LLMs as Cognitive Partners in Shipping 4.0: An Extend AI Approach to Maritime Predictive Maintenance Training

Tsigkris, N. [1], Sklavounou, E. [1], Tseles, D.[1] Papachristos, D. [2]*, Nikitakos, N.[3], Gegenava, A.[4], Kakhidze, A.[4] and Gegenava, T.[4]

[1]Dept. Industrial Design and Production, Univ. of West Attica, Greece

[2] Hellenic Navy Academy, Greece

[3] Dept. of Shipping, Trade & Transport, Univ. of Aegean, Greece

[4] Navigation Faculty, Batumi State Maritime Academy, Georgia

* Corresponding author: dimpap2009@gmail.com;

Abstract. This paper explores the integration of Generative AI (Gen AI) and Explainable AI (XAI) into Predictive Maintenance (PdM) within the Shipping 4.0 framework for educational use. The authors propose a conceptual framework called IPMDP (Intelligent Predictive Maintenance Decision Process for Education), which involves an architectural approach (System-of-System, SoS) for integrating LLMs (Gen AI approach) into the predictive maintenance of marine engines and systems, in order to support decision-making in the maintenance of systems at sea. The central idea of the conceptual framework lies in the utilization of a decision-making model using Gen AI techniques through the ExtendAI framework, which functions as a “cognitive partner” to a ship’s engine room crew. The ultimate goal is the educational use of such a framework in the training of Merchant Marine engineers and other ship engine room personnel. Essentially, this study is a conceptual framework paper, integrating existing models and literature to propose a novel application domain, rather than presenting original empirical findings. This cognitive maintenance approach improves operational safety and ensures human-centric, accountable decision-making, which is crucial for regulatory compliance and effective adoption in the maritime industry.

Key Words: AI, Gen AI, PdM, Shipping Maintenance, XAI.

1. Introduction.

The global industrial landscape is currently undergoing a transformative period known as the "Fourth Industrial Revolution" or Industry 4.0 (I4.0). This era is defined by the digitization of traditional processes, effectively bridging the physical and virtual worlds through the deployment of cyber-physical systems (CPS), the Internet of Things (IoT), and Artificial Intelligence (AI). Within this context, the maritime sector is evolving toward "Shipping 4.0," a framework aimed at creating smart, interconnected, and autonomous industrial systems.

Developments in genetic artificial intelligence (GenAI), such as large language models (LLMs), generative adversarial networks (GANs), and diffusion models, are finding significant applications, including in the field of industrial maintenance, such as the PdM sector. As highlighted in the literature, GenAI offers a range of new capabilities, such as synthetic data generation, human-AI interaction through the use of natural language, and multimodal data analysis, thereby addressing long-standing challenges and creating new fields of application. Particularly in the shipping sector, the application of GenAI for PdM remains in its early stages, although there are constantly new efforts to implement it. A critical gap identified is the lack of an established standard for integrating it not merely as a recommendation engine but as a cognitive partner, a mechanism that respects the expertise of marine engineers, provides explainable results, and is suitable for educational use in training future merchant marine personnel (Li et al. 2026; Gan et al. 2023; Reicherts et al. 2025).

To address the gaps identified in the literature, this paper proposes a conceptual framework for maritime education called IPMDP-MME (Intelligent Predictive Maintenance Decision Process for Merchant Marine Education). The framework adopts a System-of-System (SoS) architectural approach to integrate GenAI, Explainable AI (XAI) & PdM in shipboard engine room personnel training. The central methodological innovation lies in the adoption of the ExtendAI decision-making model (Reicherts et al., 2025), where an LLM acts as a “cognitive partner” assisting the reasoning process rather than providing direct recommendations (RecommendAI). Unlike conventional systems, the adoption of the ExtendAI model offers the learner the ability to first formulate his own hypothesis, through which the system provides local suggestions, reflective questions and XAI-verified technical justifications based on real sensor data and knowledge graphs (KGs). This approach is the most pedagogically optimal approach so that the final maintenance decision remains under human control, while enhancing the cognitive ability of the user-learner. Furthermore, the use of the proposed framework in the educational practice provides advantages in terms of learning outcomes as well as operational security.

The structure of this paper is organized as follows: *Section 2* presents a literature review covering maintenance strategies, PdM in shipping, GenAI applications in maintenance, AI-assisted decision-making models, and the role of LLMs in education, *Section 3* discusses the adoption of Explainable AI (XAI) as a critical parameter for trust and accountability in maritime PdM, **S***ection 4* details the proposed IPMDP-MME conceptual framework, including its SoS architecture, and a step-by-step methodological roadmap. Finally, *Section 5* concludes a summary of findings, limitations, and future research directions.

2. Literature Review

2.1 Maintenance, PdM & AI in Shipping Sector

Maintenance plays an important role in the industrial sector, as its costs can account for a significant percentage of a company's total production costs (Bevilacqua and Braglia, 2000). Effective maintenance strategies prevent unexpected production interruptions, reduce costs, and can even increase the useful life of industrial machinery. For these reasons, maintenance approaches have undergone transformations as a result of the concerns and efforts of researchers, engineers, technicians, and specialists. Figure 1 illustrates this evolution over time (Nunes et al., 2023). Furthermore, maintenance approaches in industries are changing in response to the higher risks posed by complex equipment, greater demand for production, and the need to keep costs as low as possible. Over time, thinking about maintenance has focused mainly on three basic ideas: *reacting to problems, working to prevent them, and using predictions*.

