Potential of Generative Artificial Intelligence in Knowledge-Based Predictive Maintenance for Aircraft Engines

Meriem HAFSI

CESI LINEACT, 15C Av. Albert Einstein, 69100 Villeurbanne, France
mhafsi@cesi.fr

ABSTRACT

Predictive maintenance based on remaining useful life (RUL) estimation is widely recognized as a promising strategy for monitoring the health of critical systems such as aircraft engines, anticipating failures, and optimizing maintenance planning. A variety of approaches have been proposed in the literature, including data-driven, physics-based, and knowledge-based methods. Among them, deep learning-based methods have shown strong performance and gained the most traction, but industrial adoption remains limited due to challenges in interpretability, scalability and adaptability. Recent advances in generative artificial intelligence (GAI) offer new opportunities to address challenges related to data scarcity and variability but issues of model transparency persist. In this context, our paper highlights how these recent advances could open new opportunities, especially when integrated within hybrid frameworks combining data-driven and knowledge-based reasoning. By clarifying industrial requirements and open challenges, this work provides a comprehensive synthesis of current needs and outlines a framework establishing a methodological foundation for producing interpretable RUL estimates along with the rules guiding the reasoning process.

1. INTRODUCTION

In recent decades, predictive maintenance (PdM) has become a strategic solution to detect anomalies and anticipate failures in industrial equipment and sophisticated machines. This strategy relies on the continuous collection of multi-sensor data and performance indicators to feed machine learning (ML) algorithms capable of identifying early signs of malfunction, thereby enabling preventive interventions and reducing downtime (de Pater, Reijns, & Mitici, 2022). In the literature, three main approaches are described. Physics-based methods require accurate mathematical modeling of the physical degradation processes involved, but they are often difficult to apply to complex systems where physical laws are hard to formalize. Data-driven methods dominate current implementations due to their ability to learn complex patterns from large datasets. However, they suffer from several limitations, such as a lack of interpretability, dependence on large amounts of labeled data, and poor generalization to new operating conditions. Knowledge-based approaches, on the other hand, are more explainable and rely on expert-defined rules, but they struggle with adaptability and scalability in dynamic or uncertain environments. This study investigates the potential of GAI to address these limitations by supporting the creation of hybrid PdM models that combine the strengths of both data-driven and knowledge-based approaches. GAI offers new capabilities to simulate realistic failure scenarios, augment limited datasets, and extract structured knowledge from unstructured technical sources (Lang, Peng, Cui, Yang, & Guo, 2021). It can also support the construction of domain-specific knowledge graphs or ontologies —formal representations of concepts and their relationships in a given domain— by identifying relevant concepts and semantic relationships, as well as generating logical reasoning rules based on expert input or technical resources. In addition, GAI shows promising potential in assisting the physical modeling of complex systems by proposing plausible approximations or surrogate models when traditional analytical modeling proves difficult. To support this investigation, the study addresses a use case on RUL estimation for aircraft engines and introduces a methodological framework that combines data-driven approaches with domain knowledge and symbolic reasoning to enable interpretable and reliable predictions.

2. PROBLEM STATEMENT

The emergence of Industry 4.0 has led to the development of several innovative applications, such as PdM. This trend is further driven by the vast amounts of data produced by modern systems and IoT-based monitoring technologies. While these advancements are promising, their implementation often depends on accurately estimating the RUL of components or systems. However, deploying such solutions comes with a range of challenges. First, several data-related issues arise: large volumes of data are required to train robust models capable of generalizing. The datasets used must be rich, containing diverse and rare failure events, elements that are often missing in existing datasets. Additionally, the data must be complete and of high quality. However, in industrial settings, data are frequently noisy and incomplete. This makes preprocessing necessary, but it also introduces uncertainty, which can compromise the training of reliable models. The lack of complete data and the absence of detailed descriptions of various faults and anomalies represent a major obstacle to the implementation of such reliable approaches.

ML and deep learning (DL) have enabled the development of advanced solutions for RUL prediction. Despite the high levels of accuracy achieved by some methods, their deployment in real-world scenarios remains limited due to a lack of robustness, generalization, and scalability. Beyond the challenges related to training data previously mentioned, the black-box nature of these models, limiting their explainability, as well as their failure to account for non-stationary environments and their rigidity regarding specific data formats, significantly constrain their applicability (Benatia, Hafsi, & Ayed, 2025). As a result, the adoption of these solutions by domain operators is hindered by a lack of trust in the decisions made by such systems, particularly in complex systems.

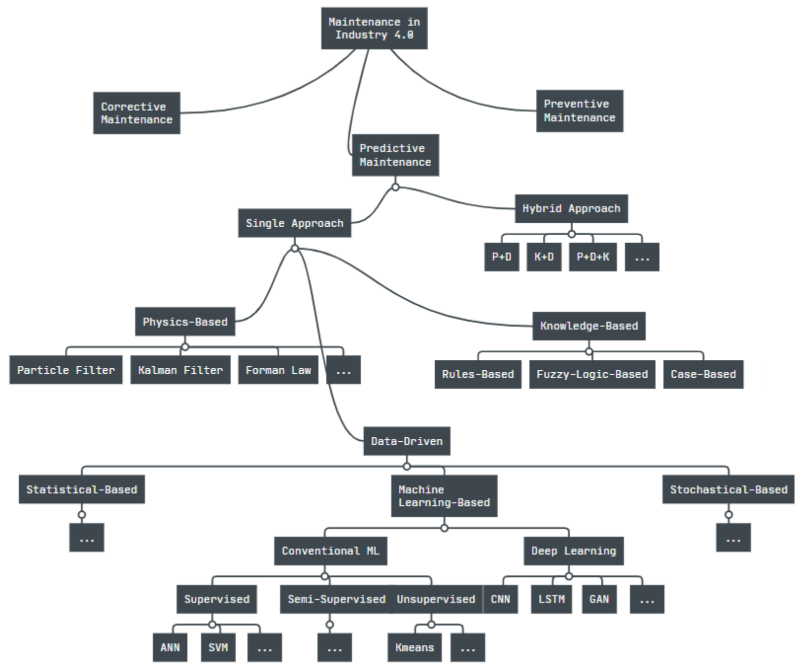

Figure 1. PdM approaches classification.

