

## PERSPECTIVE ARTICLE

# Rethinking industrial artificial intelligence: A unified foundation framework

 Jay Lee  and Hanqi Su\* 

Center for Industrial Artificial Intelligence, Department of Mechanical Engineering, A. James Clark School of Engineering, University of Maryland, College Park, Maryland, United States of America

## Abstract

Recent advancements in industrial artificial intelligence (AI) are reshaping the industry by driving smarter manufacturing, predictive maintenance, and intelligent decision-making. However, existing approaches often focus primarily on algorithms and models while overlooking the importance of systematically integrating domain knowledge, data, and models to develop more comprehensive and effective AI solutions. Therefore, the effective development and deployment of industrial AI require a more comprehensive and systematic approach. To address this gap, this paper reviews previous research, rethinks the role of industrial AI, and proposes a unified industrial AI foundation framework comprising three core modules: the knowledge module, data module, and model module. These modules help to extend and enhance the industrial AI methodology platform, supporting various industrial applications. In addition, a case study on rotating machinery diagnosis is presented to demonstrate the effectiveness of the proposed framework, and several future directions are highlighted for the development of the industrial AI foundation framework.

---

**\*Corresponding author:**

 Hanqi Su  
 (hanqisu@umd.edu)

**Citation:** Lee J, Su H. Rethinking industrial artificial intelligence: A unified foundation framework. *Int J AI Mater Design*. 2025;2(2):56-68. doi: 10.36922/IJAMD025080006

**Received:** February 21, 2025

**1st revised:** March 23, 2025

**2nd revised:** March 28, 2025

**Accepted:** April 2, 2025

**Published Online:** April 15, 2025

**Copyright:** © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

**Publisher's Note:** AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Keywords:** Industrial artificial intelligence; Industry 4.0; Machine learning; Deep learning; Large language model; Domain knowledge

---

## 1. Introduction

The rapid advancement of industrial artificial intelligence (AI) is reshaping industries worldwide.<sup>1-4</sup> Recent breakthroughs in technologies such as deep learning,<sup>1,2</sup> industrial internet of things (IIoT),<sup>5,6</sup> large language models (LLMs),<sup>7,8</sup> prognostics and health management,<sup>9,10</sup> big data analytics,<sup>11,12</sup> and cyber-physical systems (CPS)<sup>13,14</sup> have accelerated the adoption of industrial AI, enabling industrial systems to extract actionable insights from vast amounts of industrial data and support intelligent decision-making. However, current approaches often overemphasize algorithms and models while lacking a unified framework that systematically integrates domain knowledge, data, and models. To unlock the full potential of industrial AI, a structured framework is needed – one that integrates domain knowledge, high-quality data, and intelligent AI models to address complex challenges in real-world industrial settings.

Recognizing this gap, this paper proposes a unified industrial AI foundation framework composed of knowledge, data, and model modules to enhance the industrial AI methodology platform. The remainder of the paper is structured as follows:

Section 2 reviews related work; Section 3 redefines the role of industrial AI; Section 4 introduces the proposed foundation framework; Section 5 provides a case study on rotating machinery diagnosis; Section 6 discusses future directions; and Section 7 concludes the paper.

## 2. Related work

This section reviews existing studies and frameworks focusing on the integration of AI within various industrial applications. Several research works have proposed frameworks aimed at enhancing industrial productivity, decision-making, and operations by leveraging AI capabilities. Table 1 provides a clear summary of these previous industrial AI frameworks, highlighting their primary contributions. For instance, Lee *et al.*<sup>4</sup> introduced an industrial AI ecosystem framework that integrates enabling technologies, such as data technology, analytic technology, platform technology, and operation technology, within the established CPS (5C) architecture, guided by key ABCDE elements. Building on this foundation, Peres *et al.*<sup>2</sup> expanded the framework into a more comprehensive conceptual framework, emphasizing challenges, design principles, essential technologies, capabilities, attributes,

and industrial application areas. On the other hand, Zhang *et al.*<sup>15</sup> developed a comprehensive reference framework structured around seven critical dimensions (object, domain, stage, requirement, technology, function, and solutions). In addition, Yang *et al.*<sup>16</sup> proposed an intelligent manufacturing framework combining human-machine cooperation with autonomous intelligent control systems specifically tailored to the process industry. Ahmed *et al.*<sup>17</sup> highlighted AI and explainable AI (XAI) methodologies, categorizing various XAI methods and their applicability in Industry 4.0 contexts, while Jan *et al.*<sup>1</sup> structured a four-stage AI data pipeline, outlining data acquisition/validation, data processing/fusion, model training/testing, and model interpretation. Furthermore, Leng *et al.*<sup>3</sup> introduced a four-layer technical reference framework encompassing layers from hardware to industrial applications. Most recently, Lee and Su<sup>7</sup> proposed a unified industrial large knowledge model (ILKM) framework emphasizing domain-specific knowledge integration through LLMs and machine learning (ML) approaches.

Although these frameworks have successfully identified important aspects of AI integration across different industrial applications, several limitations remain.

**Table 1. Summary of existing frameworks for industrial artificial intelligence applications**

Authors	Year	Proposed industrial artificial intelligence (AI) frameworks
Lee <i>et al.</i> <sup>4</sup>	2018	Proposed an industrial AI ecosystem framework to integrate enabling technologies – data technology, analytic technology, platform technology, and operation technology – within the Cyber-Physical System (5C) architecture. The framework systematically guides AI implementation in smart manufacturing under the guidance of five key ABCDE elements: analytics technology, big data technology, cloud or cyber technology, domain know-how, and evidence.
Zhang <i>et al.</i> <sup>15</sup>	2019	Presented a comprehensive industrial AI reference framework consisting of seven key dimensions: Object (who), domain (where), application stage (when), application requirement (why), intelligent technology (which), intelligent function (what), and solutions (how); and offers a detailed overall planning for systematically integrating AI across diverse industrial scenarios.
Peres <i>et al.</i> <sup>2</sup>	2020	Proposed a conceptual industrial AI framework highlighting essential enabling technologies (data, analytics, platforms, operations, and human-machine interaction) while systematically identifying critical challenges, attributes, capabilities, design principles, and common application domains in Industry 4.0.
Yang <i>et al.</i> <sup>16</sup>	2021	Presented a two-tier intelligent manufacturing framework designed for the process industry, integrating human-machine cooperation, and intelligent autonomous control systems to achieve intelligent optimal decision-making.
Ahmed <i>et al.</i> <sup>17</sup>	2022	Conducted a comprehensive survey on AI and explainable AI (XAI) methodologies within Industry 4.0, categorizing various XAI approaches (model-specific, model-agnostic, local/global, visualization-based methods, and surrogate models), highlighting their applicability, benefits, and challenges in industrial contexts.
Jan <i>et al.</i> <sup>1</sup>	2023	Proposed a structured, four-stage industrial AI data pipeline (data acquisition/validation, data processing/fusion, model training/testing, and model interpretation). They identified common themes, issues, and industry-specific solutions related to AI integration in various sectors, providing insights into practical challenges and opportunities within Industry 4.0.
Leng <i>et al.</i> <sup>3</sup>	2024	Presented a technical reference framework structured in four layers (hardware infrastructure, computing engine, AI algorithms, and industrial application layer together with related empowering technologies across layers), identifying three core opportunities (collaborative intelligence, self-learning intelligence, and crowd intelligence) crucial for realizing Industry 5.0's vision of human-centric, resilient, and sustainable manufacturing.
Lee and Su <sup>7</sup>	2024	Proposed a unified industrial large knowledge model framework consisting of four systematic steps: (i) construction of a large knowledge library; (ii) preparation of domain-specific instruction data; (iii) development of domain-specific knowledge large language models; and (iv) establishment of intelligent domain expert machine learning systems; guided by the “6S Principle” (specialized knowledge, scrutability, safety, scalability, sustainability, and systematization).

