Dutch case study. *J Build Eng* 111 (2025), 113527. doi:10.1016/j.jobe.2025.113527

[8] Ziao Wang, Chujie Lu, Arjen Meijer, Shalika Walker, and Laure Itard. 2026. Fault detection and diagnosis for heat recovery ventilation using 4S3F method: Impact of diverse sensor configurations. *Energ Build* 345 (2026), 116913. doi:10.1016/j.enbuild.2025.11691

[9] Yang Zhao, Jin Wen, Fu Xiao, Xuebin Yang, and Shengwei Wang. 2017. Diagnostic Bayesian networks for diagnosing air handling units faults – part I: Faults in dampers, fans, filters and sensors. *Applied Thermal Engineering* 111 (2017), 1272–1286. doi:10.1016/j.applthermaleng.2015.09.121