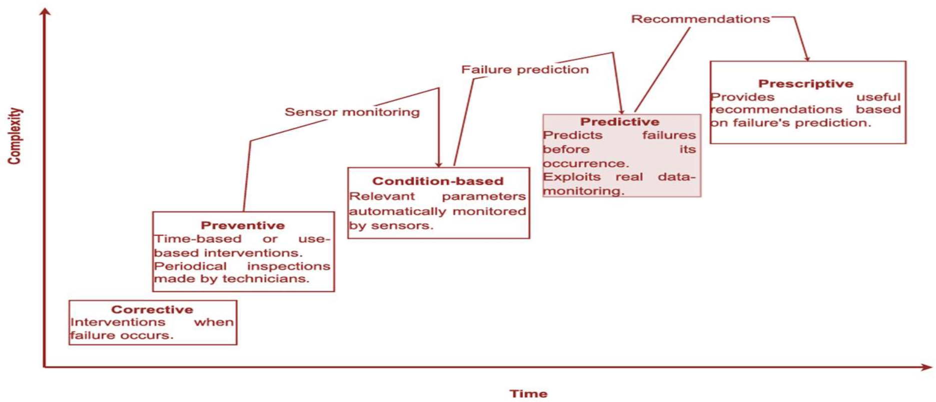

Figure 1. Historical Development of Maintenance Strategies (Nunes et al. 2023)

Maintenance strategies have evolved from a basic reactive model to more sophisticated proactive ones. The most elementary approach is Corrective Maintenance (CM), also known as "run-to-failure," where repairs are only made after a component has failed and disrupted equipment functionality. Driven by the high cost of unscheduled downtime, the field progressed to proactive strategies, beginning with Preventive Maintenance (PM). PM aims to prevent failures through regularly scheduled inspections by specialists and the replacement of parts based on fixed time intervals or a set number of operational cycles, as per manufacturer guidelines. However, this schedule-driven nature can result in two suboptimal outcomes: the premature replacement of components that still have substantial service life, leading to increased material costs (Mobley, 2002), or the unexpected failure of a part before its scheduled maintenance, requiring a costly emergency CM action that may have severe operational consequences (Hao et al., 2010).

The integration of Industry 4.0 and the Internet of Things (IoT) has shifted maintenance from scheduled approaches to Condition-Based Maintenance (CBM), which uses smart sensors to monitor equipment in real-time and trigger actions when parameters deviate from normal. As an evolution of CBM, Predictive Maintenance (PdM) employs Cyber-Physical Systems (CPS), Big Data, and artificial intelligence (AI) to forecast failures and estimate a component's Remaining Useful Life (RUL)(Table 1). This enables optimized, dynamic maintenance planning. Thus, PdM performs maintenance only when needed, making it highly economical. It relies on continuous monitoring and AI-driven analysis of historical and real-time data (using machine learning, statistical inference, and industrial indicators) to diagnose future failures and determine the optimal time for intervention. This field is competitively driven, with manufacturers striving to perfect these technologies (Bagheri et al., 2015).

Table 1. PdM Characteristics.

Description	
Key points	- The operational state of machinery is evaluated through the analysis of key indicators, including vibration, temperature, and acoustic measurements. - By predicting the likelihood of future equipment failures, PdM solutions facilitate maintenance actions at the most opportune moment. - The application of machine learning and analytics provides maintenance personnel with valuable, implementable information.
Advantages	- Minimized production downtime: Early detection of irregularities significantly lowers the risk of catastrophic equipment breakdowns. - Appropriate maintenance interventions extend the operational lifespan and preserve the functionality of machinery. - Optimized maintenance costs: Conducting solely essential servicing activities leads to more effective resource allocation. - Avoidance of accidents and environmental mishaps is made possible by predicting problems. - Contemporary Predictive Maintenance (PdM) minimizes waste generation and energy consumption by optimizing asset performance.

The shipping industry encompasses numerous activities, which are expected to undergo significant changes in the coming years as a result of technological developments and, in particular, the widespread expansion of AI use. According to the Organisation for Economic Co-operation and Development (OECD) report "*The Ocean Economy in 2030*", shipping industries will double their contribution to global value creation by 2030, due to the ongoing need for maritime trade. However, these estimates are directly influenced by the effective adoption of new technological innovations introduced under Industry 4.0 and inextricably linked to the framework of Shipping 4.0. Industry 4.0, as a consequence of the fourth industrial revolution, is expected to pave the way for smart, interoperable, and fully autonomous industrial systems in real time. This vision is based on the convergence of information and communication technologies, combined with cyber-physical systems (CPS), the Internet of Things (IoT), and cloud computing (CC) (Trakadas et al., 2019; Mahale et al., 2025).

Ship maintenance is a critical link in the chain of ship operations, ensuring their continued functionality, efficiency, and safety. However, it is also one of the biggest cost factors for shipping companies, accounting for 20% to 30% of a ship's total operating expenses (OPEX). In addition, this cost is directly linked to prolonged periods of downtime, resulting in reduced operational availability and financial returns. Predictive Maintenance (PdM) in shipping aims to proactively forecast equipment failures by analyzing real-time sensor data (e.g., from engine vibration and temperature). This data-driven approach identifies patterns that indicate when maintenance is truly needed, optimizing scheduling, minimizing both planned and unplanned downtime, reducing costs, and eliminating unnecessary tasks. As part of Shipping 4.0, PdM is expected to replace traditional strategies by enabling early detection of machinery issues. Despite this need, predictive approaches for ship reliability remain under-explored. This gap is critical, as studies show machinery damage (e.g., engine failure) was the leading cause of maritime accidents from 2009 to 2018, accounting for over a third of incidents, underscoring the urgency for predictive solutions (Kalafatelis et al., 2024; Aiello et al., 2019).

Thus, notable instances of machine failure incidents are prevalent within the maritime industry, highlighting the pressing need for proactive PdM strategies. In recent years, several publications in the international literature have analyzed different aspects of the diagnosis and prognosis of PdM-related faults. For example, several articles emphasize fault diagnosis for various ship subsystems e.g., battery system assessment for their Remaining Useful Life (RUL) and State of Health, and corrosion detection in marine structures (e.g., ship hull)(Kalafatelis et al., 2024).

Several key reviews exist in the field of Vessel Prognostics and Health Management (VPHM). These include overviews of methods for forecasting equipment lifespan (Remaining Useful Life), systematic comparisons of maintenance strategies (from Reactive to Predictive), and examinations of the role of classification societies in setting standards. While these studies provide a solid foundation for understanding VPHM, they leave a specific research gap unaddressed. The existing literature has not adequately explored the challenges and limitations in current real-world implementations of Artificial Intelligence (AI), particularly issues related to data transmission and privacy (Karatuğet al., 2023; Liang et al., 2024).

The next table shows a recent summary of AI methods in ship maintenance (Kalafatelis et al., 2024).

Table 2. AI Methods in Ship Maintenance.

