International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

Predictive Maintenance for Industrial Machinery Using LLM

K R Prabha[1], B Nataraj[2], S Sathish[3], C Sujith[4], G Suthakarr[5]
[1,2]Assistant professor, Dept. of ECE, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India.
[3,4,5]UG Scholar, Dept. of ECE, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India.
Emails: *prabha.kr@srec.ac.in[1], nataraj.b@srec.ac.in[2], sathish.2102204@srec.ac.in[3], sujith.2102226@srec.ac.in[4], suthakar.2102230@srec.ac.in[5]*

Abstract
In industrial environments, unexpected machinery failures can result in expensive downtime along with production delays and increased maintenance expenses. Maintenance strategies, such as reactive and preventive maintenance, which are the traditional methods, often fail to provide timely interventions, leading to inefficiencies in industrial operations. This paper presents a predictive maintenance system that integrates Internet of Things (IoT) technology and Large Language Models (LLMs) to monitor and analyze real-time machine health data. The proposed system utilizes Raspberry Pi Pico W, interfaced with temperature, vibration, and current sensors to collect real time operational data. This data is transmitted to Firebase Realtime Database, where it is processed using Lang Chain powered LLMs. The system employs FAISS-based similarity search to retrieve past sensor patterns and generate predictive insights on potential failures. A Flask-based web interface enables real-time monitoring, while an automated alert system via Brevo notifies users of high-risk anomalies, allowing for proactive maintenance decisions. The results demonstrate that integrating IoT-driven real-time monitoring with AI-powered predictive analytics enhances the accuracy of failure detection, reduces unplanned downtime, and improves machine reliability. This work highlights a scalable and cost-effective approach to intelligent predictive maintenance, paving the way for more efficient industrial operations. Future enhancements include the integration of advanced deep learning models and additional sensor modalities to further refine predictive accuracy.

Keywords: *Predictive Maintenance, IoT, Large Language Models, FAISS, Firebase, Industrial Automation.*

1. Introduction
Industrial machinery plays a critical role in manufacturing, production, and automation processes. Unexpected failures in such equipment cause notable downtime and operational inefficiencies which lead to a huge number of financial losses. Traditional maintenance strategies, including reactive maintenance (repairing machinery only after the failure is being noted) and preventive maintenance (scheduled servicing during a fixed intervals of time), often result in either delayed interventions or unnecessary repairs, increasing costs and reducing overall efficiency. The growing complexity of industrial systems demands an intelligent, data-driven approach to machine maintenance. Predictive maintenance (PdM) has emerged as an effective solution by integrating continuous and real-time monitoring and machine learning techniques to forecast potential failures before they occur. The integration of Internet of Things (IoT) sensors enables continuous data collection on key machine parameters, such as temperature, vibration, and current variations. However, traditional predictive maintenance models face challenges in real-time anomaly detection, context-aware failure prediction, and adaptive decision-making. This paper presents a novel predictive maintenance system that combines IoT-enabled sensor networks with Large Language Models (LLMs) powered by Lang Chain to provide real-time failure detection and proactive maintenance recommendations. The proposed system utilizes Raspberry Pi Pico W interfaced with temperature, vibration, and current sensors to capture machine health data. The collected data is stored in Firebase Realtime Database and analysed using LLM-based AI models, enabling pattern recognition, anomaly

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

detection, and automated maintenance insights. Additionally, FAISS-based vector search allows retrieval of historical failure cases for enhanced predictive analysis. A Flask-based web interface provides real-time monitoring, while Brevo-powered email alerts notify users of critical conditions, enabling timely interventions. The primary objective of this work is to develop a scalable, cost-effective, and AI-driven predictive maintenance system that reduces unplanned downtime and enhances industrial efficiency. This paper explores the architecture, implementation, and experimental evaluation of the proposed system, demonstrating how LLM-powered insights combined with IoT-based real-time monitoring can transform industrial maintenance strategies.