2.1. Existing Approaches

To anticipate failures, optimize maintenance operations, and reduce costs associated with breakdowns (Davari et al., 2021), PdM aims to predict the RUL of a component or system using four main approaches (Zio, 2022):
(1) Knowledge-based approach Uses expert rules and past failure patterns to anticipate issues, refining models over time to improve decisions and adapt to new conditions (Işık, 2023). (2) The physics-based approach estimates degradation through physical laws and simulations, requiring deep mathematical knowledge for precise component analysis (Nunes, Santos, & Rocha, 2023; Hagmeyer, Zeiler, & Huber, 2022). (3) data-driven approach leverages large datasets and uses statistical models, ML or DL algorithms to predict failures and estimate RUL. It identifies patterns in sensor data without relying on physical models. However, its effectiveness heavily depends on the quality and availability of the data (Eker, Camci, & Jennions, 2019; Santiago et al., 2024). (4) The hybrid approach combines multiple methods to improve accuracy and address data limitations or complex scenarios. It seeks to balance the strengths and mitigate the weaknesses of the individual approaches (Traini, Bruno, & Lombardi, 2021; Barry & Hafsi, 2023). Figure 1 provides a summary of maintenance strategies, as well as the various approaches and methods for PdM within the context of Industry 4.0.

Each PdM approach offers distinct advantages but also faces notable limitations as summarized in Table 1. Data-driven models are highly valued for their adaptability and scalability, especially in modern industrial environments where real-time predictions are essential. Their ability to process large datasets allows them to uncover subtle patterns in sensor data, often missed by traditional approaches. However, their performance heavily depends on the quality, quantity, and completeness of the data. In practice, industrial data are often noisy, incomplete, or imbalanced, which can significantly affect model reliability. Moreover, these models typically lack interpretability, making it difficult to explain predictions or build trust in critical applications. Physics-based models, on the other hand, provide high accuracy and are grounded in a deep understanding of the physical mechanisms of failure. Yet, they require extensive domain expertise and are often difficult to scale or adapt to varying conditions. Their development can be time-consuming and cost-intensive, especially when applied to complex systems with dynamic behaviors. Hybrid approaches attempt to combine the strengths of two or more models to improve prediction accuracy and robustness. While promising in theory, they often come with increased complexity in terms of conception, integration and validation. Amid these challenges, the knowledge-based approach emerges as a complementary solution. It incorporates expert knowledge drawn from operators, technicians, and domain specialists, into the prognostic process. Models such as rule-based systems, case-based reasoning, ontologies, or knowledge graphs offer interpretability, transparency, and traceability of the reasoning behind predictions. This is especially valuable in high-stakes environments where understanding the rationale behind a decision is as important as the deci-

Approach	Advantage	Limitations
Knowledge Based	Interpretability and understanding of the decision-making process (Rosati et al., 2023)	Expensive and time-consuming for implementation (Cao, Zanni-Merk, Samet, De Beuvron, & Reich, 2020)
Physics Based	Physical models provide the most accurate and precise estimation of the RUL (Sikorska, Hodkiewicz, & Ma, 2011). Explainability.	Deep understanding of failure mechanism (Cao et al., 2022). High mathematics and physics skills needed.
Data Driven	No knowledge of physical models or expert knowledge needed. Make use of the I4.0 technology.	Takes time and resources to train. Black-box effect.
Hybrid	Improves the accuracy by leveraging the strengths of different approaches	Difficult to implement. Fusion and uncertainty management.

Table 1. Overview of PdM approaches and their effectiveness.

sion itself. However, this approach is not without challenges. The main difficulty lies in capturing accurate, structured, and comprehensive expert knowledge. Access to domain experts can be limited, and formalizing their experience into usable models is often complex. Despite this, the explicit nature of the knowledge-based approach and its ability to function with limited data make it a strong candidate for reliable and explainable prognostics, especially when combined with other approaches in a human-in-the-loop framework. Finally, while no single approach is universally superior, the knowledge-based method stands out for its interpretability, adaptability in data-scarce environments, and potential to complement other models, making it a cornerstone in the development of robust, transparent PdM systems.

3. GENERATIVE AI FOR PDM

3.1. Principles

GAI is a class of AI techniques capable of automatically generating content such as synthetic data instances, insights, or decisions from existing datasets, supporting tasks like design optimization, PdM, and intelligent automation in industrial systems. The three common generative models in literature are Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Transformers.

Generative Adversarial Networks (GANs): are models composed of two neural networks: a generator, which creates synthetic data samples, and a discriminator, which evaluates their authenticity. GANs are commonly used in industrial applications to generate synthetic sensor data and simulate fault scenarios, thereby improving the performance of predictive models. They offer the ability to explore distribution features from real data and extrapolate them to unseen operation conditions through a two-player game between the generator and the discriminator (X. Y. Li, Cheng, Fang, Zhang, & Wang, 2024).

Variational Autoencoders (VAEs): are probabilistic models that learn latent representations by encoding data inputs into a lower-dimensional space and decoding them back allowing the generation of new data samples. VAEs are particularly effective in modeling normal system behavior, capturing degradation patterns, and generating operational states for simulation in industrial applications.

Transformers: are models that use attention mechanisms to model complex dependencies in sequential data or text. Originally designed for natural language processing (NLP), the Transformer architecture has opened the way for a new generation of Large Language Models (LLMs) that demonstrate remarkable capabilities in understanding and generating human-like text. Notable examples include OpenAI's GPT series (GPT-3, GPT-4), Anthropic's Claude, and emerging models like DeepSeek. These LLMs excel in a variety of tasks, from text generation and summarization to code synthesis and conversational AI (Klekowicki, Szymański, Waligórski, & Misztal, 2024). Another inspiration concerns autoregressive models like GPT for text generation, encoder-decoder models such as BART and T5 for translation and summarization, and specialized variants for handling long sequences (Transformer-XL), multimodal data fusion (e.g., Vision Transformers), and time-series forecasting. This diversity enables Transformers to address complex generative tasks including language, vision, audio, and sensor data processing, enabling powerful industrial applications like PdM and enhanced human-machine interaction.

