Identifying LLM Applications for Predictive Maintenance

Patrick Timm
Department of Economics, Magdeburg-Stendal University of Applied Sciences, Germany
patrick.timm@h2.de

Fabian Behrendt
Department of Economics, Magdeburg-Stendal University of Applied Sciences, Germany

Sebastian von Enzberg
Department of Engineering and Industrial Design, Magdeburg-Stendal University of Applied Sciences, Germany

Abstract
Predictive maintenance (PdM) offers significant potential for reducing downtime and costs, but in small and medium-sized enterprises (SMEs), its implementation often fails due to heterogeneous equipment landscapes, limited and inconsistent data, and restricted data science resources. At the same time, maintenance-relevant knowledge is distributed across documents and service reports, making it only partially accessible and difficult to use effectively. This preliminary study analyzes the roles that large language models (LLMs) can assume as an assistive technology along the PdM process. Methodologically, the paper is based on a qualitative framework analysis of relevant literature in order to structure the problem space, systematize key fields of application for LLMs in the PdM context, and derive design implications as well as potential development paths toward integrated and agentic support systems.

1. Introduction
A key characteristic of the fourth industrial revolution is the end-to-end connectivity of production through cyber-physical systems, that is, the interaction of physical machines with their digital representations for real-time data acquisition, communication, and (semi-) autonomous process control (Xu et al., 2018). Digital transformation affects not only production itself, but also logistics and value chains in particular, as flexible, connected, and resilient value creation systems are discussed as a central target vision. In this context, methods, models, and AI-based approaches are increasingly seen as key enablers for supporting decision-making and control processes in a data-driven manner (Behrendt et al., 2023). Thanks to the connectivity of diverse machines and their integration with sensor technology, there is considerable potential for data-driven maintenance strategies such as predictive maintenance (PdM) (Zhu et al., 2019). PdM can be applied directly in production environments, but also opens up opportunities in logistical scenarios, for example in maritime contexts. At the same time, practical implementation is often limited by scarce and imbalanced fault data, since relevant failure states occur only rarely and can therefore be observed only in small datasets (Scharr et al., 2023). Another challenge arises from the upcoming generational shift and the retirement of the baby boomer generation. As experienced professionals leave the workforce, there is a risk of losing tacit knowledge unless it is systematically captured and transferred (Anklam, 2002). These challenges increase the need for systems that can consolidate existing information, contextualize it, and translate it into concrete decision support. As an assistive technology, large language models (LLMs) potentially open up new approaches, for example through context-sensitive knowledge retrieval, understandable explanations of PdM results, and the derivation of decision-oriented recommendations along the PdM process. At the same time, the generative capabilities of language models have the potential to substantially support and accelerate development time (Nguyen-Duc et al., 2025).

Against this background, this paper examines which roles and fields of application for LLMs in the PdM context can be derived from the existing literature and how these can be structured within an integrated conceptual framework. While recent reviews have mapped LLM and foundation model applications in machinery health monitoring (Tsallis et al., 2026; Li et al., 2024) and surveyed agentic LLM architectures for PdM (Di Maggio, 2025), these contributions focus on algorithmic capabilities and system architectures rather than on how LLM support functions relate to specific PdM tasks. This paper addresses that gap by

http://dx.doi.org/10.25673/123528

systematically positioning LLM roles along the PdM task taxonomy of Pashami et al. (2023), distinguishing their function as algorithm development support, knowledge formalisation tool, and reasoning engine. The aim is to qualitatively explore the research field, systematize central patterns, and derive design implications as well as development perspectives for future work. The paper is structured as follows. First, the essential foundations of PdM and LLMs are outlined (Chapter 2). Chapter 3 explains the methodological approach of the study. Chapter 4 presents the developed framework as the paper's central contribution. Building on this, Chapter 5 discusses design implications and future development paths, including two illustrative application contexts and an indicative research trajectory.

2. Theoretical Background

This section provides the theoretical foundation of the study. It introduces predictive maintenance as the application context and large language models as the technological basis of this work. Together, these concepts form the conceptual framework for the subsequent analysis.

2.1. Predictive Maintenance

PdM analyzes the expected failures of assets and determines the required maintenance actions to ensure an optimal balance between maintenance frequency and cost. The underlying idea of PdM is that the actual operating condition of systems and components is used to optimize operation and maintenance processes (Zhu et al., 2019). Predictive analysis is based on data collected by sensors attached to machines and tools. Examples of such sensors include vibration sensors and thermal imaging cameras. Prediction algorithms play a central role in information processing within the predictive model (Zhu et al., 2019). They identify trends and thereby determine when equipment needs to be repaired or taken out of service.

PdM can be divided into two primary task areas: diagnostics and prognostics.