Category of Methods	Indicative Models	Field of Application	Main Benefits
Statistics / Linear models	Ordinary Least Squares (OLS), LASSO, Polynomial Regression, Exponentially Weighted Moving Average (EWMA)	Fuel consumption prediction, Exhaust Gas Temperature (EGT) monitoring etc.	Simplicity, interpretability, low resource requirements
Kernel-based models	SVR, Kernel Ridge Regression (KRLS), Relevance Vector Machines (RVM)	RUL batteries, propulsion decay etc.	High accuracy in nonlinear systems
Tree-based models	Random Forest, Bagging, Gradient Boosting, XGBoost	Fault diagnosis piston rings etc.	Robustness, resilience in Noise & disturbances
Neural Networks (NN) & Deep Learning (DL)	FNN, DNN, CNN, DBN, Autoencoder, GAN	Fault classification, RUL prediction, EGT forecasting	High accuracy, complex learning capability
LSTM / Sequence Models	LSTM, Gated Recurrent Unit (GRU), Nonlinear Autoregressive Network with Exogenous Inputs (NARX), Weibull-Time-to-Event (WTTE)- Recurrent Neural Network (RNN)	Time-series prediction, RUL etc.	Holistic understanding of dynamic systems
Digital Twins (DT)	Hybrid DT-ML models	Hull/propeller fouling, speed loss, maintenance scheduling	Real-time optimization & simulation
Federated Learning	Federated Averaging (FedAvg), Asynchronous FL	FL-based PdM in Fleets	Privacy, lower bandwidth, collaborative learning

2.2 *Gen AI in Maintenance*

The introduction of Artificial Intelligence (AI) offers significant prospects for Predictive Maintenance (PdM), as it can detect potential issues (anomalies) within sensor data. This is why it has been the subject of extensive research activity. For instance, Wang et al. (2018) examined how deep learning supports various levels of analytical processing, from fault diagnosis to corrective recommendations. In the relative literature, several review articles have highlighted the value of AI in PdM, while also pointing out challenges related to the selection and integration of models (Carvalho et al., 2019; Zonta et al., 2020; Zonta et al., 2022). Moreover, in industry, various AI applications have been developed for fault detection and maintenance optimisation (Li et al., 2026).

The emergence of Gen AI (Generative AI) in AI is considered a turning point in the history of AI, comparable to the invention of the internet or smartphones. Gen AI is already affecting dozens of sectors: from education and health to marketing and programming. It creates new opportunities but also challenges (issues of employment, misinformation, ethics). LLMs force us to reinterpret what "intelligence," "creativity," and "understanding" mean. Essentially, Gen AI & LLMs are leading to a radically transformative era that opens up new possibilities but also requires responsibility and a response to the new challenges that arise.

Gen AI encompasses various types; each tailored for specific tasks or forms of media generation. The following are some of the more well-known types: Generative Adversarial Networks (GANs), Transformer-based Models (TRMs), Variational Autoencoders (VAEs), and Diffusion models (DMs), to name a few (Sengar et al., 2024).

Generative AI (Gen AI) excels at solving complex real-world problems like as Predictive Maintenance (PdM). Specifically, The use of GenAI in industrial maintenance is expected to provide a future model of Predictive Maintenance (PdM) in which human operators, databases, and complex systems will be fully and seamlessly integrated, sensor data and maintenance records will be converted into interpretable, actionable knowledge, and a series of high-fidelity failure scenarios will be generated to overcome data gaps. The figure below illustrates the potential implications of GenAI in the field of PdM according to (Figure 2) (Li et al., 2026):

- Collaborative Diagnosis use GenAI tools (LLMs) to facilitate human-machine interactions (HMI);
- Instruction Generation utilizes GenAI tools (LLMs) to create task-specific maintenance guidance;
- KG-Augmented PdM integrates GenAI with domain-specific KGs to apply existing knowledge bases in PdM; and
- Data Augmentation employs image generative models to synthesize diverse, high-fidelity failure data.

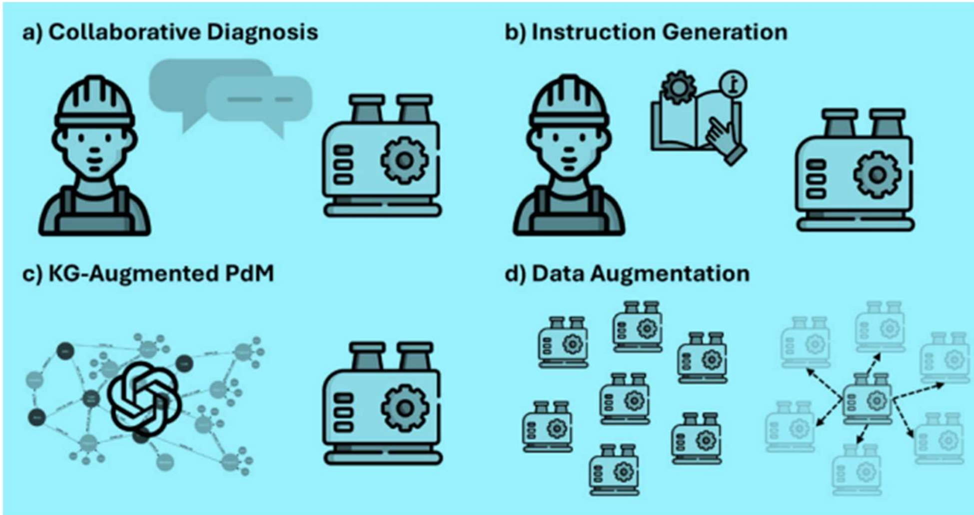

Figure 2. Envisioned Impacts of GenAI on PdM (Li et al., 2026)

The following table summarises the application of GenAI to PdM, where LLMs appear to be particularly effective for PdM tasks that require reasoning ability and text processing, aiming to improve information retrieval, generate domain Knowledge Graphs (KGs), and produce context-aware instructions across a wide range of industrial applications (robots, gearboxes, control systems etc.).

Table 3. Summary of GenAI of PdM applications (Li et al., 2026).

Models	Training	Impact
LLM	PT, FT	Collaborative Diagnosis
LLM	PT, FT	Instruction Generation
LLM	PT, FT	KG-Augmented Diagnosis
VAE	TfS	Data Augmentation
GAN	TfS	Data Augmentation
DM	TfS	Data Augmentation

Note: PT=Prompt tuning, FT=Fine-tuning, TfS=Training from scratch

GenAI (with LLMs) acts as an intelligent intermediary—capturing the domain knowledge from heterogeneous data and human expertise while augmenting the insights from limited historical records, ultimately delivering transparent, timely, and context-aware decision support. Thus, it won't be long before we see a PdM model in which operators, relevant knowledge bases, historical databases, and real-time machine signals are seamlessly integrated with the help of intuitive interfaces (Li et al., 2026).

