



Delft University of Technology

Document Version

Final published version

Citation (APA)

Lu, C., Struck, C., Miller, C., Saelens, D., & Itard, L. (2026). Artificial Intelligence for HVAC Diagnostics: Towards the Era of Large Language Models. *REHVA European HVAC Journal*, 2026(2), 59-62.

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Artificial Intelligence for HVAC Diagnostics: Towards the Era of Large Language Models

KEY WORDS: FAULT DETECTION AND DIAGNOSIS, HVAC, ARTIFICIAL INTELLIGENCE, LARGE LANGUAGE MODEL, DIAGNOSTIC BAYESIAN NETWORK



CHUJIE LU

Delft University of Technology, The Netherlands, KU Leuven, Belgium
c.j.lu@tudelft.nl



CHRISTIAN STRUCK

Saxion University of Applied Sciences, The Netherlands



CLAYTON MILLER

Singapore Management University, Singapore



DIRK SAELENS

KU Leuven, Belgium, EnergyVille, Belgium



LAURE ITARD

Delft University of Technology, The Netherlands

Faults silently degrade HVAC performance, wasting energy and diminishing indoor well-being. How can artificial intelligence help us diagnose them? This paper shares insights into challenges of large-scale practical HVAC diagnostics and presents efforts from the Brains4Buildings project, specifically highlighting the emerging potential of Large Language Models (LLMs) as intelligent assistants toward self-learning and adaptive diagnostics.

The Dilemma of HVAC Diagnostics

In line with the EU Smart Readiness Indicator (SRI) under the Energy Performance of Buildings Directive (EPBD), Fault detection and diagnosis (FDD) is an essential component of smart building operation and maintenance. By continuously monitoring system performance, FDD enables the early detection of abnormal behaviour, identification of root causes, and ensures efficient HVAC operation, occupant comfort, and reduced energy waste.

Over the past several decades, artificial intelligence (AI) has evolved into a powerful tool to support HVAC FDD, with two distinct paradigms emerging: **Symbolic AI** and **Sub-symbolic AI**, as illustrated in **Figure 1**. Symbolic AI, often referred to as rule-based expert systems, relies on explicit human expertise by interpreting HVAC documents such as piping and instrument diagrams (P&IDs) and defining “If-Then” rules that link observable symptoms to

possible faults. For example, a rule might state, “If the supply air pressure is too low, then a fan failure may have occurred.” Such rule-based diagnosis is easy to understand, verify, and refine when system conditions change. By contrast, sub-symbolic AI represents data-driven models such as machine learning and deep learning techniques, which learn fault patterns and system behaviour directly from operational data. While often more accurate, these models act as “black box”, making their diagnosis difficult to interpret.

However, a dilemma exists in the current practice of HVAC diagnostics. Sub-symbolic AI methods have become increasingly popular in research due to their high accuracy, accounting for nearly 70% of recent publications [1]. Yet, applications in practice are still dominated by rule-based systems, even though they are frequently criticized in the literature for producing many false alarms, being time- and labour-intensive, and heavily reliant on domain expertise.

Data Quality Challenge in Data-driven HVAC Diagnostics

Why do sub-symbolic AI methods fall short in practice? Aside from the limited interpretability caused by its “black box” nature, sub-symbolic AI is often constrained by the availability of high-quality data. But what does *high-quality data* actually mean in the context of HVAC diagnostics? It generally includes:

- **Labeled faulty data.** Unlike building energy prediction tasks, where labels naturally exist, HVAC diagnostics require each fault type to be properly identified and annotated to train supervised learning models. In practice, building operation data remains naturally unlabeled.
- **Sufficient sensor configuration.** Missing or insufficient sensors can directly lead to model failure. Our investigation examined 18 air handling units (AHUs) from a Dutch building service company. It revealed that they had varying sensor configurations, with most AHUs failing to comply with the sensor configuration standards recommended by ASHRAE or ISSO [2].
- **Balanced faulty labels.** Models trained on imbalanced datasets tend to bias diagnostics toward normal operation and more frequent fault types. Our analysis of historical maintenance records from a Dutch building service company revealed that fault frequencies varied significantly across fault types, highlighting the inherent imbalance in real-world HVAC datasets [3].

- **Representative data distributions.** Training data should capture the real HVAC operational conditions to ensure that the model generalizes well beyond specific cases. Yet most studies rely on short-term or commissioning fault experiments to collect data that fails to reflect the actual distribution.

Therefore, in reality, collecting such high-quality data is nearly impossible. This limitation remains a significant barrier to the large-scale practical implementation of sub-symbolic AI in HVAC diagnostics.