Diagnostics can be further divided into anomaly detection, fault detection, fault isolation, and fault identification. Anomaly detection is used to identify deviations from normal operating behavior. It provides an indication of potentially critical changes without assigning a specific fault and serves as an early warning mechanism for detecting unusual patterns. Fault detection determines whether a detected deviation actually represents a fault (Pashami et al., 2023). Within the system, this corresponds to the classification of fault-free and faulty states. Fault isolation refers to narrowing down the faulty component or subsystem. In essence, this involves identifying the affected functional area within the overall system. Fault identification is concerned with determining the specific type and characteristics of the fault. This includes assessing its severity, nature, and stage of development (Pashami et al., 2023).

One task within prognostics is remaining useful life (RUL) estimation. This involves estimating the remaining operating time until functional failure occurs (Pashami et al., 2023). It is a prediction based on degradation models and trends and is continuously updated during asset operation. Another task within prognostics is survival analysis. Its objective is the statistical modeling of failure probabilities over time using historical operating and failure data. In this way, risk profiles for components and systems can be determined (Pashami et al., 2023). Finally, prescriptive maintenance refers to the derivation of concrete recommendations for action based on diagnostic and prognostic results. Its goal is to optimize maintenance timing and resource allocation (Pashami et al., 2023).

2.2. Large Language Models

LLMs are large systems based on neural language models that are trained on extensive text corpora and are therefore capable of processing and generating natural language. The technological foundation of many modern LLMs is the transformer architecture, which organizes sequential language processing through attention mechanisms and thereby established the basis for powerful language modeling (Vaswani et al., 2017). Owing to this capability, LLMs can not only generate text but also extract information, structure content, linguistically organize relationships, and make complex issues accessible in natural language. In industrial contexts, LLMs are therefore particularly relevant because they can mediate between unstructured knowledge sources, technical data repositories, and human interaction (Qu et al., 2025).

In contrast to traditional machine learning methods, which are typically designed for clearly structured input data and narrowly defined target variables, LLMs realize their potential primarily in knowledge-intensive and language-based tasks. In the context of PdM, they should therefore not be understood as a replacement for classical diagnostic and prognostic models, but rather as a complementary assistive technology that enables knowledge processing, contextualization, and decision support along the PdM process (Zou et al., 2025).

For practical deployment, it is also relevant that current LLM applications generally do not consist of the model alone, but are extended by additional system components (Zheng & Xv, 2025). A central role is played by retrieval-augmented generation

(RAG), in which external knowledge sources such as manuals, maintenance logs, or database entries are incorporated into the response process. This can reduce the limitations of model-internal knowledge and tie responses more closely to specific documents and contexts (Chen et al., 2024). In addition, tool-use mechanisms are becoming increasingly important, allowing LLMs to interact with external tools, databases, or software systems. When such capabilities are combined with multi-step workflows, role logic, and goal-directed task execution, the transition toward agentic systems emerges (Di Maggio, 2025). These systems are able to orchestrate different subtasks, integrate information from multiple sources, and derive actionable support from it. This is particularly relevant for PdM, since analytical models, technical experience, documents, and operational constraints must all be considered jointly (Di Maggio, 2025).

At the same time, the use of LLMs is associated with considerable challenges. These include hallucinations, that is, outputs that appear plausible but are factually incorrect, limited explainability, biases in model responses, and security risks arising from the integration of external systems and tools. Especially in industrial application contexts, where recommendations for action may be safety-critical or cost-relevant, this creates the need to consider LLMs as embedded within appropriate governance, security, and control mechanisms (Karne et al., 2025).

3. Methodology

To provide a functional frame for the objective formulated in the introduction, three fields of action were defined in advance, into which different application possibilities can be categorized. These are:

1. LLMs for supporting the development of PdM algorithms,
2. Formalization/structuring of existing knowledge,
3. Reasoning/deriving PdM action recommendations based on existing knowledge.

Building on the objective of developing a big-picture concept and conducting a qualitative analysis of the research field, the framework analysis proposed by Ritchie and Spencer (1994) was selected as the methodological approach. Although this method originates in empirical social research, it has since been applied across a broader range of academic disciplines. Framework analysis follows several steps, beginning with familiarization with the available material. In this study, the selected literature was reviewed in depth to identify central statements, recurring concepts, and typical roles of LLMs in the PdM context.