2.3 AI-Assisted Decision Making in Maintenance

Decision-support tools (DS tool) that utilize AI are designed to work in collaboration with humans with the aim of optimizing performance beyond the limitations of either the human - or AI as the sole decision-making tool. This is referred to as human-AI complementarity (Bansal et al., 2021). Human decision-making is considered a fundamental cognitive function of the brain. Essentially, Decision-making is an extremely complex process, with AI support tools varying significantly in terms of their service delivery context. Research on AI-assisted decision-making focuses primarily on the provision of recommendations generated by AI through model predictions (Bao et al. 2023; Lai et al., 2023). Users must then evaluate and consequently accept or reject the machine's recommendation.

The emergence of GenAI and LLMs is enhancing productivity across various areas of human activity (e.g., software development). However, their use in decision-making presents a host of challenges and limitations, such as (Reicherts et al., 2025):

- Human factor superiority: In complex fields (e.g., investment management), humans perform better than AI tools, despite the vast amount of data that AI processes.
- Human skills: Human judgment, experience, and intuition are still considered decisive factors in decision-making.

Therefore, AI is not yet a universal solution ("panacea"). Any success it achieves depends on the context of its application and requires continuous evaluation and adaptation. However, AI, through GenAI and LLMs, is evolving from a simple tool into a "cognitive partner" that actively engages in reasoning, with two key interaction models, as identified by research (Figure 3)(Reicherts et al., 2025):

1. RecommendAI (Recommendation-centric AI)

   Function: The system analyzes data and directly suggests a ready-made final solution or course of action.

   Mode of thinking: It leads the user into backward reasoning, where they are merely called upon to evaluate and verify the pre-packaged proposal.

   Challenges: It reduces effort but carries the risk of diminished engagement and blind or excessive overreliance, even by experts.

   Countermeasure: Adding explanations or model confidence thresholds has limited success, as people often do not mentally process such information, while models may err with high certainty.

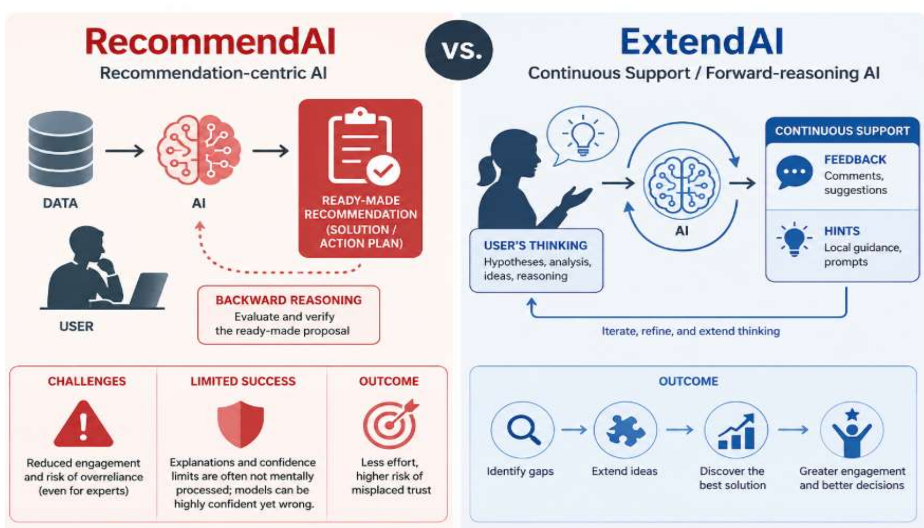

Figure 3. Decision-making models (Reicherts et al., 2025)

2. Extend AI (Continuous Support / Forward-reasoning AI)
   Function: It starts from the user's own reasoning (human-centric approach). The human first articulates their hypotheses or analysis.
   Mode of thinking: It acts as an extension of human reasoning lika a "partner."
   Outcome: Instead of a ready-made solution, the system offers feedback and localized hints, helping the user identify gaps, expand their ideas, and discover the optimal solution on their own.

2.4 LLMs in Education

In recent years, LLMs have represented a significant innovation. For example, OpenAI's GPT series, as well as Google Gemini and Anthropic Claude models, have had a profound impact on the field of AI thanks to their powerful capabilities regarding text generation and comprehension. In the field of education, their introduction is gradually transforming traditional teaching methods and driving innovation in educational technology. Specifically, the rise of GenAI over the last five years, and especially of LLMs as a result of advances in Deep Learning, has led to significant applications in the field of education. In particular, LLMs address key challenges of modern education by offering real-time feedback, reducing educational barriers, and enhancing personalized learning (promoting the Education 4.0 framework). Overall, they demonstrate significant results with continuously expanding prospects. In the near future, they are expected to become a central tool for transforming the educational system toward smarter and more personalized learning (Shang, 2025).

With the continuous development and optimization of LLM technology, it is expected that they will become smarter, capable of understanding and adapting to the personalized learning needs of different learners, and providing more accurate and effective learning support. On the other hand, the spread and development of educational technology may significantly narrow the educational gap between urban and rural areas, and offer more people opportunities for high-quality education. In the future, education may not only involve interaction between teachers and students as individuals, but also collaborative learning between humans and an artificial intelligence model (in the case of LLMs). In this case, teachers transform into learning guides, cultivators of critical thinking, and agents of ethical and humanistic care. As an integral part of education, LLMs will offer learners personalized learning programs. At the same time, those in charge of the respective educational systems must maintain constant attention and reflection regarding humanitarian and ethical issues (Shang, 2025; Gan et al., 2023).