These include the lack of systematic domain-specific knowledge integration, a tendency to focus on theoretical conceptualization without clear and objective guidelines, and an overemphasis on data and models without sufficient attention to systematic thinking that connects knowledge, data, and models. Motivated by these gaps, the following two sections revisit the role of industrial AI and present a unified industrial AI foundation framework aimed at addressing these critical shortcomings.

### 3. Rethinking the role of industrial AI

The role of industrial AI can be best understood through a two-layer perspective, as shown in Figure 1, where the physical layer consists of humans, things, and systems, while the digital layer consists of knowledge, data, and models. Based on recent emerging technologies such as IIoT,<sup>5,6</sup> digital twin,<sup>18,19</sup> CPS,<sup>13,14</sup> industrial big data analytics,<sup>11,12</sup> deep learning,<sup>1,2</sup> and LLMs,<sup>7,8</sup> industrial AI acts as a dynamic bridge between the physical and digital layers, continuously refining and applying AI-driven insights to improve and optimize the real-world industrial systems. More specifically, human expertise forms the

basis of domain knowledge, which contains accumulated experience, research insights, and best practices from AI-driven approaches. With the rise of IoT technology, things, including machines, sensors, and other devices, generate vast amounts of data, which is an essential prerequisite for conducting industrial AI analytics, modeling, and decision-making in modern industry. Moreover, systems such as CPS, IIoT systems, and other intelligent industrial systems use AI models to automate operations, enable intelligent decision-making, and improve real-time control.

While the two-layer perspective provides a conceptual understanding of how industrial AI bridges the physical and digital domains, real-world industrial settings present several challenges. In practice, developers and engineers face persistent difficulties in transforming heterogeneous data from IIoT devices into reliable, AI-ready formats due to inconsistent data quality, lack of standard preprocessing pipelines, and fragmented metadata management.<sup>5</sup> CPS, although widely discussed, often remains difficult to operationalize, as maintaining consistency between real-world operations and virtual models requires continuous data synchronization and robust model updating.<sup>13</sup> Despite its promise, digital twin development is challenged by the complexity of integrating multi-source data streams, real-time analytics, and simulation modeling requirements.<sup>18</sup> Furthermore, the increasing availability of LLMs introduces new opportunities but also pose practical challenges in domain knowledge understanding, extraction, and interpretability for industrial use cases.<sup>7</sup> Based on our observations from industry collaborations and project experiences, these complexities frequently result in ad-hoc solutions, isolated development efforts, and inefficiencies in scaling AI-driven applications. Therefore, these observations and challenges motivate the need for a structured foundation framework that systematically connects knowledge, data, and models, as presented in the following section.

### 4. Industrial AI foundation framework

In this section, we propose a unified industrial AI foundation framework designed to systematically guide the development and deployment of industrial AI solutions, as illustrated in Figure 2. The framework consists of three core modules: (i) knowledge module; (ii) data module; and (iii) model module. Within each module, we identify four key components, resulting in 12 important aspects that collectively provide structured direction for industrial AI developers and practitioners. These modules enhance and extend the industrial AI methodology platform by providing a structured, modular foundation for existing and emerging industrial AI methodologies. Figure 2