To build the analytical corpus, an exploratory literature search was conducted in Scopus. Its purpose was to identify relevant contributions that could serve as the material basis for the qualitative structuring of the research field. The search string was:

"Predictive Maintenance" AND ("Large Language Model" OR "LLM")

The search yielded a total of 126 results for the period from 2023 to 2026. Since the aim of this study was not to provide a systematic literature review in the sense of exhaustive coverage, the corpus was not constructed for completeness. Instead, the identified records were screened iteratively based on title, abstract, and, where necessary, full text, and those studies were selected that made a substantial contribution to the previously defined support functions. In addition, backward and forward snowballing were applied to identify key publications that were not fully captured by the database query due to differing terminology or publication venues. This process continued until no substantially new themes or application patterns emerged across additional records, indicating that the selected corpus was sufficiently representative of recurring patterns for the purposes of this qualitative framework analysis. The final analytical corpus comprises 25 publications.

To construct the analytical corpus, the identified records were assessed using a set of pragmatic inclusion and exclusion criteria. Included were studies that showed a clear connection to predictive maintenance and large language models, provided a substantial contribution to potential LLM roles in the PdM context, could be assigned to at least one of the three defined fields of action, and offered sufficient depth for qualitative interpretation. Excluded, by contrast, were contributions without a clear PdM focus, studies in which LLMs were mentioned only marginally, texts with insufficient analytical substance, as well as duplicates or publications of only peripheral relevance to the research field.

In the second step of the framework analysis, recurring and relevant themes were identified. Based on the familiarization process, the analysis was guided by a category system that combined deductive elements—the three pre-defined fields of action—with inductively derived themes that emerged from the material itself. To guard against the risk of circularity, studies were not excluded solely because they did not align with the three fields of action; rather, contributions that qualified or complicated the initial categories were deliberately retained and informed the final framing. The third step involved indexing. Here, the selected contributions were assigned to the relevant

fields of action as well as to core PdM tasks. Where a study fit multiple fields or introduced cross-cutting themes, this was noted explicitly. This was followed by a structured mapping of the field. The categorization of PdM tasks introduced in the theoretical background served as the basis for this representation. Finally, the results were analyzed and interpreted. Based on the framework matrix, the identified patterns were interpreted and consolidated into a coherent conceptual frame. The result is the framework presented in Chapter 4 (Figure 1), which systematically maps LLM roles along PdM core tasks and support functions.

4. Proposed Framework of LLM Roles for PdM

Figure 1 illustrates the support functions of LLMs in the context of predictive maintenance. The three previously identified fields of action can be systematically positioned within this diagram according to their functional role. The machine icons represent the physical assets that are the subject of PdM, while the reference to sensors highlights the data-driven foundation of classical PdM processes. The document symbols at the bottom represent textual and documentation-based knowledge sources, such as maintenance reports or manuals, which can be accessed and structured through LLMs. Finally, the human-in-the-loop element at the end

emphasizes that the derived information and recommendations are to be understood as decision support and remain embedded in human evaluation and decision-making processes. While both the PdM task breakdown (Pashami et al., 2023) and the characterisation of LLM capabilities draw on established literature, the value of the framework lies in their systematic combination: it makes visible, for the first time in a single representation, which specific LLM functions correspond to which PdM task—and thereby identifies where integration potential and research gaps exist along the full diagnostic-to-prescriptive process chain.