Consequently, the education sector has begun to explore how to integrate LLMs with education to enhance teaching quality and effectiveness. The following Figure illustrates the importance of introducing LLMs and the current practical areas in Education:

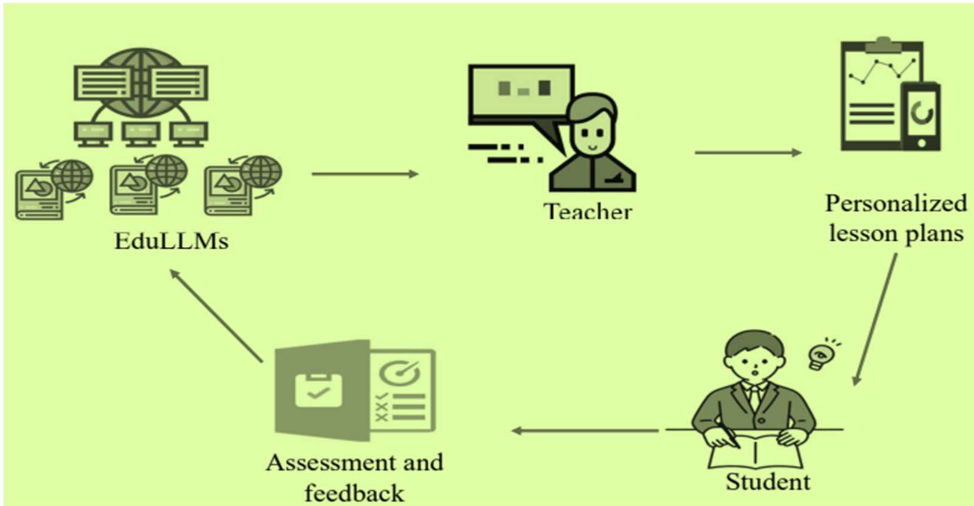

Figure 4. Architecture of LLMs for education (Gan et al., 2025)

In maritime training, AI is being used more and more, particularly in simulator-based training. Related research examines potential anomalies in simulation log data to identify discrepancies in trainee performance. At the same time, eye-tracking technology is used to analyze visual attention, comparing experienced and inexperienced officers via heatmaps, in order to improve electronic navigation skills. Furthermore, AI-powered speech analysis enhances maritime training by combining advanced language models and voice recognition techniques, enabling more precise communication in high-pressure and stressful environments. Finally, another important area of application is the analysis of cognitive load and stress, where biometric indicators such as pupil diameter, electrodermal activity, and heart rate are used to assess the impact of stress on trainee performance during maritime scenarios conducted in maritime simulators (Munim and Kim, 2023; Xue et al., 2024).

3. Adoption Explainable AI- XAI and PdM in Shipping

In the context of the transition to Shipping 4.0, Explainable Artificial Intelligence (XAI) is emerging as a critical factor for the acceptance of Generative AI (Gen AI) models in ship maintenance. While traditional deep learning models (such as GANs and VAEs) offer high accuracy, they often function as black boxes, making it difficult for engineers to understand the logic behind a prediction failure.

One of the most critical challenges for the adoption of Generative AI (Gen AI) in maritime operations is the “black-box” nature that characterizes complex models such as GANs and VAEs. Although these systems demonstrate exceptional accuracy in anomaly detection, their inability to explain the reasoning behind a specific failure prediction creates trust gaps between the system and the ship’s engineers. In the context of Maritime 4.0, where an incorrect diagnosis in critical equipment can lead to costly delays or safety risks, research into Explainable AI (XAI) methods becomes imperative.

The integration of XAI techniques enables the transformation of abstract predictions into understandable technical justifications, for example by linking a detected deviation in vibration patterns to the specific degradation of a bearing. Moreover, the use of XAI functions as a safeguard against LLM "hallucinations," allowing the crew to verify the generated maintenance recommendations based on actual sensor data and manufacturer manuals. In this way, AI ceases to be an isolated tool and is transformed into a reliable troubleshooting copilot, ensuring that decision-making remains under human control.

The integration of XAI can be analyzed through the following axes (Adadi and Berrada, 2018; Arrieta et al., 2020; Gunning and Aha, 2019):

A.Trust Building

- Ship engineers and crew demand transparency and justification for diagnoses concerning critical equipment, such as main engines and generators.
- Research into XAI methods is essential to understand the rationale behind model predictions, thereby strengthening human trust in the system.
- Without explainability, the adoption of advanced PdM systems remains limited due to users' reluctance to rely on unverifiable decisions.

B.Addressing "Hallucinations"

- Large Language Models (LLMs), if not properly fine-tuned, can produce inaccurate content or incorrect diagnoses.
- XAI allows these "hallucinations" to be identified, ensuring that automatically generated maintenance instructions are based on actual data and technical manuals.

C. Improving Decision Making

- Explainable AI transforms sensor data (vibration, temperature) into understandable information, allowing the crew to act proactively.
- By linking the anomalies detected by VAEs to specific operating parameters, XAI helps determine the exact cause of an impending failure.

D. Integrating into the PdM Roadmap

- The implementation of XAI is the connecting link between Failure Diagnosis and Mitigation in the predictive maintenance pyramid.
- It allows systems to act as a "troubleshooting copilot", providing not only the prediction but also the "why" behind the proposed maintenance action.
- The table (Tab.4) below highlights the added value of XAI for decision-making by the crew (Human-in-the-loop).

Table 4. Comparative Analysis between Traditional AI and Explainable AI (XAI) in Maritime Maintenance.

Feature	Traditional AI (Black-Box)	Explainable AI (XAI / Glass-Box)
Output Type	Provides only the prediction (e.g., "Failure Probability: 85%")	Provides prediction accompanied by rationale (e.g., "Due to thermal deviation in Bearing X")
Transparency	Low/Opaque. Internal processes and decision logic are hidden	High. Utilizes interpretability methods (e.g., SHAP, LIME) to highlight key features
User Trust	Limited. Engineers may hesitate to intervene without technical justification	High. Engineers understand the "why," leading to confident and timely actions
Error Management	Difficult to identify the root cause of a false positive/negative	Streamlined identification and mitigation of model bias or data drifts
Regulatory Compliance	Challenges in meeting safety standards and Class society requirements	Facilitates certification through traceability and accountability of AI decisions

3.1 Integrating XAI into the PdM Roadmap: Bridging the "Black-Box" Gap

To successfully transition from traditional Predictive Maintenance to *Cognitive Maintenance,* Explainable AI (XAI) must be integrated as a strategic layer within the existing PdM framework. In the proposed roadmap, XAI acts as the critical bridge between algorithmic output and human action:

- *Stage I: Anomaly Detection:* Gen AI models (e.g., VAEs) monitor normal operational behavior and identify deviations.
- Stage *II: Failure Diagnosis & The XAI Layer:* This is where the "Black-Box" is opened. XAI techniques translate complex neural network weights into human-understandable technical justifications (e.g., attributing a vibration spike to specific piston wear).
- *Stage III: Degradation Prognosis:* By understanding the *why* behind a diagnosis, the prediction of Remaining Useful Life (RUL) becomes more reliable for the Chief Engineer.
- *Stage IV: Mitigation & Decision Support:* LLMs acting as "troubleshooting copilots" use the XAI-verified data to generate precise maintenance actions, reducing the risk of "hallucinations".