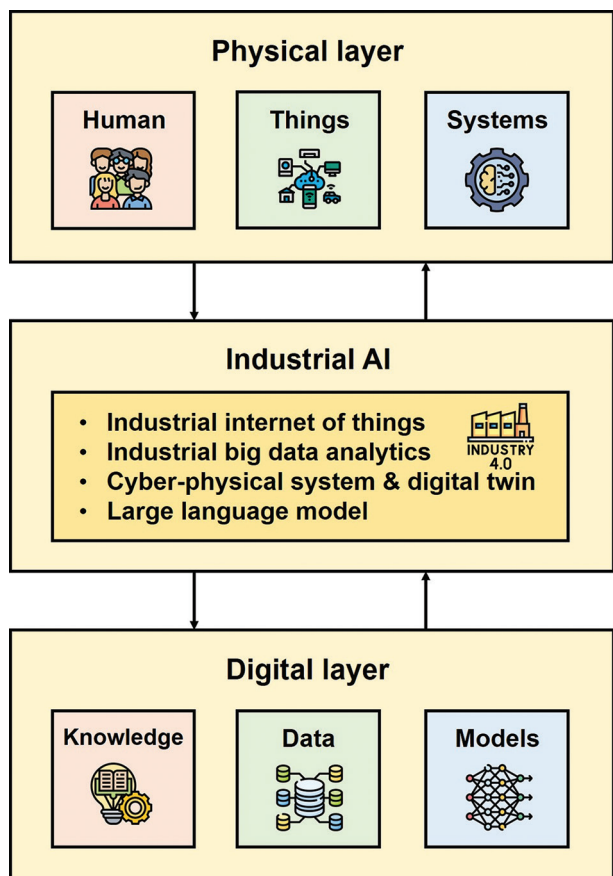
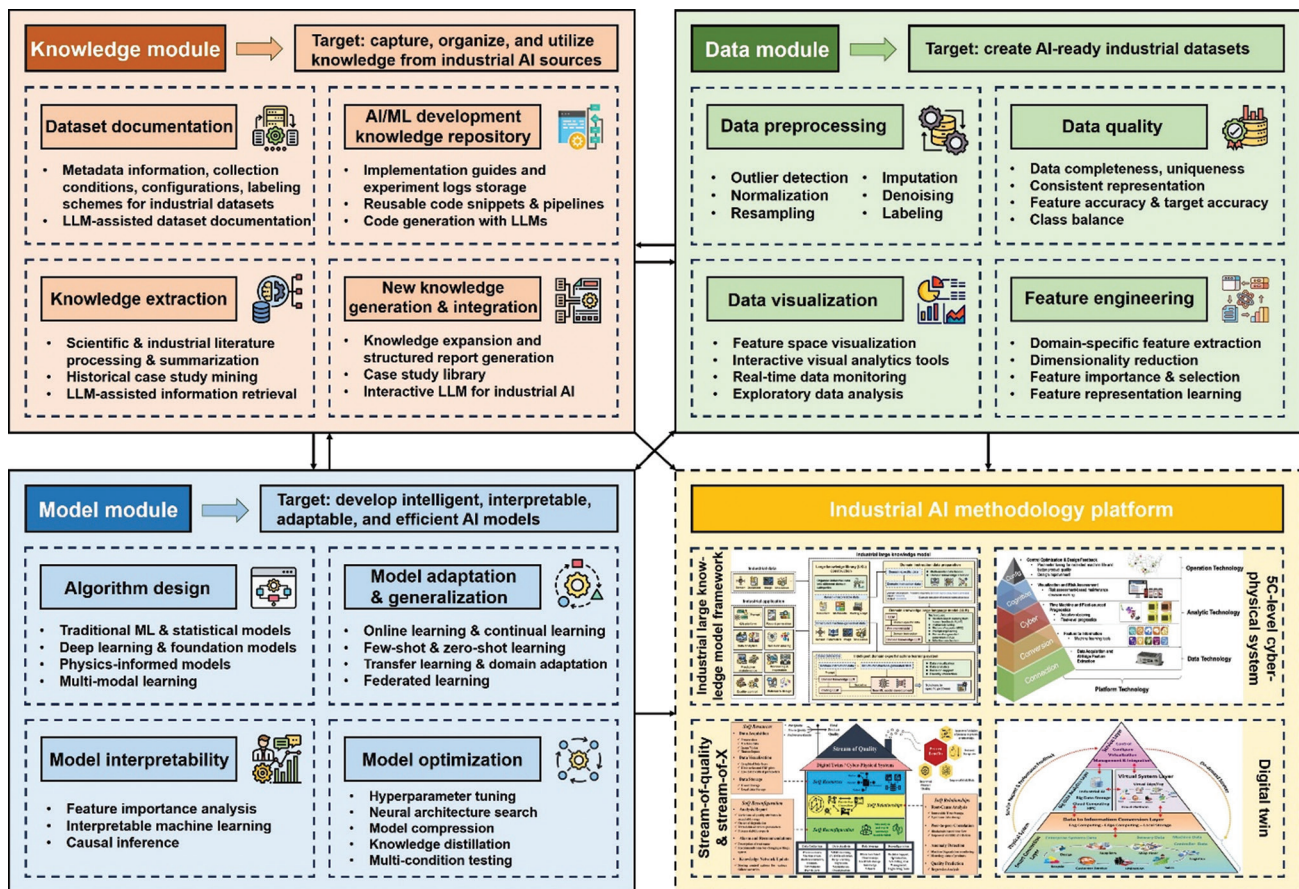


Figure 1. The role of industrial artificial intelligence in bridging the physical and digital layers. Image created by the authors.



**Figure 2.** A unified industrial artificial intelligence (AI) foundation framework. The industrial large knowledge model framework image is reprinted with permission from Lee and Su.<sup>7</sup> The stream of quality image is reprinted with permission from Lee *et al.*<sup>20</sup> Copyright © 2022 Society of Manufacturing Engineers (SME). The 5C-cyber-physical system figure is reprinted with permission from Lee *et al.*<sup>4</sup> Copyright © 2018 SME. The digital twin image is reprinted with permission from Lee *et al.*<sup>19</sup> Copyright © 2020 The Institution of Engineering and Technology. Abbreviations: LLM: Large language model; ML: Machine learning.

illustrates the overall structure of the proposed framework, highlighting the interconnections and dynamic feedback between knowledge, data, and model modules. Each pair of modules is connected by forward and backward arrows, highlighting that these modules continuously inform and refine one another, rather than existing in a linear sequence. In addition, forward arrows from all three modules point toward the industrial AI methodology platform, indicating that the platform is strengthened by these three modules. In Sections 4.1-4.4, the details of the knowledge, data, and model modules are outlined, followed by an explanation of how they enhance and extend the capabilities of the industrial AI methodology platform.

#### 4.1. Knowledge module

The knowledge module focuses on capturing, organizing, and utilizing domain knowledge to support the entire industrial AI development cycle. In complex industrial settings, domain expertise often exists in unstructured

forms, such as reports, manuals, research papers, and maintenance logs. Rather than reflecting an anthropocentric design philosophy in the traditional sense, the knowledge module aims to transform these scattered resources into structured, accessible, and reusable formats that guide data analytics and model development – serving as a foundation that enables scalable, automated, and context-aware AI system development. It focuses on the following four aspects.

##### 4.1.1. Knowledge extraction

Knowledge extraction refers to the systematic process of converting unstructured information into structured, searchable knowledge representations. This is essential for enabling automated reasoning and guiding data preprocessing and model design. Techniques such as knowledge graph construction<sup>21-23</sup> and LLM-assisted information extraction<sup>24</sup> are commonly used. For instance, tools such as OpenKE,<sup>25</sup> TransOMCS,<sup>26</sup> and gBuilder<sup>27</sup> can

help with knowledge graph construction, while LLMs, such as GPT-3/4,<sup>28,29</sup> Llama1/2,<sup>30,31</sup> PaLM,<sup>32</sup> and DeepSeek,<sup>33</sup> can automate the summarization and extraction of key technical information from technical reports and research articles.