2. Literature Survey

Predictive maintenance has significantly evolved with the advancement of artificial intelligence, particularly with the combination of the large language models (LLMs) and domain-specific knowledge. Traditional machine learning techniques have been widely utilized for fault detection, but recent developments in generative AI and multimodal learning have enhanced predictive maintenance capabilities. One of the key innovations in this domain is the application of Generative AI for predictive maintenance. Unlike conventional statistical models, generative AI techniques like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) enable the synthesis of synthetic failure data. This improves fault detection accuracy while addressing the challenges posed by limited historical failure records [1]. These methods also enhance anomaly detection by generating realistic failure scenarios, allowing predictive models to anticipate potential breakdowns more effectively [1]. Another promising approach involves Multimodal Large Language Models (LLMs) for the detection of problems. These models take the input data from various sources like sensor readings, historical logs, and images, to improve decision-making in industrial IoT environments [2]. The fusion of LLMs with smaller models has proven effective in real-time anomaly detection, enhancing the interpretability and accuracy of predictive maintenance frameworks [3]. Further advancements have been made by incorporating domain-specific knowledge bases (DSKBs) into LLMs. This integration enables AI models to leverage industry-specific datasets, improving the reliability and precision of predictive maintenance recommendations [4]. By aligning with structured knowledge bases, LLMs can provide maintenance insights that are tailored to specific industrial applications, thereby reducing downtime and optimizing maintenance schedules. Additionally, the use of user-defined prompts in LLMs has been explored to create flexible, reliable data processing systems. By allowing domain experts to define custom prompts, LLMs can adapt to specific predictive maintenance scenarios, ensuring a more dynamic and context-aware fault detection mechanism [5]. This approach enhances real-time decision-making, making predictive maintenance frameworks more robust. These studies collectively indicate that AI-driven predictive maintenance is transitioning towards automation and self-learning models. The convergence of multimodal AI, generative data synthesis, and domain-specific expertise ensures optimized industrial operations with reduced downtime. Our research aligns with these advancements by integrating sensor based real-time monitoring, FAISS for anomaly detection, and Lang Chain-powered AI analysis, creating a holistic, AI-enhanced predictive maintenance solution for industrial applications.

3. Proposed Method

The proposed predictive maintenance system is designed to continuously monitor industrial machinery using sensor data and analyse it with a Large Language Model (LLM) to predict failures and recommend maintenance actions. The system follows a structured workflow, as illustrated in Fig. 1.

3.1.Sensor Data Collection

The system begins with real-time data acquisition using temperature, vibration, and current sensors. These sensors are deployed on a 5V DC motor to monitor critical operational parameters. The collected data helps detection accuracy while in detecting potential faults that may lead to system failures.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

3.2.Data Transmission to Microcontroller

Once acquired, the sensor data is transmitted to the Raspberry Pi Pico W, which serves as the edge processing unit. The microcontroller processes the incoming signals and sends structured data to a cloud-based database for further analysis.

3.3.Preprocessing of Sensor Data

Before analysis, the raw sensor data undergoes pre-processing to remove noise and anomalies. This step includes filtering out irregular values, normalizing data formats, and structuring it for efficient computation. The cleaned data ensures accurate fault detection and system performance evaluation.

3.4. Integration with Large Language Model (LLM)

The processed data is then integrated with an LLM using Lang Chain. The model performs advanced data interpretation, utilizing a FAISS-based vector similarity search to compare new sensor readings with historical failure patterns. This enables early detection of potential issues.

3.5.Failure Prediction and Diagnosis

The LLM assesses the incoming data patterns and predicts possible failures based on its trained dataset. It identifies the root causes of anomalies and provides a detailed diagnostic report. This real-time failure prediction helps in minimizing downtime and reducing maintenance costs.

3.6.Maintenance Recommendations

Based on the failure analysis, the system generates automated maintenance recommendations. These recommendations provide actionable insights to maintenance personnel, helping them take proactive measures to prevent unexpected breakdowns.