3.2 Ensuring Accountability through Explainable AI (XAI)

As the maritime industry transitions toward *Shipping 4.0*, the deployment of Generative AI (Gen AI) in maintenance operations introduces a critical requirement for Autonomy. While traditional deep learning models, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), offer high accuracy in detecting engine anomalies, they often operate as "black-box" systems. This lack of transparency poses a significant barrier to adoption, as ship engineers require a clear understanding of *why* a specific failure prediction was made before initiating costly or mission-critical repairs.

To address this, the integration of *Explainable AI (XAI)* serves as a vital bridge between algorithmic output and human decision-making. By utilizing interpretability methods such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations), the system can translate complex sensor deviations into actionable technical justifications (e.g., attributing a vibration spike to specific piston ring wear). Furthermore, XAI plays a crucial role in mitigating the risks associated with *AI hallucinations*—instances where Large Language Models (LLMs) might generate inaccurate maintenance instructions. By providing a traceable rationale for every diagnosis, XAI ensures that the "troubleshooting copilot" remains a reliable tool for the crew. This transparency not only builds trust among shipboard personnel but also facilitates compliance with stringent international safety regulations and classification society standards

4. The Proposed Conceptual Framework (IPMDP-MME)

The proposed framework is based an Intelligent Predictive Maintenance Decision Process integrating GenAI, XAI and PdM for the educational training of Merchant Marine engineers (IPMDP-MME). The baseline of proposed framework defines the system as a GenAI/XAI-supported PdM framework for marine engines and engine-room systems, with ExtendAI acting as a "cognitive partner" rather than a replacement for the Marine engineer's judgement. The main pillars of the proposed framework are (Adadi and Berrada, 2018; Arrieta et al., 2020; Gunning and Aha, 2019; Reicherts et al., 2025; Li et al., 2026; Sengar et al., 2024; Munim and Kim, 2023):

- *Human-Centric Approach*: The trainee first formulates their own hypotheses about a fault, acting as the primary decision-maker.
- *Cognitive Partner*: GenAI does not provide ready-made solutions but expands the user's thinking through feedback and local hints.
- *Explainability (XAI Layer*): Every suggestion/recommendation from the system is documented based on real sensor data and technical manuals, avoiding hallucinations.

The system should be implemented as a *human-in-the-loop educational decision-support System-of-System* **(***SoS)*, not as an maintenance decision-maker without human implication. The trainee remains responsible for the final diagnosis and maintenance plan, while the GenAI provides traceable, evidence-based support. This is consistent with the article's emphasis on accountable, human-centric cognitive maintenance and with the ExtendAI evidence that reasoning-extension integrates better with the user's decision process than direct recommendation.

The architecture of proposed conceptual framework follows the System-of-Systems (SoS) approach (system Engineering methodology) for the integration of heterogeneous technological entities. The systems comprising the SoS are:

- *Data Acquisition System***:** Collects real-time data (e.g., vibrations and temperatures) from the ship's machinery.
- *Knowledge Base System***:** Includes manufacturers' technical manuals and maintenance history records, organized through Knowledge Graphs (KGs).
- *ExtendAI Core (LLM):* The "brain" of the system, which receives the user's reasoning process and interprets it through Retrieval-Augmented Generation (RAG).
- *XAI Interface***:** Visualizes the reasoning behind each recommendation (e.g., through SHAP or LIME) in order to establish trust.

For example, to ensure that the system operates as an ExtendAI model rather than as a RecommendAI system, the following prompt is proposed:

ROLE: "Act as a Cognitive Partner for a Merchant Marine Engineer within a Predictive Maintenance (PdM) environment.

Your mission: Do not provide direct diagnoses or ready-made solutions.

Process:

1. Wait for the trainee to express their initial analysis or fault hypothesis.
2. Analyze the trainee's reasoning and identify gaps based on sensor data (e.g., exhaust gas temperature) and technical manuals.
3. Provide local hints and reflective questions that encourage deeper reasoning (e.g., 'Did you consider the deviation in lubrication pressure?').
4. Every response must be accompanied by technical justification (XAI) in order to minimize hallucinations and ensure transparency.

The GOAL is for the learner to discover the best course of action on their own.

In addition, the next table (Tab.5) shows a use case for Merchant Marine Engineers in a Main Engine Failure.

Table 5. Use Case: Training Merchant Marine Engineers in a Main Engine Failure Scenario.

Scenario	The trainee observes abnormal vibrations in a Main Engine bearing through the simulator environment
User Action	The trainee enters the following assumption into the system: "I believe there is bearing shell wear due to excessive temperature."
ExtendAI Response	The system does not respond with a direct recommendation such as, "Yes, replace the bearing shell." Instead, it replies: "Your assumption regarding the temperature is valid. However, observe the vibration spectrum at frequency X. How does this relate to the lubrication condition?"
Outcome	The trainee reviews the XAI visualizations, correlates the available data, and ultimately reaches a more comprehensive and holistic diagnosis independently

The methodological approach adopted for the transition toward Cognitive Maintenance follows four (4) sequential stages:

- *Stage I: Anomaly Detection*. Utilization of GenAI models (e.g., Variational Autoencoders – VAEs) for the identification of deviations from normal operational conditions.
- *Stage II: Failure Diagnosis & XAI Layer*. Transformation of "black-box" predictions into interpretable technical justifications through Explainable AI mechanisms.
- *Stage III: Degradation Prognosis*. Estimation of the Remaining Useful Life (RUL) based on the validated and explainable diagnostic outcomes.
- *Stage IV: Mitigation & Decision Support*. The trainee, supported by GenAI (LLM) acting as a cognitive "copilot," makes the final maintenance decision regarding the appropriate corrective or preventive action.