3.2. GAI in Data-Driven PdM

The capabilities of GAI have naturally led to its integration in addressing data-related challenges in RUL estimation and PdM tasks. (Mohapatra, 2024) explores the transformative potential of GAI for anomaly detection and failure prediction. GANs and VAEs are identified as well-suited for these applications, with several key advantages highlighted, including the simulation of equipment behavior, the generation of synthetic data, and the optimization of maintenance schedules. (Khan, Nasim, & Rasheed, 2025) examines various aspects of GAI-powered PdM in industrial systems, especially in aircraft systems. The study demonstrates how GAI can be leveraged to extract key features for building classifiers and predictive models, thereby facilitating effective data representation learning. GAI's potential to address data scarcity and imbalance is highlighted through three key tasks: data augmentation, missing data imputation, and feature extraction for predictive modeling. To this end, GANs, VAEs, or hybrids combining both are identified as the most commonly used GAI models in this context. Despite their advantages, challenges remain in model interpretability, integration, and knowledge transfer, which are critical for real-world deployment and the effective adoption of such models in aeronautics.

3.2.1. Data augmentation

In real industrial contexts, limited or imbalanced data hinder the development of robust PdM models. Data augmentation mitigates this challenge by generating new samples from existing ones, thereby increasing dataset size and diversity (Khan et al., 2025). It improves model generalization and robustness, especially under data scarcity. With the advent of GAI, augmentation extends beyond simple perturbations to produce realistic, semantically meaningful variations. Leveraging models such as GANs or VAEs, GAI enables task-specific, distribution-aware synthesis that supports the training of high-performing predictive models. (Lang et al., 2021) address the lack of run-to-failure data for accurate RUL prediction model and propose a data augmentation approach. A GAN learns the distribution of the original dataset and generate a synthetic training set. The original and generated datasets are then fused to train a Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM) hybrid model for RUL prediction. Experimental on the original and augmented C-MAPSS dataset demonstrate the effectiveness of the proposed method. (Campbell, Ilangovan, Gregory, & Mikaelian, 2022) explores the integration of GAI for synthetic generation of realistic flight data within NASA's Integrated Cognitive Modeling (ICM) program. The objective is to augment flight dynamics datasets and reduce simulation costs while covering critical flight scenarios. GANs and advanced variants such as BigGAN are suggested for their ability to produce diverse, high-quality synthetic data. These models are proposed as potential substitutes for traditional simulation pipelines, pending validation. The study also examines other architectures, including DCGANs, which learn complete flight maneuvers via convolutional layers, and cGANs, which generate scenario-specific data reflecting conditions like loss of control, system failures, or envelope violations. (X. Y. Li et al., 2024) highlights the challenges faced by GAN-based data augmentation methods in real environments, where time-series data are often multivariate, multiscale, and structurally complex. In such contexts, the inherent instability of the generative process frequently results in synthetic data that fail to meet both feature diversity and usability requirements. To address these limitations, a framework for aeroengine multitask prognosis is proposed, grounded in degradation behavior extrapolation and a trade-off between diversity and usability. A First Prediction Time (FPT) identification method is introduced, leveraging a combination of the Health Index (HI) and its volatility. The augmentation strategy is built upon the Dual Discriminator Time-series GAN (DDTGAN) coupled with a Negative Sample Elimination (NSE) mechanism. Specifically, DDTGAN is designed to extrapolate degradation behaviors while capturing multiscale temporal features across global and local domains. NSE subsequently filters out low-quality samples, ensuring that only high-value synthetic data are retained. Finally, the authors us an adaptive Transformer-Multi-gate Mixture-of-Experts (T-MMOE) to incorporates a gradient normalization-based joint loss function to dynamically balance RUL prediction and fault diagnostic tasks. GANs have proven effective for addressing data scarcity and imbalance (Konig, Wagner, Liebschner, & Kley, 2025) and are widely applied to fault diagnostic in rotating machinery (W. Huang, Zhang, Jiang, Shao, & Bai, 2025), with recent methods such as CBAM-MVACGAN combining a Multi-Scale Convolutional Block Attention Mechanism with Minimum Variance-Assisted Classification GANs. In this context, (Z. Li, Jiang, & Wang, 2025) integrates an Equilibrium Deep Q-Network based agent with a Variational Autoencoder–Wasserstein GAN with Gradient Penalty (VAE-WGAN-GP) to address the challenges of data imbalance and limited sample availability caused by constraints in sampling and labeling. The hybrid model enables the generation of high-quality synthetic samples while ensuring class distribution balance and enhanced feature representation, thereby improving diagnostic performance under real-world constraints.

3.2.2. Synthetic data generation

Synthetic data generation has undergone significant advancements with the emergence of GAI, which enables the creation of entirely new data instances based on learned distributions. By generating realistic time series data, this approach is particularly valuable in scenarios where real data are scarce, expensive to collect, sensitive in nature, or inaccessible. Moreover, synthetic data generation contributes to mitigate the class imbalance issue in real-world data and facilitates the simulation of rare or extreme events, thereby enhancing the robustness and generalizability of predictive models. In aeronautics, (Stanton, Munir, Ikram, & El-Bakry, 2024) addresses the challenge of limited publicly available data, primarily due to the proprietary nature of aircraft system information, by generating synthetic datasets. Specifically, six time-series datasets were synthesized using the DoppelGANger model, trained on real Airbus landing gear system data. While the resulting datasets contain no proprietary content, they preserve the statistical properties and temporal patterns of the original data, making them suitable for evaluating novel PdM models and enabling broader exploration across various aircraft systems. The proposed approach offers a replicable framework for industry practitioners to generate and share realistic, privacy-preserving datasets. To address the lack of comprehensive life-cycle degradation data for bearings, (Wen, Su, Li, Ding, & Feng, 2024) proposes a hybrid approach that first uses a Loop GAN (Loop-GAN) to generate high-quality pseudo data. Subsequently, a hybrid GRU-AE-Wiener model is introduced, combining a Gated Recurrent Unit (GRU) as an autoencoder with a Wiener process as its hidden layer. Both components are jointly optimized with Loop-GAN, emphasizing collaborative information fusion. Validated on the PHM 2012 and XJTU-SY datasets, the proposed model outperforms existing deep learning methods in RUL prediction, demonstrating the effectiveness of this fusion-based approach.