Diagnostic Bayesian Networks: Integrating Expertise into HVAC Diagnostics

Diagnostic Bayesian Networks (DBNs) provide a promising solution to integrate decades of accumulated engineering knowledge, such as traditional rule-based diagnostics and expert experience, into modern HVAC diagnostics [1].

As shown in **Figure 2**, DBNs are constructed through expert analysis of P&IDs, where causal reasoning defines the relationships between faults and their observable symptoms (either predefined rules or data-driven symptom detection) [5,6]. Experts assign prior probabilities to represent the likelihood of faults and conditional probabilities to describe how symptoms depend on specific faults, forming the basis for diagnostic inference. Once established, DBNs use prior and conditional probabilities to perform diagnostic

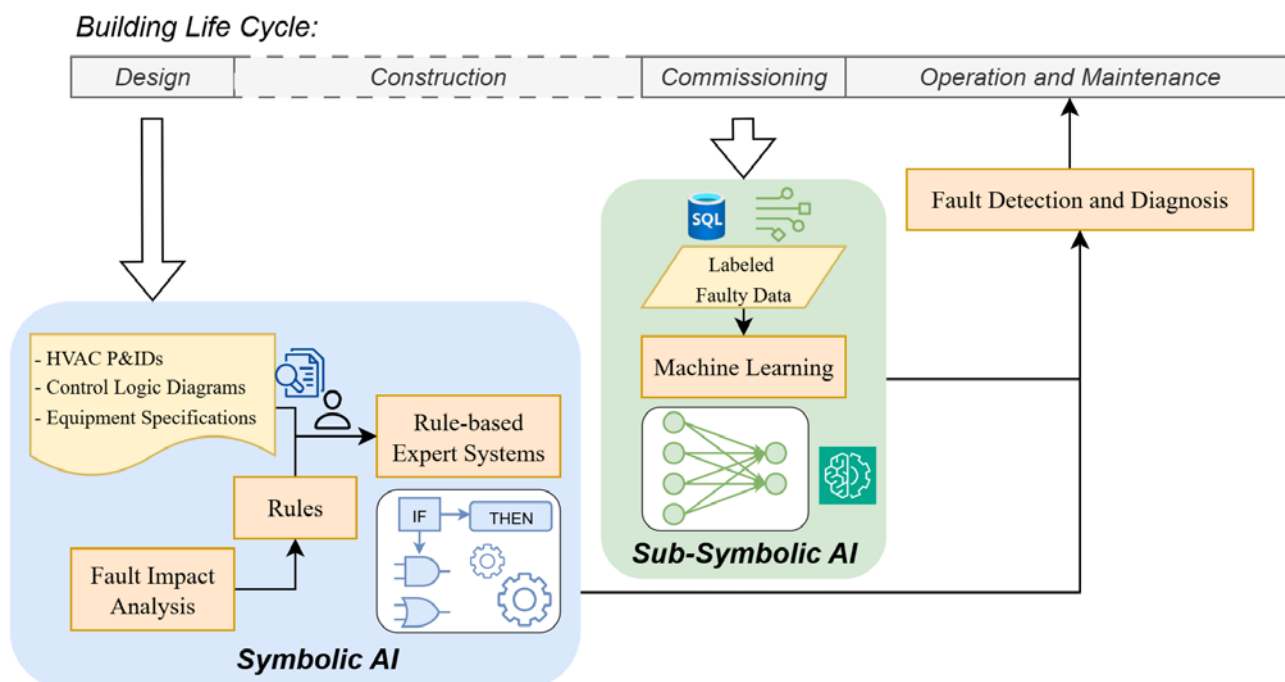


Figure 1. AI Modelling Framework for HVAC Diagnostics.

inference, identifying the most probable faults from detected symptoms derived from real-time sensor data streams. DBNs can also be flexibly combined with sub-symbolic AI methods and can easily integrate occupant feedback and expert observations as additional symptom inputs [1,7].

In short, DBNs align well with HVAC design and implementation practices and can provide high diagnostic accuracy, strong interpretability, and robustness to uncertainty. While DBNs have demonstrated their effectiveness, their development remains highly dependent on expert experience and manual input, which makes the process labor-intensive and time-consuming across different HVAC systems [1,8].

LLM Opportunities in HVAC Diagnostics

Recent advances in LLMs have introduced new possibilities for addressing persistent challenges in HVAC diagnostics. Pretrained on massive datasets, LLMs such as ChatGPT can process extensive linguistic and technical knowledge, enabling them to perform complex tasks involving language comprehension, reasoning, knowledge abstraction, and even vision-based interpretation. This emerging capability raises an essential question for the building industry: *Can LLMs interpret HVAC documents (e.g., P&IDs) and support engineers in developing diagnostic models (e.g., DBNs)?* To explore this potential, we present three promising LLM-assisted applications in HVAC diagnostics.