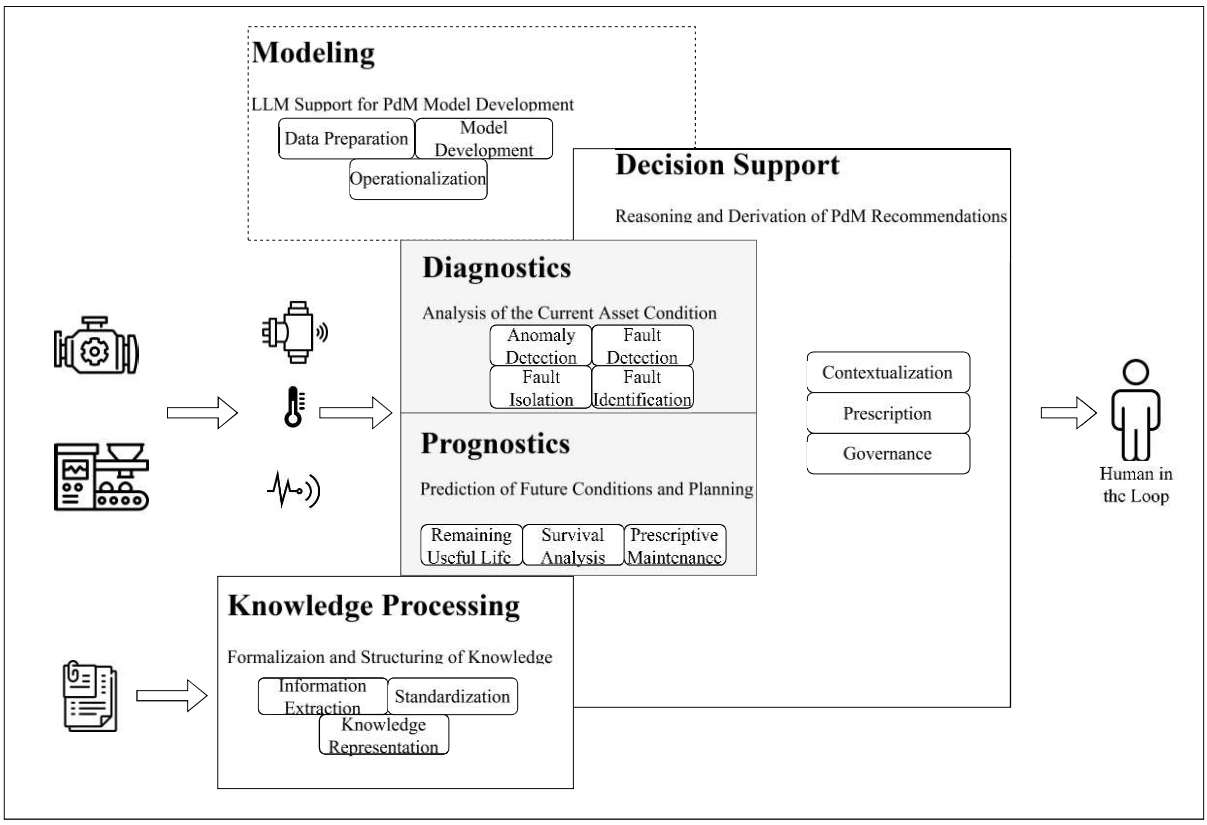

Figure 1: Framework for illustrating the support functions of LLMs in predictive maintenance

Field of action 1: Support for PdM development
This field of action includes all contributions aimed at making PdM models and related pipelines more efficient, more robust, and more suitable for operational deployment. It therefore refers not only to model development itself, but also to upstream and downstream tasks in data, feature, and system processing. In this context, LLMs realize their potential primarily as cognitive assistive systems and in the sense of agent-based orchestration, for example by supporting analysis processes, code generation, documentation, and experiment planning. However, the quality of the resulting PdM models continues to be determined primarily by data quality, domain adequacy, and empirical evaluation (Caesarendra & Tjahjowidodo, 2017). Within this field of action, three central task areas can be distinguished.

First, data preparation includes the systematic qualification and preprocessing of PdM data. An industrial case study explicitly shows that data-

driven preprocessing and analysis form the basis for both decision support and the development of robust predictive models (Bekar et al., 2020). In addition, LLM-based agent systems are already being evaluated in PdM-related application areas, for example for the automated correction of typical errors in maintenance logs (Dimidov et al., 2026).

Second, model development refers to the structured creation of suitable representations and models. The condition monitoring literature shows that different fault patterns require different forms of signal representation and feature engineering (Caesarendra & Tjahjowidodo, 2017). Accordingly, model development is understood in relevant PdM frameworks as an engineering process that integrates feature engineering, model selection, and validation (Ucar et al., 2024). LLMs do not replace actual feature learning or the model's generalization capability, but they can effectively support development through systematic code generation, prototypical implementation, and result synthesis (Mumuni & Mumuni, 2025).

Third, operationalization refers to the transfer of PdM approaches into productive, operationally viable systems. The MLOps literature describes this phase as a central challenge of industrial machine learning applications (Kreuzberger et al., 2023). For PdM, cloud-to-edge architectural patterns have been proposed that integrate edge gateways, data storage, and services for analytics, visualization, and scheduling (Bowden et al., 2019). The relevance of LLMs and agentic systems increases particularly when RAG indices, document integration, and tool use are additionally embedded into the system architecture. In such scenarios, security and governance layers must be explicitly considered from the outset (Hong et al., 2024).

Field of action 2: Formalization of maintenance knowledge

The second field of action addresses the systematic extraction, structuring, and operational use of maintenance knowledge. In industrial practice, relevant knowledge is continuously generated, but it often exists in textual, heterogeneous, and semi-structured form. One contribution explicitly describes maintenance work orders as "black holes," emphasizing that they represent valuable knowledge resources that remain insufficiently accessible for decision-making processes (Sundaram & Zeid, 2024). Systematic reviews show that natural language processing in the context of industrial maintenance constitutes a distinct and growing line of research, since maintenance reports can be leveraged for classification, information extraction, similarity search, and knowledge management (Zhong et al., 2024). LLMs are particularly well suited to this field because they can transform unstructured language into knowledge units that can be systematically organized.