3.2.3. Missing data imputation

Missing data imputation is a critical preprocessing step in PdM, particularly in scenarios where sensor failures, transmission errors, or incomplete logging result in partially observed datasets. Data imputation techniques aim to "fill in the blanks" in the raw data (Khan et al., 2025), ensuring continuity and consistency in temporal patterns essential for reliable prognostic modeling. Traditional methods, such as linear interpolation or statistical estimation, often fail to capture the complex dependencies inherent in multivariate industrial time-series data. GAI offers a more robust solution by learning the underlying data distribution and synthesizing plausible values for missing entries. VAEs and GANs have demonstrated effectiveness in reconstructing incomplete sequences, preserving both local dynamics and global structure. However, in the context of imbalanced datasets, these models tend to favor the over-represented normal class, which can introduce bias and distort performance metrics such as accuracy. This challenge underscores the importance of carefully designing imputation strategies that not only restore missing values but also account for class distribution and data diversity to ensure diagnostic validity (Khan et al., 2025). Finally, (Baptista & Henriques, 2022) proposes a 1-D Denoising GAN for Prognostics and Health Management (1D-DGAN-PHM) approach to enhance the quality of condition monitoring signals in NASA's C-MAPSS Dataset.

3.2.4. Rare events generation

One of the major challenges in RUL prediction is the scarcity of data related to failures, anomalies, and rare events. This lack of representative samples makes it difficult to train robust and generalizable models. GAI techniques such as GANs and VAEs enable the synthesis of synthetic failure data, improving fault detection accuracy and addressing the limitations posed by limited historical failure records. These methods further enhance anomaly detection by generating realistic failure scenarios, allowing predictive models to better anticipate potential breakdowns (Prabha, Nataraj, Sathish, Sujith, & Suthakarr, 2025). (M. Huang, Sheng, & Rao, 2025) tackles the problem of motor fault diagnosis under long-tailed data distributions, where healthy conditions dominate and fault samples are underrepresented. To address this, the authors propose integrating a Conditional VA (CVAE) with a Conditional GAN (CGAN) to improve the generation of rare fault instances. The resulting LT-CVAE-GAN model combines the conditional generation and uncertainty modeling capabilities of CVAEs with the high-fidelity sample generation of CGANs, thereby enhancing diagnostic performance in imbalanced scenarios.

3.2.5. LLM integration for decision-making support

Recent studies have demonstrated the growing relevance of LLMs in PdM. These models offer advanced capabilities for interpreting heterogeneous data sources, detecting anomalies, and supporting maintenance decision-making processes. (Klekowicki et al., 2024) explores the capabilities of LLMs, particularly GPT-4, in interpreting sensor data, detecting anomalies, and generating maintenance guidelines for aircraft engines in the aviation sector. LLMs can also be used to provide maintenance personnel with insights and recommendations based on comprehensive data analysis, supporting better decision-making and reducing reliance solely on human judgment (Klekowicki et al., 2024). (Prabha et al., 2025) presents multimodal LLMs as a promising approach for problem detection in IoT-based real-time monitoring. These models can leverage various data sources, such as sensor readings, historical logs, and images, to enhance decision-making. The integration of LLMs with smaller models has proven effective for real-time anomaly detection, improving both interpretability and accuracy. By aligning with structured knowledge bases, LLMs can deliver maintenance insights tailored to specific industrial contexts, thereby reducing downtime and optimizing maintenance schedules. Furthermore, the use of user-defined prompts has been explored to create flexible and reliable data processing systems. Allowing domain experts to define custom prompts enables LLMs to adapt to specific PdM scenarios, supporting more dynamic and context-aware fault detection. (Prabha et al., 2025) also compared several LLMs (OpenAI GPT-4, Mistral, DeepSeek, Gemma-3, and LLaMA-3) in terms of response time and accuracy. Their evaluation showed that LLaMA-3 achieved the highest accuracy, while Mistral offered the fastest response time. In the leather tanning industry, (Palma, Cecchi, & Rizzo, 2025) explores the use of LLMs to integrate multimodal data for PdM of compressors. The objective is to uncover failure patterns often missed by traditional ML methods and to identify degradation mechanisms not captured by numerical models alone. The study presents a comprehensive performance evaluation of several LLMs, including Qwen-2.5-32B, DeepSeek-R1-Distill-Qwen-32B, Qwen-QWQ-32B, LLaMA 3.3-70B-Versatile, LLaMA 3.2-11B-Vision, and LLaMA 3.2-90B-Vision. Among these, Qwen-2.5-32B demonstrated strong performance in anomaly detection tasks. However, the study also highlights key limitations, particularly concerning the explainability and interpretability of LLM-based predictions. Despite encouraging results, the inherent complexity of LLMs poses challenges in terms of transparency and their practical deployment in industrial environments.

GAI models have shown strong potential in addressing several key challenges in PdM, fault diagnostic, and prognostics, offering promising solutions. The main issues tackled by GAI in this context are summarized in Table 2.