LLM-assisted P&ID Digitalization

Most P&IDs of existing HVAC systems are still stored as scanned images or static PDF files, which is a major barrier to automated modeling for HVAC diagnostics. We explored how LLMs can automatically convert

HVAC piping and instrumentation diagrams (P&IDs) from static images into machine-readable formats (e.g., JSON), without any task-specific training (i.e., in a *zero-shot* setting) [9]. The preliminary results show that directly applying LLMs to P&ID digitization remains highly challenging, even with state-of-the-art models such as GPT-5, which failed to recognize symbols effectively (0% mean average precision, mAP). To address this, we proposed a preprocessing strategy that segments P&IDs into local image crops and pairs them with full-diagram annotations containing bounding boxes for global context. With this approach, the LLM achieved improved symbol recognition performance, reaching a mAP of 31.05%. This demonstrates the feasibility of applying LLMs to P&ID digitalization, though further improvement is still required.

LLM-assisted DBN Symptom Reasoning

Developing effective DBNs relies heavily on expert knowledge and experience to define fault–symptom relationships, requiring proficiency in both HVAC systems and Bayesian reasoning. To evaluate the potential of LLMs in supporting this process, a comparative study was conducted involving four HVAC engineers and GPT-5 in constructing an AHU DBN. Expert opinions were obtained through a semi-structured survey covering 11 common AHU faults, where participants identified relevant symptoms and rated their severity and confidence levels. GPT-5 was evaluated through a prompt-based reasoning task under the same conditions. The survey revealed that, while some consistency existed among expert opinions, there were considerable differences in how they perceived the strength of fault–symptom relationships, reflecting the inherent subjectivity of human reasoning. Due to these differences, the diagnostic performance of expert-constructed DBNs also differed considerably, correctly

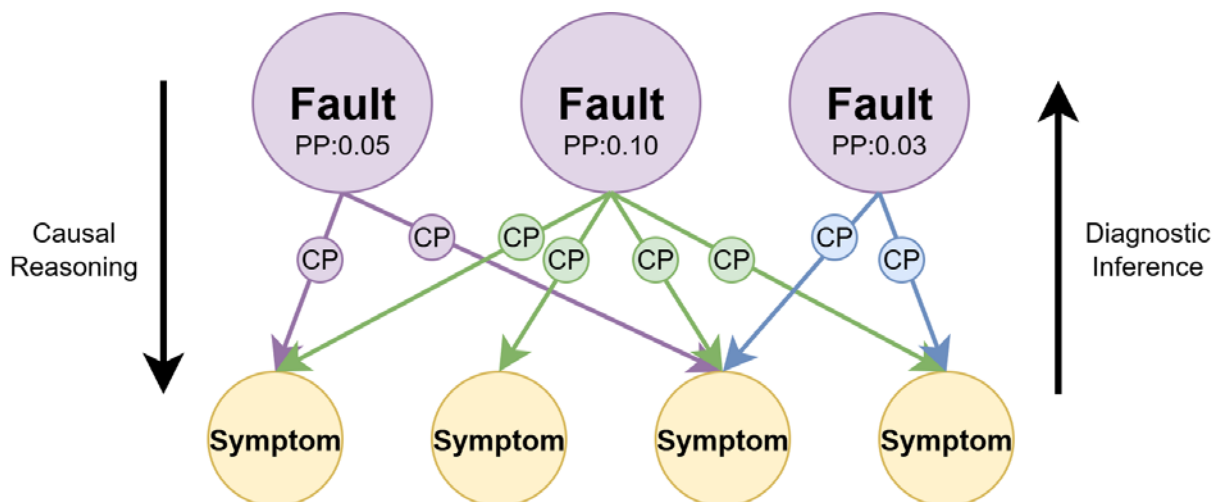


Figure 2. Illustration of DBNs (PP: Prior Probability; CP: Conditional Probability). Adapted from. [4].

diagnosing between 2 and 10 fault cases out of 15. In comparison, the DBN based on GPT-5-derived symptom relationships achieved comparable accuracy, correctly diagnosing 10 fault cases and matching the best-performing expert.