Three central task areas can also be identified within this field of action.
First, information extraction aims to derive relevant information on fault patterns, causes, and resulting actions from textual sources such as service reports or manuals. An application example from the wind energy sector shows that the extraction of key performance indicators (KPIs) can both improve efficiency and support the calculation of reliability metrics.

Second, standardization refers to the harmonization of terms, synonyms, abbreviations, and codes. Without such harmonization, analyses often remain limited to site-specific or team-specific contexts and are only transferable to a limited extent. The respective study highlights this issue and points, among other things, to the linking of maintenance problems with standardized codes in the aviation context (Sundaram & Zeid, 2024).

Third, knowledge representation refers to the structured transformation of extracted knowledge into ontologies, knowledge graphs, or RAG-based knowledge structures. One study explicitly argues that the "liberation and formalization" of knowledge from work orders can significantly increase the intelligence of maintenance management systems through knowledge reuse (Tahsin et al., 2024). It describes a process that includes ontological modeling as well as LLM support for the further development of PdM. In addition, the maintenance reference ontology iof-maint highlights the importance of semantic interoperability for industrial maintenance data and related applications (Hodkiewicz et al., 2024). Beyond that, the smart manufacturing literature emphasizes that knowledge graphs can serve as a unifying semantic layer between heterogeneous data sources, processes, and decision logics (Wan et al., 2024).

Field of action 3: Reasoning and decision support

The third field of action focuses on translating PdM insights into actionable decisions. Although PdM generates predictions in the form of anomaly indications, condition indicators, or remaining useful life (RUL) estimates, its operational value is only realized when this information can be transformed into concrete actions, for example with regard to maintenance windows, prioritization, personnel and material allocation, or risk control. A literature review on decision-making in PdM emphasizes that, alongside failure prediction, decision algorithms are increasingly moving into focus, as they address precisely this bridge between prediction and action (Bousdekis et al., 2019). Agentic LLM approaches explicitly address this

gap by combining RAG and tool use to generate context-specific recommendations for action (Deng et al., 2024).

Three task areas can also be distinguished within this field of action.
First, contextualization refers to the integration of observations from sensors and symptom patterns with technical background knowledge, such as manuals, fault codes, or historical cases. One study shows that troubleshooting is heavily shaped by manuals and expert experience. At the same time, it compares different RAG configurations with LLM-only and retrieval-only approaches and emphasizes the relevance of an evaluation-first principle. It also becomes clear that retrieval-based approaches can, in some more label-bound tasks, outperform purely generative LLM approaches (Lukens et al., 2025).
Second, prescription refers to the derivation of concrete actions from available predictions while taking technical and operational trade-offs into account. One paper identifies exactly this translation process as a central challenge and uses agentic LLMs in combination with RAG and tool use for the strategic generation of recommendations for action (Deng et al., 2024). In addition, the agentic framework MindRAG integrates LLM-based reasoning with multimodal retrieval for condition monitoring in order to reduce false alarms, assess criticality, and provide explainable user interfaces, among other functions (Löwenmark et al., 2025).

Third, governance addresses the requirements of explainability, auditability, and safety in industrial application contexts. One relevant contribution points to the fundamental risks of large language models and thereby justifies the need for a cautious and reflective integration of such systems (Bender et al., 2021). In addition, security research shows that LLM-integrated systems can be vulnerable to indirect prompt injection, since the boundary between data and instructions becomes blurred. For PdM agents with tool calls, this represents a concrete architectural risk (Greshake et al., 2023). With the AI Risk Management Framework (AI RMF 1.0) and the corresponding profile for generative AI, practical guidelines already exist that can be adapted as a governance layer for industrial LLM systems (Tabassi, 2023).

5. Design Implications and Future Development Paths

The framework analysis shows that, to date, LLMs in the context of PdM have primarily been examined in clearly delimited support functions. In particular, contributions can be identified that focus on supporting PdM development, formalizing maintenance knowledge, and reasoning-based decision support. However, the actual value for operational PdM processes is likely to lie in the interaction of these functions along real process and decision chains (Romero & Suyama, 2025). The developed framework can therefore be read, first, as a structuring of the current state of research and, second, as a starting point for a broader development perspective.

Against this background, a development trajectory can be outlined from isolated LLM applications toward agentic PdM systems. Such a system would not be limited to the role of a language model with supplementary analytical functions, but should instead be conceived as a socio-technical support architecture in which different functions are coordinated and integrated (Di Maggio, 2025). These functions include, for example, the processing and contextualization of sensor data, the extraction of document-based knowledge from manuals, maintenance reports, and service histories, the use of external tools and interfaces, and the target-group-specific preparation of action-relevant information. The transition toward agentic systems thus lies primarily in the architectural and functional integration of components that have so far largely been considered separately (Di Maggio, 2025).