#### **4.1.2. Dataset documentation as industrial knowledge**

Dataset documentation is a critical part of building structured industrial knowledge. It involves systematically recording dataset metadata, collection conditions, sensor configurations, labeling schemes, and linking datasets with domain knowledge. This ensures that datasets evolve from isolated assets into long-term, reusable knowledge resources that can be properly understood, reused, and referenced in future work. Platforms such as GitHub and Hugging Face demonstrate best practices in dataset documentation, providing clear descriptions, structured metadata fields, and version histories. In addition, LLMs could enrich dataset documentation.<sup>34</sup>

#### **4.1.3. AI/ML development knowledge repository**

An AI/ML development knowledge repository provides a centralized space to store reusable code templates, implementation guides, hyperparameter tuning records, and experiment logs. This repository accelerates development by allowing engineers to reuse proven techniques and implementation methods. Examples include maintaining shared code repositories on GitHub combined with experiment documentation platforms and using platforms such as MLflow or Weights and Biases for tracking experiments, packaging code, managing models, and sharing results. Moreover, LLMs also enhance multiple aspects, including code generation, information retrieval, and interactive AI-assisted exploration.<sup>35-38</sup>

#### **4.1.4. New knowledge generation and integration**

As models are developed and validated, they produce new insights and interpretations that should be continuously integrated into the existing knowledge base. This process involves capturing lessons learned, model interpretation outputs, model results, and deployment feedback, feeding them back into structured documentation and knowledge graphs. This iterative process ensures that the knowledge module remains dynamic and evolves alongside technological progress and deployment experiences.

### **4.2. Data module**

The data module focuses on transforming raw industrial data into AI-ready datasets to improve data usability and reliability, enhanced by the domain knowledge from the knowledge module. In processing data, this module

focuses on four key aspects: data preprocessing, data quality, feature engineering, and data visualization.

#### **4.2.1. Data preprocessing**

This is a fundamental prerequisite for successful industrial AI applications, as raw data often contains issues such as noise, missing values, anomalies, class imbalances, and labeling inconsistencies. A well-structured data preprocessing pipeline typically involves several key techniques, including outlier detection methods (such as isolation forests, local outlier factor, and statistical thresholding), data imputation methods (including mean, median, multiple imputation, k-nearest neighbors, and regression-based approaches), signal denoising methods (such as wavelet transforms and filtering), and resampling methods (including synthetic minority over-sampling technique, random over-sampling, and under-sampling).<sup>39</sup> In addition, researchers are encouraged to adopt, design, and document reusable preprocessing pipelines using standardized tools such as Pandas, scikit-learn, PyTorch, and TensorFlow to ensure reproducibility and scalability.

#### **4.2.2. Data quality**

Data quality directly impacts the reliability of AI models, as poor-quality data can degrade model performance and lead to poor decision-making. Key dimensions of data quality include consistent representation, completeness, uniqueness, feature accuracy, target accuracy, and target class balance.<sup>40</sup> Therefore, developers should adopt systematic data quality checks as part of their pipelines. For instance, CleanML demonstrates how various data quality issues can significantly affect the performance of common ML models.<sup>41</sup> Furthermore, Foroni *et al.*<sup>42</sup> extended conventional data quality definitions by evaluating not only how data deviate from an ideal clean dataset but also how these deviations influence task outcomes.

#### **4.2.3. Feature engineering**

Feature engineering focuses on extracting meaningful representations from raw data, improving model performance, interpretability, and efficiency.<sup>43</sup> In industrial applications, well-designed features can greatly improve model accuracy. Common techniques include domain-specific feature extraction, such as time-domain statistical measures, frequency-domain features derived from Fourier transforms, and time-frequency domain features. Dimensionality reduction methods, including principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and linear discriminant analysis, help reduce complexity while maintaining important information. Feature importance ranking methods are used to identify key variables that most

influence model predictions, guiding feature selection. In addition, deep learning-based feature representation learning can discover complex patterns in datasets. Tools such as tsfresh, PyCaret, and Shapley Additive Explanations (SHAP) provide automated pipelines for feature extraction, feature importance, and feature selection. Therefore, effective feature engineering ensures that AI models focus on the most informative inputs while minimizing redundancy and noise.

#### **4.2.4. Data visualization**

Data visualization plays an important role in making complex datasets interpretable for both humans and AI systems.<sup>44</sup> Visualization supports tasks such as exploratory data analysis, feature space visualization, anomaly detection, and real-time monitoring. For instance, feature space visualization using t-SNE or PCA projections can help detect outliers or clusters in high-dimensional sensor data. Real-time dashboards that visualize data streams from machines and production lines help operational teams identify issues early and improve system reliability. Moreover, industrial practitioners are encouraged to utilize interactive visualization tools, such as Plotly, Tableau, and D3, to enable dynamic data exploration.

### **4.3. Model module**

With structured knowledge and AI-ready data in place, the model module is responsible for developing intelligent, adaptable, interpretable, and efficient AI models to drive industrial AI applications in Industry 4.0. This module provides four key aspects to help developers systematically design, adapt, and refine models that meet complex industrial requirements.

#### **4.3.1. Algorithm design**

Algorithm design serves as the foundation for developing AI models. Developers should select or design models that align with data characteristics, domain constraints, and application requirements. Typical options range from traditional ML and statistical methods – such as decision tree, random forest, support vector machine, gradient boosting, k-nearest neighbors, Gaussian mixture model, and various clustering algorithms – to deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks, graph neural networks, transformers, and their variants.<sup>45,46</sup> These algorithms are applicable to problems involving both continuous variables (e.g., vibration signals, temperature, and pressure) and discrete variables (e.g., operational states, fault modes, and control events). Beyond traditional ML and deep learning architectures, physics-informed neural networks (PINNs) have emerged as valuable tools in industrial AI.<sup>47,48</sup>

PINNs incorporate physical laws and equations into the learning process, improving model reliability in scenarios where data are sparse but prior knowledge is available. In addition, multi-modal learning approaches, which combine information from multiple modalities – such as images, text, audio, tabular data, and sensor measurements – enable the development of more comprehensive models suitable for complex industrial scenarios.<sup>49,50</sup> Furthermore, foundation models, which are pre-trained on large-scale diverse datasets, offer a promising approach for transferable and adaptable AI across different industrial domains. These models can be fine-tuned for specific tasks, allowing for efficient deployment with reduced training time and improved generalization.<sup>51</sup> In the process of algorithm design, developers may utilize prior knowledge and experience to create new model architectures or iteratively refine existing models through systematic experimentation and testing. When open-source implementations are available, they can be reproduced and further extended based on established methods.