LLM-assisted DBN Code Generation

LLMs have recently demonstrated strong capabilities in producing structured outputs such as executable code. We explored LLM-assisted DBN code generation for HVAC diagnostics using machine-readable files extracted from P&IDs as inputs [10]. Claude 3.5 Sonnet was employed for its strong ability in code generation tasks. To enhance the reliability of the generated code, we employed Chain-of-Thought (CoT) reasoning to guide the model through a four-step DBN modeling framework [1], thereby decomposing the complex task of DBN construction into explicit reasoning steps. The results demonstrated that the LLM was able to generate functional DBN code in Python. A qualitative analysis confirmed that the generated fault-symptom relationships and prior probabilities were largely consistent with those of the existing DBNs in the literature, demonstrating the LLM's capability to capture HVAC-relevant causal logic. In contrast, the quantitative analysis based on experimental data revealed that only one fault ("supply fan stuck") was correctly identified. The limited diagnostic accuracy was mainly attributed to inappropriate symptom thresholds and model hallucinations within the generated code. These findings underscore both the potential and current limitations of LLM-assisted DBN modelling, emphasizing the need for expert-in-the-loop supervision and further refinement to ensure accurate and robust fault diagnosis.

Outlook

These studies highlight the emerging potential of LLMs as intelligent assistants in HVAC diagnostics, from interpreting visual-textual engineering documents to supporting symptom reasoning and automating model generation. Although the results remain preliminary, they show that LLMs can meaningfully apply engineering expertise in automatically developing diagnostic models that extend beyond traditional sensor data-driven methods. Future work will first focus on structured prompt design and domain-specific fine-tuning to enhance LLM performance in HVAC diagnostics. Meanwhile, emerging paradigms such as retrieval-augmented generation and AI agents offer new possibilities for developing self-learning diagnostic systems that integrate document-based engineering knowledge with real-time operational data. These advances lay the foundation for *neuro-symbolic*

AI frameworks, which combine the interpretability of symbolic reasoning with the adaptability of sub-symbolic learning, to enable self-learning HVAC diagnostics capable of deeper understanding, causal reasoning, and continuous learning.

Acknowledgements

The authors acknowledge the support from the Brains4Buildings project (No. MOOI32004) by the Dutch program for Mission-Driven Research, Development and Innovation, and the Junior Postdoctoral Fellowship (No. 1255626N) by Research Foundation - Flanders (FWO).

References

- [1] C. Lu, Z. Wang, M. Mosteiro-Romero, L. Itard, Diagnostic Bayesian network in building energy systems: Current insights, practical challenges, and future trends, *Energy Build* 341 (2025) 115845. <https://doi.org/10.1016/j.enbuild.2025.115845>.
- [2] Z. Wang, C. Lu, A. Meijer, S. Walker, L. Itard, Fault detection and diagnosis for heat recovery ventilation using 4S3F method: impact of diverse sensor configurations, (2025). <https://doi.org/10.2139/ssrn.5511688>.
- [3] S. Gopalan, A. Rijs, S. Chitkara, A. Thamban, R. Kramer, Fault prioritisation for Air Handling units using fault modelling and actual fault occurrence data, *Energy Build* 319 (2024) 114476. <https://doi.org/10.1016/j.enbuild.2024.114476>.
- [4] R. Kramer, S. Walker, Automatische Fout Detectie en Diagnose binnen bereik?, *TVVL Magazine* (2025) 12–17.
- [5] A. Taal, L. Itard, W. Zeiler, A reference architecture for the integration of automated energy performance fault diagnosis into HVAC systems, *Energy Build* 179 (2018) 144–155. <https://doi.org/10.1016/j.enbuild.2018.08.031>.
- [6] A. Taal, L. Itard, P&ID-based symptom detection for automated energy performance diagnosis in HVAC systems, *Autom Constr* 119 (2020). <https://doi.org/10.1016/j.autcon.2020.103344>.
- [7] M. Mosteiro-Romero, Z. Wang, C. Lu, L. Itard, Whole-Building HVAC Fault Detection and Diagnosis with the 4S3F Method: Towards Integrating Systems and Occupant Feedback, in: *The 5th Asia Conference of International Building Performance Simulation Association 2024*, Osaka, Japan, 2024.
- [8] L. van Koetsveld van Ankeren, C. Lu, L. Itard, Implementing diagnostic Bayesian networks for heat recovery ventilation in real-world scenarios: A Dutch case study, *Journal of Building Engineering* 111 (2025) 113527. <https://doi.org/10.1016/j.jobee.2025.113527>.
- [9] C. Lu, S. Walker, C. Struck, L. Itard, D. Saelens, 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 (BuildSys'25)*, Association for Computing Machinery, Golden, CO, USA, 2025.
- [10] C. Lu, L. Itard, 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, 2025. ■