IPMDP-MME prioritizes the user's initial reasoning, providing feedback and local hints to guide the learner toward discovering the optimal maintenance path. The system operates in a *Forward-Reasoning* loop:

1. *Human Initiation*: The trainee observes an anomaly and proposes an initial hypothesis (e.g., "Vibration indicates bearing wear").
2. *System Verification*: The ExtendAI engine checks the hypothesis against the *Ship Sensor Node* and *Knowledge Graph*.
3. *Cognitive Extension*: Instead of confirming/denying, the system provides a "Local Hint" (e.g., "Check the lubrication pressure trend before concluding").
4. *Accountable Decision*: The trainee refines their reasoning based on hints and XAI data to reach the final maintenance decision

Also, the following table describes the core nodes of the SoS architecture and their functional roles in the predictive maintenance process:

Table 6. Comparative Analysis between Traditional AI and Explainable AI (XAI) in Maritime Maintenance.

SoS Node	Function Description	Data Interaction
Ship Sensor Node	Real-time data acquisition from marine engines (vibration, temp, pressure).	Streamed to the Analysis Engine.
Knowledge Graph (KG)	Structured repository of technical manuals and historical maintenance logs.	Provides context for RAG.
ExtendAI Engine (LLM)	The cognitive core that evaluates human reasoning against sensor data.	Processes User Inputs + KG context.
XAI Layer	Translates complex AI outputs into human-understandable technical justifications.	Outputs technical evidence for hints.

Finally, the following diagram illustrates the hierarchical relationship between the various systems (SoS), with the user (trainee engineer) and the “cognitive partner” GenAI (LLM) playing central roles. This structure ensures that the educational process remains human-centered, with artificial intelligence serving to augment human thought rather than replace it. More specific, in Figure 5 shows:

- User Interface (UI) / Cognitive Extension Loop: represents the user’s thought process (“User’s Reasoning & Hypothesis”) and initiates the continuous interaction loop (learner-user & AI system/LLM).
- AI & Knowledge Core: The central control core, where the ExtendAI Engine (LLM) processes the user’s input, using Retrieval-Augmented Generation (RAG) to access external knowledge, while the Explainable AI (XAI) Layer provides accurate (reliable, to the extent possible) technical documentation.
- Data Sources & Knowledge Base: The database and knowledge base are fed (RAG input) by information sources related to ship sensor data (Ship Sensor Node) and organized knowledge (Knowledge Graph, KG).

- Methodological Roadmap: At the bottom of the image, the four-stage methodology is visualized: Anomaly Detection, Failure Diagnosis & XAI, Degradation Prognosis (RUL), and Mitigation & Decision Support.

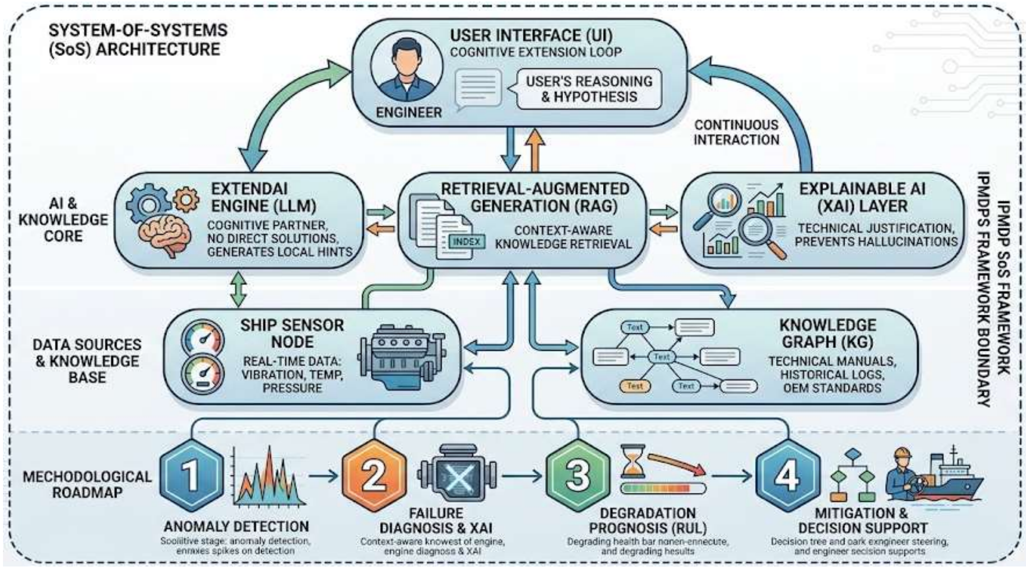

Figure 5. Conceptual framework IPMDP-MME Architecture (by authors)

5. Conclusions

The transition from traditional maintenance models to the Shipping 4.0 era represents a fundamental paradigm shift in maritime operations. It overcomes traditional AI's data scarcity by generating synthetic data for rare failures. In summary, while Gen AI offers transformative potential for the shipping industry, its successful deployment requires a balanced approach that addresses technical integration, data privacy, and the human-centric requirement of interpretability. Future research should focus on refining *Edge Computing* solutions to run these complex models within the resource-constrained environment of sea-going vessels.

As regard the IPMDP-MME the following arise:

- The integration of Explainable Artificial Intelligence (XAI) is critical to making predictions understandable and reliable. Additionally, the use of XAI in combination with Retrieval-Augmented Generation (RAG) from technical manuals and ship sensors is concluded as a necessary strategy to mitigate the inaccurate diagnoses produced by LLMs, ensuring that every indication is substantiated by real data.
- The main conclusion of the proposed IPMDP-MME framework is that Large Language Models (LLMs) should not function as direct recommendation systems (RecommendAI) that provide ready-made solutions. Instead, they should function as "extended intelligence" (ExtendAI), extending the user's reasoning through suggestions, while keeping the human in the central role of decision-making.