Problem	**Propositions**
Lack of historical data	• Generate synthetic data using GAN variants: BigGAN, DCGAN, cGAN (Campbell et al., 2022) • Data augmentation by generating synthetic data (Stanton et al., 2024) (Khan et al., 2025)
Limited run-to-failure data	• Generate synthetic failure data (Khan et al., 2025; Prabha et al., 2025) • Equipment behavior simulation with GANs and operating scenarios generation using VAEs for subtle anomaly detection (Mohapatra, 2024)
Imbalanced data	• Extraction of meaningful patterns and insights from large volumes of data using LLMs (Klekowicki et al., 2024) • Data augmentation to overcome both insufficient data per condition class and an unbalanced amount of data (Konig et al., 2025)
Noisy or poor-quality data	• Signal denoising using a GAN model (Baptista & Henriques, 2022) • Missing data imputation (Khan et al., 2025) • Analysis and processing of large volumes of text data by LLMs (Klekowicki et al., 2024)
Integration of heterogeneous data sources	• Using Multimodal LLMs to detect problems and improve decision-making (Prabha et al., 2025).
Maintenance schedules optimization	• Simulation of different maintenance scenarios to identify the most efficient schedules (Mohapatra, 2024)

Table 2. Key issues in PdM addressed by GAI.

3.3. Current Challenges and Limitations

Despite their increasing versatility, GAI models exhibit several critical limitations that constrain their deployment in sensitive and critical domains. One of the most prominent challenges is the phenomenon of hallucination, particularly observed in LLMs, where the system generates outputs that are syntactically plausible yet factually incorrect or fabricated (Khan et al., 2025). This issue undermines the reliability of generative models, especially in contexts where factual accuracy is essential. Another persistent concern involves data bias, as these models tend to replicate or amplify, the biases present in their training data, leading to fairness and ethical concerns across various application domains. Furthermore, GAI systems often lack transparency and interpretability. Their black-box nature makes it difficult for users and practitioners to trace, justify, or understand the decision-making process, thereby limiting trust in high-stakes environments (Mohapatra, 2024). The explainability gap is particularly pronounced in architectures such as GANs and transformer-based models, where the complexity of internal representations obscures causal reasoning. In addition to interpretability, generative models are resource-intensive, requiring significant computational power for both training and inference. This imposes constraints on scalability and limits integration into real-time or embedded systems (Mohapatra, 2024). Finally, effective deployment often depends on preprocessing tasks such as data cleaning, quality assurance, and format conversion. For instance, in aviation diagnostics, raw heterogeneous data must be transformed into vector embeddings compatible with LLMs, a process that introduces additional complexity and potential sources of error (Klekowicki et al., 2024). Finally, security considerations including data protection and access control, represent a key challenge for implementing the proposed approach and must be carefully addressed alongside other infrastructure requirements.

Despite recent advances, GAI in its current form remains insufficient for enabling fully reliable PdM. A primary limitations lies in the difficulty of root cause analysis: while black-box models may predict failure events, they rarely provide insight into the underlying causative factors, which is essential for informed diagnostic and maintenance decisions (Mohapatra, 2024). This lack of interpretability not only limits operational utility but also presents significant challenges in highly regulated sectors such as aviation, where transparency is crucial for compliance and accountability. Regulatory audits often require clear justification of decisions, which opaque AI models struggle to provide. Moreover, issues related to integration, domain knowledge transfer, and contextual understanding further hinder the effective deployment of generative models in real-world maintenance systems (Khan et al., 2025). These constraints underscore the need for approaches that combine the generative capacity of AI with structured domain knowledge, enabling more interpretable, traceable, and actionable outputs. In this context, we propose a knowledge-based integration of GAI to enhance PdM frameworks, aiming to bridge the gap between data-driven predictions and expert-driven reasoning.

4. PROPOSED APPROACH: TOWARD HYBRID MODELS

To address the aforementioned limitations, particularly in high-stakes PdM applications like in aviation or for aircraft engine (Hafsi, 2024), there is an increasing need for hybrid frameworks that combine the generative capabilities of modern AI with structured domain expertise. Data-driven approaches, while powerful, suffer from several well-documented constraints, including black-box behavior, sensitivity to non-stationary data requiring frequent retraining (Benatia et al., 2025), difficulty in integrating heterogeneous data sources, and a lack of interpretability and transparency. Conversely, knowledge-based approaches offer the potential to address these limitations by enabling explainable and structured reasoning. However, they introduce their own challenges, such as the dependency on domain experts for knowledge extraction, the difficulty of reaching consensus on rule formalization, the need for interoperability standards, and the time-consuming nature of rule definition and evaluation (Barry & Hafsi, 2023).

To overcome the respective drawbacks of both paradigms, we propose a hybrid framework that integrates GAI within a knowledge-based reasoning architecture. In this approach, GAI is not used as a mere black-box predictor, but as an active component in the construction, enrichment, and evolution of the knowledge base. Specifically, generative models are employed to automatically extract and generate domain concepts, relationships, and ontology elements from heterogeneous datasets and technical documentation. Furthermore, we leverage GAI to propose reasoning rules, whose relevance is then evaluated using data-driven metrics. This enables a weighted, context-aware activation of rules, allowing the system to dynamically adapt to changing operational conditions, thus addressing the issue of data non-stationarity. The hybridization of generative and knowledge-based AI enhances system transparency, supports root cause analysis, and facilitates compliance with regulatory requirements by enabling traceable and interpretable decision-making processes. Additionally, it provides a scalable way to integrate expert input and heterogeneous data formats into a unified reasoning framework, making it well-suited for complex and evolving maintenance environments.

4.1. Motivation

The integration of GAI into ontology engineering offers significant opportunities across the ontology lifecycle, including creation, enrichment, and reasoning. LLMs, in particular, have demonstrated strong capabilities in generating initial concept hierarchies, extracting knowledge from unstructured text, aligning heterogeneous ontologies, and even generating SPARQL queries or logical rules from natural language (Westerinen & Bennett, 2023). Techniques such as DRAGON-AI leverage Retrieval-Augmented Generation (RAG) to dynamically synthesize both textual and logical components of ontologies, drawing from existing ontologies and large knowledge corpora (Toro et al., 2024). These models greatly enhance scalability and interactivity, enabling ontology engineers to build and refine knowledge structures more efficiently. Beyond LLMs, other generative models like GGANs and Autoencoders (AEs) also offer potential, albeit in more specialized roles. GANs can be used to generate synthetic data or embeddings that simulate plausible instances of knowledge graphs, which may assist in ontology population or feature augmentation. Autoencoders, especially VAEs, can learn compact latent representations of concepts and relations, supporting tasks like ontology alignment, concept similarity detection, or reconstruction of incomplete knowledge structures. While these models are less suited for symbolic reasoning, their ability to capture semantic regularities and generate new candidates for concepts or relations makes them valuable complements in data-driven ontology development.