#### **4.3.2. Model interpretability**

Model interpretability is crucial for ensuring that AI models are transparent and explainable. One important direction is interpretable ML, which focuses on developing models that are transparent by design, such as regression models, decision trees, and generalized additive models. These models, along with techniques such as rule-based learning and sparse linear models, can provide clear and consistent explanations, making them suitable for industrial applications.<sup>52,53</sup> However, with the increasing use of deep learning models, which are often treated as black boxes, it has become challenging to understand and explain how such models make predictions. To address this, developers are encouraged to use *post hoc* explanation methods, such as the SHAP values,<sup>54</sup> Local Interpretable Model-agnostic Explanations,<sup>55</sup> and integrated gradients. These methods enable feature importance analysis and provide explanations of model predictions in both global and local contexts. In addition, causal inference techniques can enhance model interpretability by identifying cause-and-effect relationships between input variables and model outcomes. Approaches such as structural causal models, counterfactual analysis, and invariant causal prediction can be used to provide more robust and stable explanations, particularly in dynamic industrial settings.<sup>56</sup>

#### **4.3.3. Model adaptation and generalization**

Model adaptation refers to the ability of an AI model to adjust its parameters or structure to accommodate new data distributions, domains, or operational scenarios without requiring complete retraining. Meanwhile, model

generalization refers to the model's ability to maintain accurate predictions when exposed to previously unseen data or tasks. They are essential to ensure that models remain robust, reliable, and effective in real-world industrial settings. Several key techniques support model adaptation and generalization. Transfer learning and domain adaptation allow models trained in one context or domain to be fine-tuned or adjusted for use in another, reducing the need for large amounts of newly labeled data.<sup>57,58</sup> Federated learning enables decentralized model training across multiple industrial sites, allowing collaborative improvement of models while preserving data privacy and security.<sup>59</sup> In addition, few-shot and zero-shot learning techniques allow models to make accurate predictions in new domains with limited (few-shot) or no (zero-shot) task-specific training data.<sup>60,61</sup> These are especially valuable when collecting and labeling new industrial data is time-consuming or expensive. Online learning allows models to continuously update as new data streams become available, maintaining relevance and accuracy in real-time applications.<sup>62</sup> Likewise, continual learning techniques enable models to incrementally incorporate knowledge from new data without catastrophic forgetting of previously learned information, supporting long-term model evolution in dynamic environments.<sup>63</sup>

#### **4.3.4 Model optimization and deployment readiness**

Model optimization is essential for improving the performance, efficiency, and reliability of AI models in industrial applications. It involves tuning model parameters, selecting appropriate architectures, and improving computational efficiency. Common techniques include hyperparameter optimization methods, such as grid search, random search, and Bayesian optimization, which systematically explore the parameter space to find optimal configurations.<sup>64</sup> In addition to hyperparameter tuning, model optimization also addresses structural design and computational efficiency to meet industrial requirements. Neural architecture search (NAS) methods, including differentiable NAS and evolutionary algorithms, can automate the discovery of architectures that achieve optimal trade-offs between accuracy, latency, and memory consumption.<sup>65</sup> Model compression techniques, such as pruning (magnitude-based or structured pruning), weight quantization, and low-rank factorization, are commonly used to reduce inference time and deployment costs without significant loss of performance.<sup>66</sup> Knowledge distillation further enhances deployment readiness by transferring knowledge from large teacher models to smaller student models.<sup>67</sup> Furthermore, multi-condition testing could be performed to ensure that models perform reliably across different operating scenarios.<sup>68</sup>

#### **4.4. Industrial AI methodology platform**

The industrial AI methodology platform provides structured approaches, architectures, and guiding principles to ensure AI applications are systematically integrated into industrial systems. This platform includes multiple well-established methodologies that shape AI integration in industrial settings. Examples include: (i) ILKM: ILKM bridges LLMs with domain-specific industrial knowledge to support reasoning, explanation, and contextual decision-making in industrial AI;<sup>7</sup> (ii) 5C CPS: CPS provides a structured approach for integrating AI with industrial systems through five levels (connection, conversion, cyber, cognition, and configuration), enabled by operation technology, analytic technology, data technology, and platform technology;<sup>4,14</sup> (iii) stream-of-quality and stream-of-X: This shows a paradigm that defines a structured methodology for continuous monitoring, optimization, and decision-making in industrial systems, particularly for multi-stage manufacturing processes;<sup>20</sup> and (iv) digital twin: It integrates data, simulation, and services to create a virtual representation of physical entities, achieving industrial intelligence by utilizing AI, IoT, and ML techniques.<sup>19</sup>

While existing methodologies have played an important role in advancing industrial AI, their implementation can be further enhanced by the proposed industrial AI foundation framework. Rather than redefining the core methodologies, the knowledge, data, and model modules offer comprehensive and systematic views that support both existing and emerging industrial AI approaches. By promoting more structured development processes, they ensure that the design, implementation, and deployment of industrial AI solutions are more effective, reliable, and scalable. A concrete example is using ILKM to enable intelligent question-answering (QA) systems for maintenance personnel, who frequently need to query complex maintenance manuals and troubleshooting guides during equipment servicing. The knowledge module enables the structured extraction of key procedures, fault descriptions, and parameter settings from maintenance manuals and technical documentation. This content is organized into knowledge graphs and indexed repositories that can be efficiently queried by LLMs. The data module ensures that domain-specific datasets – including historical maintenance logs and annotated QA pairs – are properly documented, versioned, and linked to the knowledge base. Standardized metadata (e.g., machine type, operating conditions, and fault categories) allows the LLM to retrieve relevant answers aligned with specific equipment and scenarios. The model module guides the fine-tuning of LLMs on domain-specific QA datasets. It also defines best



AI foundation framework, guiding knowledge extraction, data preparation, model development, and evaluation in a structured and systematic manner.

The process began with the knowledge module. Under its guidance, domain knowledge was extracted from existing publications related to the dataset by leveraging the latest LLMs, GPT-4o, through OpenAI's application programming interface. Researchers interacted with the LLM using targeted research questions such as: "Can you provide the background information about this dataset?;" "In [target] paper, how did the authors perform data processing?;" and "How can one develop specific ML models such as 1D-CNN, LSTM, or Transformer for this task?" GPT-4o provided structured responses, summaries, and code generation examples, helping researchers quickly understand the dataset's background, signal characteristics, data preparation requirements, feature extraction strategies, and model development practices.

The data module was applied to ensure systematic preprocessing. A structured pipeline was designed under the guidance of domain knowledge from prior studies. The process began with data cleaning to improve data quality, followed by data segmentation to increase the number of usable samples for model training. Next, feature engineering was performed using the fast Fourier transform to convert raw time-series signals into frequency domain features. The dataset was then split into training, validation, and test sets in a 3:1:1 ratio to ensure robust model evaluation. Throughout this stage, researchers continued to interact with LLMs to validate the soundness of their preprocessing strategies or to quickly obtain code examples for implementing new ideas.