From this perspective, several design implications emerge for the conception of agentic PdM systems. First, a process- and context-oriented development approach appears necessary (Farahani et al., 2025). Second, agentic PdM systems must integrate different forms of knowledge, including structured condition and failure data, unstructured documents, and tacit domain knowledge (Wulf et al., 2025). Third, requirements such as explainability, safety, governance, and human oversight must be treated as constitutive design dimensions (Borghoff et al., 2025). To render these implications more tangible, two potential application contexts are briefly outlined. A first context is a sheet-metal manufacturing company operating a heterogeneous machine fleet, where maintenance decisions are currently largely experience-driven and relevant knowledge is distributed across paper-based service logs. An LLM-based system could support data preparation (field 1), structure fault patterns from historical work orders (field 2), and generate plain-language maintenance recommendations for operators (field 3). A second context is a company that develops and distributes field measurement devices used, for example, for soil compaction testing. Maintenance-relevant knowledge is embedded in technical documentation and service histories; LLMs could support the structuring of this distributed knowledge (field 2) and assist in deriving context-sensitive service recommendations (field 3). Both contexts represent potential application partners for future empirical work, though the concrete research design remains to be developed. Building on this, an indicative research trajectory can be sketched. An initial step would involve translating the framework into a more

concrete system concept, including a domain-specific requirements analysis and a modular architecture integrating the three fields of action. A subsequent step could involve the iterative development and evaluation of a functional prototype, assessing both technical performance and user-oriented aspects such as comprehensibility and acceptance. In a further step, the approach could be validated in real operational settings with the identified application partners. A key challenge for future work therefore lies not only in the technical integration of the identified LLM support functions, but in developing adequate methods for specifying agentic PdM systems – capturing functional, organisational, and governance requirements in a form suitable for systematic design and evaluation. From the analysis, three overarching research needs can be derived: first, the clarification of concepts and requirements for agentic PdM systems; second, methodological support for their specification and design; and third, guidance for their integration and evaluation within existing production, maintenance, and IT contexts. The present paper addresses the first of these needs and thereby provides a conceptual starting point for subsequent research.

Building on this, the analysis points to a broader development trajectory from isolated LLM applications toward integrated, agentic PdM systems. Such systems would need to be specified along multiple dimensions – including goals, agent roles, degrees of autonomy, decision and action spaces, tools, safety boundaries, and human-agent interactions – that go well beyond the scope of classical PdM model development. This raises fundamental questions of requirements engineering, system architecture, and organisational integration that existing approaches do not yet address in a domain-specific manner.

6. Limitations and Conclusion

This paper examines which roles and fields of application of LLMs in the context of PdM can be derived from the existing literature and how they can be structured within an overarching conceptual framework. Based on a qualitative framework analysis, three central fields of action were identified: support for PdM development, the formalization of maintenance knowledge, and reasoning-based decision support. The resulting framework shows that LLMs in the PdM context should not be viewed solely as isolated assistive tools, but can instead be systematically positioned across different task areas.

Beyond this structuring function, the analysis also points to an integration potential between support functions that have so far been considered largely in isolation. This suggests a development trajectory from isolated LLM applications toward more integrated and, potentially, agentic support systems. However, the contribution of this paper does not lie in the design of such a system architecture, but rather in the conceptual structuring of the field and the identification of starting points for further research.

At the same time, the explanatory power of the study is limited by several factors. First, the analysis is based on an exploratory qualitative literature synthesis rather than a fully systematic review. Accordingly, no claim can be made to completeness or representativeness of the literature corpus considered. It must be acknowledged that the patterns identified and the development trajectory outlined carry a degree of interpretive inference that exceeds what the evidential basis strictly warrants; the conclusions should therefore be read as reasoned propositions rather than empirically established findings. Second, although measures were taken to include studies that qualified or complicated the initial framework categories, a confirmation bias cannot be fully excluded. Third, the research field is characterized by a high degree of dynamism, reflected in the emergence of new terminology, the strong role of preprints, and rapidly shifting thematic emphases. Fourth, the framework developed in this study has not been empirically validated. It should therefore not be understood as a definitive representation of the field, but rather as a conceptual frame of reference whose robustness must be examined in future empirical and design-oriented studies.

Overall, this paper should be understood as a structuring contribution to a still emerging field of research. The developed framework condenses central application areas of LLMs in the PdM context, makes their interrelationships visible, and provides a foundation for future work concerned with the design, integration, and evaluation of more advanced support systems.