4.2. Proposed framework

We propose a methodological framework based on a modular architecture that combines the strengths of GAI and knowledge-based systems in an explainable hybrid approach. The framework is designed to be robust during development and adaptable after deployment through continuous learning from real-world data. It processes heterogeneous raw inputs (e.g., measurements, sensor data, technical reports, textual data) and is organized into four main modules, as illustrated in Figure 2.

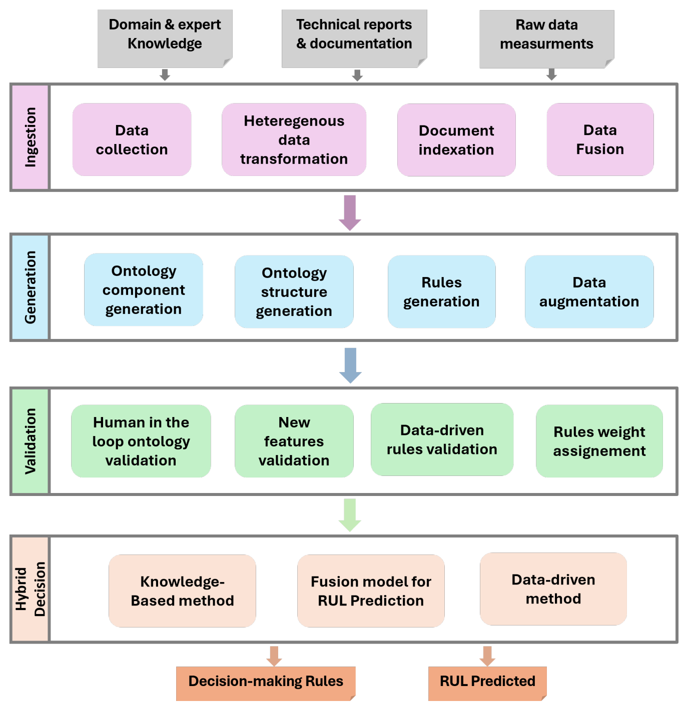

Figure 2. The proposed framework for interpretable RUL prediction integrating GAI to address multiple challenges in a hybrid approach.

Data ingestion module: This module handles the acquisition and integration of heterogeneous data sources, including sensor readings, maintenance records, and unstructured textual documents. It performs data cleaning, transformation, and feature engineering to generate enriched representations. In addition, it indexes textual documents to support semantic search and knowledge extraction in later stages.

Data and knowledge generation module: This module plays a dual role in enhancing both the dataset and the knowledge base. First, it improves the quality and robustness of the dataset through techniques such as data augmentation, synthetic data generation, data imputation, and run-to-failure data generation, particularly valuable in domains with limited labeled data. Second, it leverages GAI to construct the core components of the domain ontology. This includes the automatic extraction, generation and formalization of concepts, their properties, hierarchical and semantic relationships, and symbolic reasoning rules. These elements form the backbone of the knowledge-based system, enabling structured representation and reasoning over heterogeneous information sources.

Validation module: This module ensures the reliability and contextual relevance of the generated knowledge components. It begins with a human-in-the-loop validation phase, where domain experts review and confirm the coherence of the generated ontology, including the correctness of its hierarchical and semantic structure. Next, each generated reasoning rule is evaluated using data-driven techniques, assessing its performance and consistency against historical datasets. Validated rules are then assigned contextual weights, enabling selective activation based on operational conditions, an essential feature for adapting to non-stationary environments. Additionally, this module assesses the relevance and impact of newly engineered features, ensuring that only informative and robust variables are retained for downstream predictive tasks.

Hybrid prediction module: This module delivers interpretable RUL estimations by combining data-driven inference with symbolic reasoning. It implements two complementary prediction pathways: (1) A ML or DL model trained on the enriched and structured dataset produced by the upstream modules. and (2) a knowledge-based reasoning approach, which uses validated rules and ontological relationships to derive RUL estimates through symbolic inference. A configurable fusion mechanism integrates the outputs of both pathways, allowing for adaptive weighting or selection depending on the use case, system criticality, or data availability. Crucially, the module outputs not only the estimated RUL value but also a traceable explanation detailing which rules and knowledge elements contributed to the final prediction, enhancing interpretability and trust in the decision-making process.

5. DISCUSSION

To validate the proposed methodological framework, we suggest its application to a well-established industrial use case: aircraft engine PdM using the C-MAPSS dataset (Saxena, Goebel, Simon, & Eklund, 2008). Developed by NASA, this dataset provides run-to-failure simulations, including multivariate time series sensor data that closely reflect real-world degradation patterns. To make the validation more constructive, we propose leveraging the accompanying technical documentation and incorporating the physical rules that describe engine behavior and its interactions with the environment. This approach enables the framework to integrate domain knowledge and symbolic reasoning, providing a more robust and interpretable assessment of RUL predictions.

Aircraft engines are particularly suitable for evaluating PdM frameworks due to their high reliability requirements and abundant sensor data. A major challenge in this domain is the imbalance between normal and faulty operating data, which limits the generalization of purely data-driven models and the representation of the full range of degradation behaviors. By combining GAI techniques with knowledge-based reasoning, the framework can enrich training data, generate meaningful features, and formulate interpretable decision rules—capabilities that are especially valuable in imbalanced scenarios. The symbolic layer also supports extrapolation to previously unseen cases, which data-driven models alone may fail to handle.