After ensuring that the data were AI-ready, in the model module, researchers developed and evaluated eight different AI models based on previous methodologies and their expertise. These included: (i) tree-based model (decision tree and random forest); (ii) CNN-based model (naive 1D-CNN, residual 1D-CNN); (iii) long short-term memory (LSTM)-based model (naive LSTM, bi-LSTM, hybrid-LSTM); and (iv) transformer-based model (vanilla transformer). Deep learning models were selected for their ability to automatically extract complex patterns from large-scale, frequency-domain features and for their proven robustness in handling variations in operating conditions without the need for extensive manual feature engineering. Hyperparameter tuning and performance evaluation were performed following the experience indicated in the previous research and researchers' development experience.

The classification accuracy and confusion matrix are presented in [Figure 3](#), demonstrating high performance

across various architectures, with the transformer model achieving the highest accuracy of 99.47%. These results highlight the potential of advanced deep learning architectures in industrial fault diagnosis. Compared with previous studies on similar gearbox fault classification tasks, the proposed approach demonstrates clear improvement. For instance, Su and Lee<sup>69</sup> developed a residual CNN that achieved 96.99% accuracy, Vaerenberg *et al.*<sup>70</sup> used power spectral density preprocessing, log normalization, and a 3-layer CNN to reach 96.9% accuracy, and Gauriat *et al.*<sup>71</sup> proposed multi-class neural additive models with 92.03% accuracy. The higher performance achieved in this study is attributed to not only the model design but also the systematic application of the industrial AI foundation framework. The structured guidance from the knowledge, data, and model modules ensured consistent data preprocessing, appropriate model selection, and effective hyperparameter tuning, ultimately enhancing reliability, scalability, and real-world applicability.

## 6. Future direction

To further strengthen the industrial AI foundation framework, several key directions require further exploration. One critical aspect is talent development. Incorporating 4P-based learning (principle, practice, problem-solving, and professional) and interdisciplinary training in AI/ML, engineering, and industrial applications could benefit the next generation of industrial AI practitioners. Another promising direction is data foundry, which aims to establish a standardized framework for industrial dataset collection, annotation, benchmarking, and management. A well-structured data foundry would enhance collaborative AI research, reproducibility, and cross-industry data sharing while also facilitating the hosting of the Industrial AI Data Challenge Competitions to promote innovation and benchmark AI model performance on industrial datasets. Meanwhile, the foundation framework can be extended toward discrete event dynamic systems and hybrid control systems. This would involve adapting the knowledge, data, and model modules to better handle event-based transitions, symbolic representations, and hierarchical system logic. Furthermore, future research should explore the development of an LLM-assisted intelligent knowledge management system to better make use of historical case studies, domain expertise, and best practices. Such a system could autonomously acquire, structure, and retrieve relevant information, providing researchers and engineers with contextualized and actionable insights to improve AI model development and decision-making in industrial applications.

## 7. Conclusion

This paper proposes a unified industrial AI foundation framework, structured into three core modules – the knowledge module, data module, and model module – which collectively support and enhance the industrial AI methodology platform. The role of industrial AI is explored in detail, and a case study on intelligent rotating machinery diagnosis demonstrates the framework's potential. Furthermore, several future directions are discussed. Overall, our unified industrial AI foundation framework provides a systematic approach to developing, validating, and deploying industrial AI solutions for future industries. Despite these contributions, we acknowledge that acquiring and labeling industrial data remains a significant challenge. Furthermore, while the framework leverages LLMs for knowledge extraction, human guidance is still required, and computational costs may limit its deployment in certain industrial settings. Moving forward, we encourage further research and collaboration between academia, industry, and AI practitioners to refine and expand this framework, fostering the next generation of intelligent, knowledge-data-driven industrial AI ecosystems.

## Acknowledgments

None.

## Funding

None.

## Conflict of interest

Jay Lee is an Editorial Board Member of this journal but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

## Author contributions

*Conceptualization:* All authors

*Writing–original draft:* Hanqi Su

*Writing–review & editing:* All authors

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data

The dataset used in the case study is available for download from <https://data.phmsociety.org/phm2023-conference-data-challenge/>.

## References

1. Jan Z, Ahamed F, Mayer W, *et al.* Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Syst Appl.* 2022;216:119456. doi: 10.1016/j.eswa.2022.119456
2. Peres RS, Jia X, Lee J, Sun K, Colombo AW, Barata J. Industrial artificial intelligence in industry 4.0 - systematic review, challenges and outlook. *IEEE Access.* 2020;8:220121-220139. doi: 10.1109/access.2020.3042874
3. Leng J, Zhu X, Huang Z, *et al.* Unlocking the power of industrial artificial intelligence towards Industry 5.0: Insights, pathways, and challenges. *J Manuf Syst.* 2024;73:349-363. doi: 10.1016/j.jmsy.2024.02.010
4. Lee J, Davari H, Singh J, Pandhare V. Industrial artificial intelligence for industry 4.0-based manufacturing systems. *Manuf Lett.* 2018;18:20-23. doi: 10.1016/j.mfglet.2018.09.002
5. Sisinni E, Saifullah A, Han S, Jennehag U, Gidlund M. Industrial internet of things: Challenges, opportunities, and directions. *IEEE Trans Industr Inform.* 2018;14(11):4724-4734. doi: 10.1109/tii.2018.2852491
6. Khalil RA, Saeed N, Masood M, Fard YM, Alouini MS, Al-Naffouri TY. Deep learning in the industrial internet of Things: Potentials, challenges, and emerging applications. *IEEE Internet Things J.* 2021;8(14):11016-11040. doi: 10.1109/jiot.2021.3051414
7. Lee J, Su H. A unified industrial large knowledge model framework in Industry 4.0 and smart manufacturing. *Int J AI Mater Des.* 2024;1(2):41-47. doi: 10.36922/ijamd.3681
8. Chang Y, Wang X, Wang J, *et al.* A survey on evaluation of large language models. *ACM Trans Intell Syst Technol.* 2024;15(3):1-45. doi: 10.1145/3641289
9. Zio E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliab Eng Syst Saf.* 2021;218:108119. doi: 10.1016/j.res.2021.108119