- The proposed framework (IPMDP-MME) appears to facilitate compliance with safety regulations and the requirements of classification societies (Class societies), as it provides traceability and accountability for every decision (XAI), while keeping the trainee-user responsible for the final action regarding the maintenance of a system or subsystem.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the Black-Box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access, 6*, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

Aiello, G., Giallanza, A., & Mascarella, G. (2020). Towards shipping 4.0. a preliminary gap analysis. *Procedia Manufacturing, 42*, 24-29. https://doi.org/10.1016/j.promfg.2020.02.019

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Landa, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Bagheri, B., Yang, S., Kao, H. A., & Lee, J. (2015). Cyber-physical systems architecture for self-aware machines in industry 4.0 environment. *IFAC-PapersOnLine, 28*(10), 1622-1627. https://doi.org/10.1016/j.ifacol.2015.09.238

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M.T., & Weld, D. (2021). *Does the Whole Exceed Its Parts? The Effect of AI Explanations on Complementary Team Performance*. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). ACM, Yokohama, Japan, 81:1–81:16. https://doi.org/10.1145/3411764. 3445717

Bao, Y., Gong, W., & Yang, K. (2023). A Literature Review of Human AI Synergy in Decision Making: From the Perspective of Affordance Actualiza tion Theory. *Systems* 11, 9 (Sept. 2023), 442:1–442:20. https://doi.org/10.3390/ systems11090442

Bevilacqua, M., & Braglia, M. (2000). Analytic Hierarchy Process applied to maintenance strategy selection. *Reliability Engineering and System Safety, 70*(1), 71-83. https://doi.org/10.1016/S0951-8320(00)00047-8

Carvalho, T. P., Soares, F. A., Vita, R. , Francisco, R. d. P., Basto, J. P., & Alcala, S. G. (2019). A systematic literature review of machine learning methods applied ´ to predictive maintenance. *Computers & Industrial Engineering 137*, 106024. https://doi.org/10.1016/j.cie.2019.106024

Gan, W., Qi, Z., Wu, J., & Lin, J.C.-W. (2023). "Large Language Models in Education: Vision and Opportunities,". *2023 IEEE International Conference on Big Data (BigData)*, Sorrento, Italy, 2023, 4776-4785. 10.1109/BigData59044.2023.10386291

Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI* Magazine*, 40*(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Hao, Q., Xue, Y., Shen, W., Jones, B., & Zhu, J. (2010). A decision support system for integrating corrective maintenance, preventive maintenance, and condition-based maintenance. *Construction Research Congress 2010*, 470-479. https://doi.org/10.1061/41109(373)48

Kalafatelis, A. S., Nomikos, N., Giannopoulos, A., Krinidis, S., & Tzovaras, D. (2024). *A survey on* predictive *maintenance in the maritime industry using machine and federated learning*. TechRxiv. https://doi.org/10.36227/techrxiv.173471013.33234190/v1

Karatuğ, Ç., Arslanoğlu, Y., & Soares, C. G. (2023). Review of maintenance strategies for ship machinery systems. *Journal of Marine Engineering & Technology, 22*(5), 233-247. https://doi.org/10.1080/20464177.2021.1921323

Lai, V., Chen, C., Smith-Renner, A., Q. Liao, V., & and Tan, C. (2023). Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *Proceedings of the 2023 ACMConference on Fairness, Accountability, and Transparency* (FAccT '23). ACM, Chicago, IL, USA, 1369–1385. https://doi.org/10.1145/3593013.3594087

Li, X., Peng Wang, P., Li, X., Lee, J., Zhou, Q., Byung-Guk J.M., & Gao, R.X. (2026). Generative AI for Predictive Maintenance for Manufacturing: A Survey. *Procedia CIRP*, Volume 139, 35-40. https://doi.org/10.1016/j.procir.2025.09.012

Liang, Q., Knutsen, K. E., Vanem, E., Æsøy, V., & Zhang, H. (2024). A review of maritime equipment prognostics health management from a classification society perspective. *Ocean Engineering, 301*, 117619. https://doi.org/10.1016/j.oceaneng.2024.117619

Mahale, Y., Kolhar, S., & More, A. S. (2025). A comprehensive review on artificial intelligence driven predictive maintenance in vehicles: technologies, challenges and future research directions. *Discovery Applied Sciences, 7*, 243. https://doi.org/10.1007/s42452-024-06392-1

Mobley, R. K. (2002). *An introduction to predictive maintenance*. Elsevier.

Munim, Z., Kim, T. (2023). A Review of Learning Analytics Dashboard and a Novel Application in Maritime Simulator Training. In: *Salman Nazir (eds) Training, Education, and Learning Sciences*. AHFE (2023) International Conference. AHFE Open Access, vol 109. AHFE International, USA. http://doi.org/10.54941/ahfe1003158

Nunes, P., Santos, J., & Rocha, E. (2023). Challenges in predictive maintenance – A review. CIRP *Journal of Manufacturing Science and Technology, 40*, 53-67. https://doi.org/10.1016/j.cirpj.2022.11.002

Reicherts, L., Zhang, Z.T., von Oswald, E., Liu, Y., Rogers, Y., & Hassib, M. (2025). AI, Help Me Think—but for Myself: Assisting People in Complex Decision-Making by Providing Different Kinds of Cognitive Support. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25*). Association for Computing Machinery, New York, NY, USA, Article 255, 1–19. https://doi.org/10.1145/3706598.3713295

Sengar, S., Hasan, A., Kumar, S., & Carroll, F. (2024). *Generative Artificial Intelligence: A* systematic *review and applications*. arXiv. https://doi.org/10.48550/arXiv.2405.11029

Shang, H. (2025). The Current Status and Trends of the Application of Large Language Models in the Field of Education. *Applied and Computational Engineering*. 135. 115-121. 10.54254/2755-2721/2025.21088

Trakadas, P., Nomikos, N., Michailidis, E. T., Zahariadis, T., Facca, F. M., Breitgand, D., Rizou, S., Masip, X., & Gkonis, P. (2019). Hybrid clouds for data-intensive, 5G-enabled IoT applications: An overview, key issues and relevant architecture. *Sensors, 19*(16), 3591. https://doi.org/10.3390/s19163591

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of manufacturing systems* 48, 144–156. 10.1016/j.jmsy.2018.01.003

Xue, H., Haugseggen, Ø., Røds, J.-F., Batalden, B.-M., & Prasad, D.-K. (2024). Assessment of stress levels based on biosignal during the simulator-based maritime navigation training and its impact on sailing route reliability. Transportation Research Interdisciplinary Perspectives, Volume 24, 101047. https://doi.org/10.1016/j.trip.2024.101047

Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., Da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the industry 4.0: A systematic literature review, Computers & industrial engineering 150 (2020) 106889. https://doi.org/10.1016/j.cie.2020.106889

Zonta, T., Da Costa, C. A., Zeiser, F. A., de Oliveira Ramos, G., Kunst, R., & da Rosa Righi, R. (2022). A predictive maintenance model for optimizing production schedule using deep neural networks. *Journal of* Manufacturing *Systems, 62*, 450-462. https://doi.org/10.1016/j.jmsy.2021.12.013