Future work will involve applying this framework to C-MAPSS and evaluating its performance in terms of prediction accuracy, interpretability, and robustness under low-fault-sample conditions, as illustrated in Figure 3.

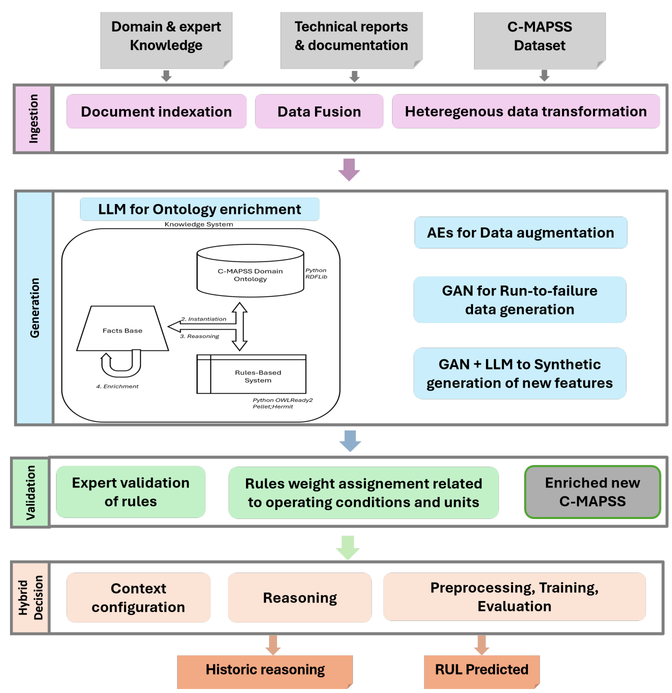

Figure 3. Proposed framework applied to the estimation of RUL for aircraft engines using the C-MAPSS dataset.

6. CONCLUSION

In PdM, RUL estimation of components in critical systems remains a highly complex task. While recent models particularly those based on data-driven have achieved impressive accuracy, several persistent challenges continue to hinder their practical deployment. These include the scarcity of high-quality failure data, the lack of interpretability, the difficulty of generalization across different operational conditions, and the complexity of integrating domain knowledge into learning systems. In this evolving landscape, the emergence of GAI opens new avenues. By enabling synthetic data generation, knowledge extraction, and advanced reasoning capabilities, generative AI brings promising contributions to overcome data limitations and enhance model explainability. However, its integration also raises new challenges, including ensuring the reliability of generated outputs, maintaining trust and transparency, and aligning generated knowledge with domain-specific constraints. Among the various approaches reviewed, the knowledge-based paradigm stands out as a promising and complementary direction. Its emphasis on human expertise, system understanding, and interpretability makes it a valuable asset especially when integrated with data-driven methods. This convergence paves the way toward neuro-symbolic approaches, which aim to combine the learning capacity of neural networks with the reasoning and transparency of symbolic systems, thereby moving toward more robust and explainable artificial intelligence.

In this work, we have provided a review of the most recent contributions in PdM that leverage GAI, identifying both their strengths and current limitations. Building on these insights, we proposed a novel hybrid framework that integrates GAI into a knowledge and data-based PdM approach. This framework is designed to enhance the reliability and interpretability of RUL estimation models. Future work will focus on implementing and evaluating this framework in a real-world use case involving aircraft engines, a critical domain where accurate and trustworthy prognostics are essential.

REFERENCES

Baptista, M., & Henriques, E. (2022). 1d-dgan-phm: A 1-d denoising gan for prognostics and health management with an application to turbofan. *Applied Soft Computing*, *131*. doi: 10.1016/j.asoc.2022.109785

Barry, I., & Hafsi, M. (2023). Towards hybrid predictive maintenance for aircraft engine: Embracing an ontological-data approach. In *2023 20th acs/ieee international conference on computer systems and applications (aiccsa)* (p. 1-6). doi: 10.1109/AICCSA59173.2023.10479328

Benatia, M., Hafsi, M., & Ayed, S. B. (2025). A continual learning approach for failure prediction under non-stationary conditions: Application to condition monitoring data streams. *Computers & Industrial Engineering*, *204*, 111049. doi: https://doi.org/10.1016/j.cie.2025.111049

Campbell, N., Ilangovan, H., Gregory, I., & Mikaelian, S. (2022, 01). Data augmentation for intelligent contingency management using generative adversarial networks. In *Aiaa scitech 2022 forum.* doi: 10.2514/6.2022-0622

Cao, Q., Zanni-Merk, C., Samet, A., De Beuvron, F. D. B., & Reich, C. (2020). Using rule quality measures for rule base refinement in knowledge-based predictive maintenance systems. *Cybernetics and Systems*, *51*(2), 161–176.

Cao, Q., Zanni-Merk, C., Samet, A., Reich, C., De Beuvron, F. D. B., Beckmann, A., & Giannetti, C. (2022). Kspmi: a knowledge-based system for predictive maintenance in industry 4.0. *Robotics and Computer-Integrated Manufacturing*, *74*, 102281.

Davari, N., Veloso, B., Costa, G. d. A., Pereira, P. M., Ribeiro, R. P., & Gama, J. (2021). A survey on data-driven predictive maintenance for the railway industry. *Sensors*, *21*(17), 5739.

de Pater, I., Reijns, A., & Mitici, M. (2022). Alarm-based predictive maintenance scheduling for aircraft engines with imperfect remaining useful life prognostics. *Reliability Engineering & System Safety*, *221*, 108341.

Eker, O. F., Camci, F., & Jennions, I. K. (2019). A new hybrid prognostic methodology. *International Journal of Prognostics and Health Management*, *10*(2).

Hafsi, M. (2024, July). Exploring a Knowledge-Based Approach for Predictive Maintenance of Aircraft Engines: Studying Fault Propagation through Spatial and Topological Component Relationships. In *8th european conference of the prognostics and health management society 2024* (p. 9). Czech Republic. doi: 10.36001/phme.2024.v8i1.4100

Hagmeyer, S., Zeiler, P., & Huber, M. F. (2022). On the integration of fundamental knowledge about degradation processes into data-driven diagnostics and prognostics using theory-guided data science. In *Phm society european conference* (Vol. 7, pp. 156–165).