10. Su H, Lee J. Machine learning approaches for diagnostics and prognostics of industrial systems using open source data from PHM data challenges: A review. *Int J Progn Health Manag.* 2024;15(2):1-26.  
doi: 10.36001/ijphm.2024.v15i2.3993
11. Yan J, Meng Y, Lu L, Li L. Industrial big data in an industry 4.0 environment: Challenges, schemes, and applications for predictive maintenance. *IEEE Access.* 2017;5:23484-23491.  
doi: 10.1109/access.2017.2765544
12. Wang J, Xu C, Zhang J, Zhong R. Big data analytics for intelligent manufacturing systems: A review. *J Manuf Syst.* 2021;62:738-752.  
doi: 10.1016/j.jmsy.2021.03.005
13. Pivoto DGS, de Almeida LFF, da Rosa Righi R, Rodrigues JJPC, Lugli AB, Alberti AM. Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. *J Manuf Syst.* 2021;58:176-192.  
doi: 10.1016/j.jmsy.2020.11.017
14. Lee J, Bagheri B, Kao HA. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf Lett.* 2014;3:18-23.  
doi: 10.1016/j.mfglet.2014.12.001
15. Zhang X, Ming X, Liu Z, Yin D, Chen Z, Chang Y. A reference framework and overall planning of industrial artificial intelligence (I-AI) for new application scenarios. *Int J Adv Manuf Technol.* 2018;101(9-12):2367-2389.  
doi: 10.1007/s00170-018-3106-3
16. Yang T, Yi X, Lu S, Johansson KH, Chai T. Intelligent manufacturing for the process industry driven by industrial artificial intelligence. *Engineering.* 2021;7(9):1224-1230.  
doi: 10.1016/j.eng.2021.04.023
17. Ahmed I, Jeon G, Piccialli F. From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where. *IEEE Trans Industr Inform.* 2022;18(8):5031-5042.  
doi: 10.1109/tii.2022.3146552
18. Tao F, Zhang H, Liu A, Nee AYC. Digital twin in industry: State-of-the-art. *IEEE Trans Industr Inform.* 2019;15(4):2405-2415.  
doi: 10.1109/tii.2018.2873186
19. Lee J, Azamfar M, Singh J, Siahpour S. Integration of digital twin and deep learning in cyber-physical systems: Towards smart manufacturing. *IET Collab Intell Manuf.* 2020;2:34-36.  
doi: 10.1049/iet-cim.2020.0009
20. Lee J, Gore P, Jia X, Siahpour S, Kundu P, Sun K. Stream-of-quality methodology for industrial internet-based manufacturing system. *Manuf Lett.* 2022;34:58-61.  
doi: 10.1016/j.mfglet.2022.09.004
21. Buchgeher G, Gabauer D, Martinez-Gil J, Ehrlinger L. Knowledge graphs in manufacturing and production: A systematic literature review. *IEEE Access.* 2021;9:55537-55554.  
doi: 10.1109/access.2021.3070395
22. Zhong L, Wu J, Li Q, Peng H, Wu X. A comprehensive survey on automatic knowledge graph construction. *ACM Comput Surv.* 2023;56(4):1-62.  
doi: 10.1145/3618295
23. Pan S, Luo L, Wang Y, Chen C, Wang J, Wu X. Unifying large language models and knowledge graphs: A roadmap. *IEEE Trans Knowl Data Eng.* 2024;36(7):3580-3599.  
doi: 10.1109/tkde.2024.3352100
24. Kirk JR, Wray RE, Lindes P, Laird JE. Improving knowledge extraction from LLMs for task learning through agent analysis. *Proc AAAI Conf Artif Intell.* 2024;38(16):18390-18398.  
doi: 10.1609/aaai.v38i16.29799
25. Han X, Cao S, Lv X, et al. OpenKE: An Open Toolkit for Knowledge Embedding. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.* Association for Computational Linguistics; 2018.  
doi: 10.18653/v1/D18-2024
26. Zhang H, Khashabi D, Song Y, Roth D. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20).* 2020.  
doi: 10.24963/ijcai.2020/554
27. Li Y, Zou L. gBuilder: A Scalable knowledge graph construction system for unstructured corpus. *arXiv.* 2023.  
doi: 10.48550/arXiv.2208.09705
28. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *arXiv.* 2020.  
doi: 10.48550/arXiv.2005.14165
29. Achiam J, Adler S, Agarwal S, et al. GPT-4 technical report. *arXiv.* 2024.  
doi: 10.48550/arXiv.2303.08774
30. Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and efficient foundation language models. *arXiv.* 2023.  
doi: 10.48550/arXiv.2302.13971
31. Touvron H, Martin L, Stone K, et al. LLaMA 2: Open foundation and fine-tuned chat models. *arXiv.* 2023.  
doi: 10.48550/arXiv.2307.09288
32. Chowdhery A, Narang S, Devlin J, et al. PaLM: Scaling language modeling with pathways. *arXiv.* 2022.