References

Anklam, P. (2002). Knowledge management: the collaboration thread. Bulletin of the American Society for Information Science and Technology, 28(6),8-8. https://asistdl.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/bult.254

Behrendt, F., Schmidtke, N., Wollert, T., & Weigert, D. (2023). Digital transformation and future impacts in logistics and value chains. In J. Boukachour & A. Benaini (Eds.), Transport and logistics planning and optimization (pp. 1–36). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-6684-8474-6.ch001

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In Proceedings of the 2021 ACM

conference on fairness, accountability, and transparency (pp. 610-623).

Bekar, E. T., Nyqvist, P., & Skoogh, A. (2020). An intelligent approach for data pre-processing and analysis in predictive maintenance with an industrial case study. Advances in Mechanical Engineering, 12(5), 1–14. https://doi.org/10.1177/1687814020919207

Borghoff, U. M., Bottoni, P., & Pareschi, R. (2025). Human-artificial interaction in the age of agentic AI: a system-theoretical approach. Frontiers in Human Dynamics, 7, 1579166. https://doi.org/10.3389/fhumd.2025.1579166

Bousdekis, A., Lepenioti, K., Apostolou, D., & Mentzas, G. (2019). Decision making in predictive maintenance: Literature review and research agenda for Industry 4.0. IFAC-PapersOnLine, 52(13), 607–612. https://doi.org/10.1016/j.ifacol.2019.11.226

Bowden, D., Marguglio, A., Morabito, L., Napione, C., Panicucci, S., Nikolakis, N., Makris, S., Coppo, G., Andolina, S., Macii, A., Macii, E., O'Mahony, N., Becker, P., & Jung, S. (2019). A cloud-to-edge architecture for predictive analytics. In Proceedings of the Workshops of the EDBT/ICDT 2019 Joint Conference. CEUR Workshop Proceedings. https://serena-project.eu/wp-content/uploads/2019/02/A-cloud-to-edge-architecture-for-predictive-analytics.pdf

Caesarendra, W., & Tjahjowidodo, T. (2017). A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing. Machines, 5(4), 21. https://doi.org/10.3390/machines5040021

Chen, J., Lin, H., Han, X., & Sun, L. (2024, March). Benchmarking large language models in retrieval-augmented generation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 16, pp. 17754-17762). https://doi.org/10.1609/aaai.v38i16.29728

Deng, H., Namoano, B., Zheng, B., Khan, S., & Ahmet Erkoyuncu, J. (2024). From prediction to prescription: Large language model agent for context-aware maintenance decision support. PHM Society European Conference, 8(1), 10. https://doi.org/10.36001/phme.2024.v8i1.4114

Di Maggio, L. G. (2025). Toward autonomous LLM-based AI agents for predictive maintenance: State of the art, challenges, and future perspectives. Applied Sciences, 15(21), 11515. https://doi.org/10.3390/app152111515

Dimidov, V. I., Hawlader, F., Jafarnejad, S., & Frank, R. (2026). Cleaning maintenance logs with LLM agents for improved predictive maintenance. Annual Conference of the PHM Society, 5(1). https://doi.org/10.36001/phmap.2025.v5i1.4486

Farahani, M. A., Khan, M. I., & Wuest, T. (2025). Hybrid Agentic AI and Multi-Agent Systems in Smart Manufacturing. *arXiv preprint arXiv:2511.18258*. https://doi.org/10.48550/arXiv.2511.18258

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection. Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, 1046–1060. https://doi.org/10.1145/3605764.3623985

Hodkiewicz, M., Woods, C., Selway, M., & Stumptner, M. (2024). Iof-maint: Modular maintenance ontology [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2404.05224

Hong, S., Shin, J., Seo, J., Lee, T., Park, J., Cho, M., Choi, B., & Lim, H. (2024). Intelligent predictive maintenance RAG framework for power plants: Enhancing QA with StyleDFS and domain specific instruction tuning. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track (pp. 805–820). Association for Computational Linguistics.

Karne, R., Pativada, P. K., & Dudhipala, A. (2025). Hallucinations in large language models (LLMs): Challenges in mitigation trust and future directions. Indian Journal of Computer Science and Engineering, 16(1). DOI : 10.21817/indjcse/2025/v16i1/251603011

Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. IEEE Access, 11, 31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138

Li, Y., Wang, H., & Sun, M. (2024). ChatGPT-like large-scale foundation models for prognostics and health management: A survey and roadmaps. Reliability Engineering & System Safety, 243, Article 109896. https://doi.org/10.1016/j.ress.2023.109896

Löwenmark, K., Strömbergsson, D., Liu, C., Liwicki, M., & Sandin, F. (2025). Agent-based condition monitoring assistance with multimodal industrial database retrieval augmented generation [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.09247

Lukens, S., Bishof, M., Siddiqui, N., & West, D. (2025). An evaluation framework for fault diagnosis using technical manuals in retrieval-augmented large language models. Annual Conference of the PHM Society, 17(1). https://doi.org/10.36001/phmconf.2025.v17i1.4549

Mumuni, A., & Mumuni, F. (2025). Automated data processing and feature engineering for deep learning and big data applications: A survey. Journal of Information and Intelligence. https://doi.org/10.1016/j.jiixd.2024.01.002

Nguyen-Duc, A., Cabrero-Daniel, B., Przybylek, A., Arora, C., Khanna, D., Herda, T., ... & Abrahamsson, P. (2025). Generative artificial

intelligence for software engineering—A research agenda. Software: Practice and Experience, 55(11), 1806-1843.