Huang, M., Sheng, C., & Rao, X. (2025). Conditional variational autoencoder and generative adversarial network-based approach for long-tailed fault diagnosis for the motor system. *Measurement*, *242*, 116116.

Huang, W., Zhang, X., Jiang, H., Shao, Z., & Bai, Y. (2025, 01). Mcba-mvacgan: A novel fault diagnosis method for rotating machinery under small sample conditions. *Machines*, *13*, 71. doi: 10.3390/machines13010071

Işık, B. (2023). Knowledge-based maintenance management system of compressed air system. In *International symposium on industrial engineering and automation* (pp. 197–208).

Khan, Z. U., Nasim, B., & Rasheed, Z. (2025, 10). Generative ai based predictive maintenance in aviation: A systematic literature review. *CEAS Aeronautical Journal*. doi: 10.1007/s13272-025-00818-1

Klekowicki, M., Szymański, G., Waligórski, M., & Misztal, W. (2024, 11). Application of large language models in diagnostics and maintenance of aircraft propulsion systems. *Advances in Science and Technology Research Journal*, *19*, 304-320. doi: 10.12913/22998624/196264

Konig, T., Wagner, F., Liebschner, M., & Kley, M. (2025). Generative adversarial networks in fault diagnosis and an implementation approach: A review. *IEEE Sensors Journal*. doi: 10.1109/JSEN.2025.3575284

Lang, P., Peng, K., Cui, J., Yang, J., & Guo, Y. (2021). Data augmentation for fault prediction of aircraft engine with generative adversarial networks. In *2021 caa symposium on fault detection, supervision, and safety for technical processes (safeprocess)* (p. 1-5). doi: 10.1109/SAFEPROCESS52771.2021.9693711

Li, X. Y., Cheng, D. J., Fang, X. F., Zhang, C. Y., & Wang, Y. F. (2024). A novel data augmentation strategy for aeroengine multitask prognosis based on degradation behavior extrapolation and diversity-usability trade-off. *Reliability Engineering & System Safety*, *249*. doi: https://doi.org/10.1016/j.ress.2024.110238

Li, Z., Jiang, H., & Wang, X. (2025). A novel reinforcement learning agent for rotating machinery fault diagnosis with data augmentation. *Reliability Engineering & System Safety*, *253*, 110570.

Mohapatra, A. (2024, 11). Generative ai for predictive maintenance: Predicting equipment failures and optimizing maintenance schedules using ai. *International Journal of Scientific Research and Management (IJSRM)*, *12*, 1648-1672.

Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance–a review. *CIRP Journal of Manufacturing Science and Technology*, *40*, 53–67.

Palma, G., Cecchi, G., & Rizzo, A. (2025). Large language models for predictive maintenance in the leather tanning industry: Multimodal anomaly detection in compressors. *Electronics*, *14*(10).

Prabha, K., Nataraj, B., Sathish, S., Sujith, C., & Suthakarr, G. (2025, 05). Predictive maintenance for industrial machinery using llm. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, *3*, 2063-2069. doi: 10.47392/IRJAEH.2025.0301

Rosati, R., Romeo, L., Cecchini, G., Tonetto, F., Viti, P., Mancini, A., & Frontoni, E. (2023). From knowledge-based to big data analytic model: a novel iot and machine learning based decision support system for predictive maintenance in industry 4.0. *Journal of Intelligent Manufacturing*, *34*(1), 107–121.

Santiago, R. A., Barbosa, N. B., Mergulhão, H. G., Carvalho, T., Santos, A. B., Medrado, R. C., ... Nascimento, E. G. S. (2024). Data-driven models applied to predictive and prescriptive maintenance of wind turbine: A systematic review of approaches based on failure detection, diagnosis, and prognosis. *Energies*, *17*(5).

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (p. 1-9). doi: 10.1109/PHM.2008.4711414

Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical systems and signal processing*, *25*(5), 1803–1836.

Stanton, I., Munir, K., Ikram, A., & El-Bakry, M. (2024). Data augmentation for predictive maintenance: Synthesising aircraft landing gear datasets. *Engineering Reports*, *6*(12), e12946. doi: https://doi.org/10.1002/eng2.12946

Toro, S., Anagnostopoulos, A., Bello, S., Blumberg, K., Cameron, R., Carmody, L., ... Mungall, C. (2024). Dynamic retrieval augmented generation of ontologies using artificial intelligence (dragon-ai). *Journal of Biomedical Semantics*, *15*.

Traini, E., Bruno, G., & Lombardi, F. (2021). Design of a physics-based and data-driven hybrid model for predictive maintenance. In *Advances in production management systems. artificial intelligence for sustainable and resilient production systems: Ifip wg 5.7 international conference, apms 2021, nantes, france* (pp. 536–543).

Wen, L., Su, S., Li, X., Ding, W., & Feng, K. (2024). Gru-ae-wiener: A generative adversarial network assisted hybrid gated recurrent unit with wiener model for bearing remaining useful life estimation. *Mechanical Systems and Signal Processing*, *220*, 111663. doi: https://doi.org/10.1016/j.ymssp.2024.111663

Westerinen, A., & Bennett, M. (2023, 10). Ontologies and large language models related but different..

Zio, E. (2022). Prognostics and health management (phm): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety*, *218*, 108119. doi: https://doi.org/10.1016/j.ress.2021.108119

BIOGRAPHIES

Meriem Hafsi was born in 1990 in Tizi-Ouzou, Algeria. She holds a Master's degree in Web Intelligence from Jean Monnet University and earned her PhD in Computer Science from Grenoble-Alpes University. She is currently a researcher and lecturer at CESI and the LINEACT Laboratory in Lyon, where she works on predictive maintenance. Her research focuses on data-related challenges and on the integration of symbolic reasoning into hybrid artificial intelligence approaches to improve explainability and reasoning capabilities in intelligent maintenance systems.