- doi: 10.48550/arXiv.2204.02311
33. Liu A, Feng B, Wang B, *et al.* DeepSeek-V2: A strong, economical, and efficient mixture-of-experts language model. *arXiv*. 2024.  
doi: 10.48550/arXiv.2405.04434
  34. Giner-Miguel J, Gómez A, Cabot J. Using large language models to enrich the documentation of datasets for machine learning. *arXiv*. 2024.  
doi: 10.48550/arXiv.2404.15320
  35. Vaithilingam P, Zhang T, Glassman EL. Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models. In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts)*. 2022.  
doi: 10.1145/3491101.3519665
  36. Fan W, Ding Y, Ning L, *et al.* A survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. In: *Proceedings of the 28<sup>th</sup> ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024. p. 6491-6501.  
doi: 10.1145/3637528.3671470
  37. Strobelt H, Webson A, Sanh V, *et al.* Interactive and visual prompt engineering for ad-hoc task adaptation with large language models. *IEEE Trans Vis Comput Graph*. 2022;29:1146-1156.  
doi: 10.1109/tvcg.2022.3209479
  38. He Q, Zeng J, Huang W, *et al.* Can large language models understand real-world complex instructions? *Proc AAAI Conf Artif Intell*. 2024;38(16):18188-18196.  
doi: 10.1609/aaai.v38i16.29777
  39. Cofre-Martel S, Lopez Droguett E, Modarres M. Big machinery data preprocessing methodology for data-driven models in prognostics and health management. *Sensors (Basel)*. 2021;21(20):6841.  
doi: 10.3390/s21206841
  40. Mohammed S, Budach L, Feuerpfeil M, *et al.* The effects of data quality on machine learning performance. *arXiv*. 2024.  
doi: 10.48550/arXiv.2207.14529
  41. Li P, Rao X, Blase J, Zhang Y, Chu X, Zhang C. CleanML: A Study for Evaluating the Impact of Data Cleaning on ML Classification Tasks. In: *2021 IEEE 37<sup>th</sup> International Conference on Data Engineering (ICDE)*. 2021.  
doi: 10.1109/icde51399.2021.00009
  42. Foroni D, Lissandrini M, Velegrakis Y. Estimating the Extent of the Effects of Data Quality through Observations. In: *2021 IEEE 37<sup>th</sup> International Conference on Data Engineering (ICDE)*. 2021.  
doi: 10.1109/icde51399.2021.00176
  43. Zheng A, Casari A. *Feature Engineering for Machine Learning Principles and Techniques for Data Scientists*. Beijing, Boston, Farnham: O'Reilly; 2018.
  44. Allen L, Atkinson J, Jayasundara D, Cordiner J, Moghadam PZ. Data visualization for Industry 4.0: A stepping-stone toward a digital future, bridging the gap between academia and industry. *Patterns (N Y)*. 2021;2(5):100266.  
doi: 10.1016/j.patter.2021.100266
  45. Bertolini M, Mezzogori D, Neroni M, Zammori F. Machine learning for industrial applications: A comprehensive literature review. *Expert Syst Appl*. 2021;175:114820.  
doi: 10.1016/j.eswa.2021.114820
  46. Sarker IH. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput Sci*. 2021;2(6):420.  
doi: 10.1007/s42979-021-00815-1
  47. Xu Y, Kohtz S, Boakye J, Gardoni P, Wang P. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliab Eng Syst Saf*. 2022;230:108900.  
doi: 10.1016/j.res.2022.108900
  48. Li H, Zhang Z, Li T, Si X. A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities. *Mech Syst Signal Process*. 2024;209:111120-111120.  
doi: 10.1016/j.ymsp.2024.111120
  49. Ektefaie Y, Dasoulas G, Noori A, Farhat M, Zitnik M. Multimodal learning with graphs. *Nat Mach Intell*. 2023;5(4):340-350.  
doi: 10.1038/s42256-023-00624-6
  50. Su H, Song B, Ahmed F. Multi-modal Machine Learning for Vehicle Rating Predictions Using Image, Text, and Parametric Data. In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Vol. 87295. American Society of Mechanical Engineers; 2023.  
doi: 10.1115/DETC2023-115076
  51. Zhang H, Semujju SD, Wang Z, *et al.* Large scale foundation models for intelligent manufacturing applications: A survey. *J Intell Manuf*. 2025:1-52.  
doi: 10.1007/s10845-024-02536-7
  52. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: A review of machine learning interpretability methods. *Entropy (Basel)*. 2020;23(1):18.  
doi: 10.3390/e23010018
  53. Li X, Xiong H, Li X., Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond.

- Knowl Inf Syst.* 2022;64(12):3197-3234.  
doi: 10.1007/s10115-022-01756-8
54. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Vol. 30. Curran Associates, Inc.; 2017.
  55. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: *Proceedings of the 22<sup>nd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. 2016. p. 1135-1144.  
doi: 10.1145/2939672.2939778
  56. Li Z, Guo X, Qiang S. A survey of deep causal models and their industrial applications. *ArtifIntell Rev.* 2024;57(11):298.  
doi: 10.1007/s10462-024-10886-0
  57. Singhal P, Walambe R, Ramanna S, Kotecha K. Domain adaptation: Challenges, methods, datasets, and applications. *IEEE Access.* 2023;11:6973-7020.  
doi: 10.1109/access.2023.3237025
  58. Li W, Huang R, Li J, *et al.* A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mech Syst Signal Process.* 2022;167:108487.  
doi: 10.1016/j.ymsp.2021.108487
  59. Banabilah S, Aloqaily M, Alsayed E, Malik N, Jararweh Y. Federated learning review: Fundamentals, enabling technologies, and future applications. *Inf Process Manag.* 2022;59(6):103061.  
doi: 10.1016/j.ipm.2022.103061
  60. Song Y, Wang T, Cai P, Mondal SK, Sahoo JP. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Comput Surv.* 2023;55:1-40.  
doi: 10.1145/3582688
  61. Wang W, Zheng VW, Yu H, Miao C. A survey of zero-shot learning: Settings, methods, and applications. *ACM Trans Intell Syst Technol.* 2019;10(2):1-37.  
doi: 10.1145/3293318
  62. Hoi SCH, Sahoo D, Lu J, Zhao P. Online learning: A comprehensive survey. *Neurocomputing.* 2021;459(1):249-289.  
doi: 10.1016/j.neucom.2021.04.112
  63. Wang L, Zhang X, Su H, Zhu J. A comprehensive survey of continual learning: Theory, method and application. *IEEE Trans Pattern Anal Mach Intell.* 2024;46(8):5362-5383.  
doi: 10.1109/tpami.2024.3367329
  64. Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing.* 2020;415:295-316.  
doi: 10.1016/j.neucom.2020.07.061
  65. Elsken T, Metzen JH, Hutter F. Neural architecture search: A survey. *J Mach Learn Res.* 2019;20(55):1-21.  
doi: 10.48550/arXiv.1808.05377
  66. Deng BL, Li G, Han S, Shi L, Xie Y. Model compression and hardware acceleration for neural networks: A comprehensive survey. *Proc IEEE.* 2020;108(4):485-532.  
doi: 10.1109/jproc.2020.2976475
  67. Gou J, Gou J, Yu B, Maybank SJ, Tao D. Knowledge distillation: A survey. *Int J Comput Vis.* 2021;129(6):1789-1819.  
doi: 10.1007/S11263-021-01453-Z
  68. Kim S, Kim I, You D. Multi-condition multi-objective optimization using deep reinforcement learning. *J Comput Phys.* 2022;462:111263.  
doi: 10.1016/j.jcp.2022.111263
  69. Su H, Lee J. An advanced diagnostic model for gearbox degradation prediction under various operating conditions and degradation levels. *Annu Conf PHM Soc.* 2024;16(1).  
doi: 10.36001/phmconf.2024.v16i1.3869
  70. Vaerenberg R, Marx D, Hosseinli SA, *et al.* Preprocessing and modeling approach for gearbox pitting severity prediction under unseen operating conditions and fault severities. *Int J Progn Health Manag.* 2024;15(1):1-12.  
doi: 10.36001/ijphm.2024.v15i1.3808
  71. Gauriat CM, Pencolé Y, Ribot P, Brouillet G. Multi-class Neural Additive Models: An Interpretable Supervised Learning Method for Gearbox Degradation Detection. In: *2024 IEEE International Conference on Prognostics and Health Management (ICPHM)*. 2024. p. 41-48.  
doi: 10.1109/icphm61352.2024.10627522