Pashami, S., Nowaczyk, S., Fan, Y., Jakubowski, J., Paiva, N., Davari, N., ... & Gama, J. (2023). Explainable predictive maintenance. arXiv preprint arXiv:2306.05120. https://doi.org/10.48550/arXiv.2306.05120

Qu, Y., Huang, S., Li, L., Nie, P., & Yao, Y. (2025). Beyond Intentions: A Critical Survey of Misalignment in LLMs. *Computers, Materials & Continua*, *85*(1). DOI:10.32604/cmc.2025.067750

Ritchie, J., & Spencer, L. (1994). Qualitative data analysis for applied policy research. In A. Bryman & R. G. Burgess (Eds.), Analyzing qualitative data (pp. 173–194). Routledge. https://doi.org/10.4324/9780203413081_chapter_9

Romero, M. L., & Suyama, R. (2025, October). Agentic ai for intent-based industrial automation. In *2025 16th IEEE International Conference on Industry Applications (INDUSCON)* (pp. 437-444). IEEE. https://doi.org/10.48550/arXiv.2506.04980

Scharr, C., Heinze, C., & Geist, M. (2023, October). Technical communication on engine fault detection with small datasets. In 2023 8th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 52–56). IEEE. https://doi.org/10.1109/ICFSP59764.2023.10372919

Sundaram, S., & Zeid, A. (2024). Technical language processing for prognostics and health management: Applying text similarity and topic modeling to maintenance work orders. Journal of Intelligent Manufacturing. Advance online publication. https://doi.org/10.1007/s10845-024-02323-4

Tahsin, R., Li, Y., Abolhasani, M. S., & Ameri, F. (2024). Generation of semantic knowledge graphs from maintenance work order data. Semantic Web Journal. https://www.semantic-webjournal.net/system/files/swj3658.pdf

Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Tsallis, C., Papageorgas, P., Munteanu, R. A., & Dellagi, S. (2026). Large language and foundation models for machinery health monitoring: A systematic review. Applied Sciences, 16(5), 2493. https://doi.org/10.3390/app16052493

Ucar, A., Karakose, M., & Kırımça, N. (2024). Artificial intelligence for predictive maintenance applications: Key components, trustworthiness, and future trends. Applied Sciences, 14(2), 898. https://doi.org/10.3390/app14020898

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wan, Y., Liu, Y., Chen, Z., Chen, C., Li, X., Hu, F., & Packianather, M. (2024). Making knowledge graphs work for smart manufacturing: Research topics, applications and prospects. Journal of Manufacturing Systems, 76, 103–132. https://doi.org/10.1016/j.jmsy.2024.07.009

Wulf, J., Meierhofer, J., & Hannich, F. (2025). Architecting Human-AI Cocreation for Technical Services--Interaction Modes and Contingency Factors. *arXiv preprint arXiv:2507.14034*. https://doi.org/10.48550/arXiv.2507.14034

Xu, L. D., Xu, E. L., & Li, L. (2018). Industry 4.0: State of the art and future trends. International Journal of Production Research, 56(8), 2941–2962. https://doi.org/10.1080/00207543.2018.1444806

Zheng, J., & Xu, D. (2025). Multi-Level Safety Filtering Method for LLM Generation Security. *Journal of Computer Science and Frontier Technologies*, *1*(3), 76-90. DOI: 10.63313/JCSFT.9025

Zhong, K., Jackson, T., West, A., & Cosma, G. (2024). Natural language processing approaches in industrial maintenance: A systematic literature review. Procedia Computer Science, 232, 2082–2097. https://doi.org/10.1016/j.procs.2024.02.029

Zhu, T., Ran, Y., Zhou, X., & Wen, Y. (2019). A survey of predictive maintenance: Systems, purposes and approaches [Preprint]. arXiv. https://arxiv.org/abs/1912.07383

Zou, H. P., Huang, W. C., Wu, Y., Chen, Y., Miao, C., Nguyen, H., ... & Yu, P. S. (2025). Llm-based human-agent collaboration and interaction systems: A survey. arXiv preprint arXiv:2505.00753. https://doi.org/10.48550/arXiv.2505.00753