3.7.User Feedback and Continuous Learning

To enhance the accuracy of predictions, user feedback is incorporated into the system. Maintenance engineers can validate or correct the model's recommendations, which helps in refining the LLM's learning process over time. This continuous learning mechanism ensures that the system adapts to evolving operational conditions. The methodology outlined above ensures a structured and efficient approach to predictive maintenance. By leveraging real-time sensor data, cloud-based processing, and AI-driven analysis, the system enables intelligent decision-making for industrial equipment maintenance. (Figure 1)

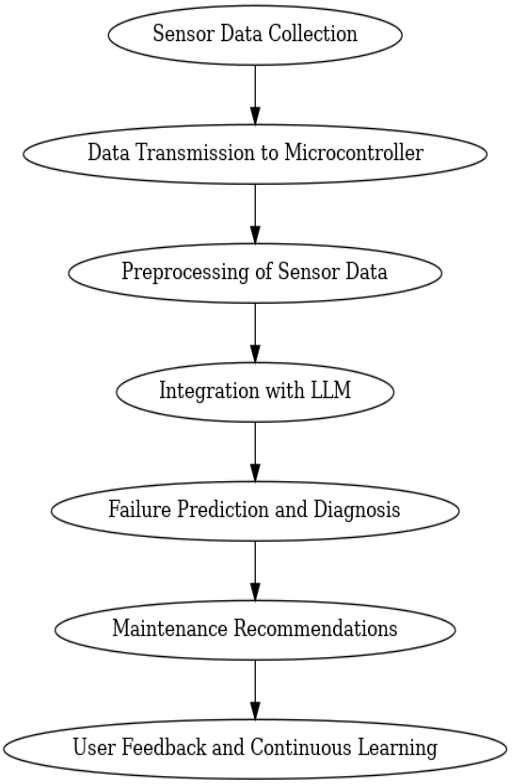

Figure 1 Flowchart of the Proposed Predictive Maintenance System

4. System Architecture

The system architecture consists of five interconnected layers, as illustrated in Fig. 1:

- **Sensor Layer:** The Raspberry Pi Pico W collects temperature, vibration, and current readings from industrial equipment.
- **Cloud Storage Layer:** Sensor data is transmitted and stored in Firebase Realtime Database.
- **Anomaly Detection Layer:** FAISS-based vector search is used to compare live sensor readings with past historical data.
- **AI Analysis Layer:** The Lang Chain-powered LLM analyzes the retrieved data and generates real-time maintenance insights.
- **Alert Mechanism Layer:** When critical anomalies are detected, notifications are sent to workers via email alerts. This multi-layered

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

architecture ensures real-time machine health monitoring and minimizes unexpected failures.

5. Implementation and Experimental Setup

5.1.Hardware Components

The system is built using the following components:

- **Raspberry Pi Pico W:** Serves as the microcontroller for collecting and transmitting sensor data.
- **DHT11 Sensor:** Measures temperature and humidity of the machinery environment.
- **ACS712 Current Sensor:** Monitors electrical consumption for anomaly detection.
- **Vibration Sensor:** Detects mechanical vibrations indicating machine wear or misalignment.

The hardware setup is as shown below in the (Figure 2)

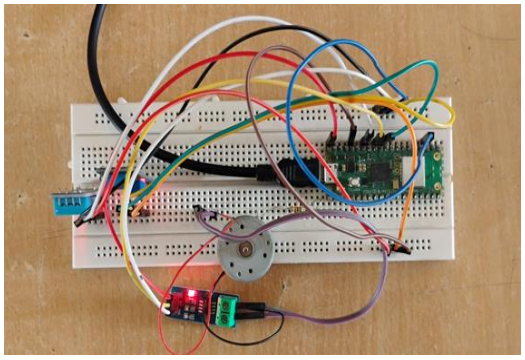

Figure 2 Hardware **Setup of the Predictive Maintenance System Using Raspberry PI Pico W and Sensors**

5.2.Software Stack

The following tools and frameworks were utilized:

- **Firebase Realtime Database:** Cloud storage for live sensor data. The sample data is as illustrates in Fig .3 **Lang Chain:** For LLM-based predictive analysis.
- FAISS (Facebook AI Similarity Search): Used to store past machine behavior patterns.
- **Flask:** Provides a lightweight server for hosting real-time monitoring results.
- **Brevo API:** Sends email alerts when anomalies are detected.

The experimental setup was tested by simulating and analyzing real-time sensor behavior. (Figure 3)

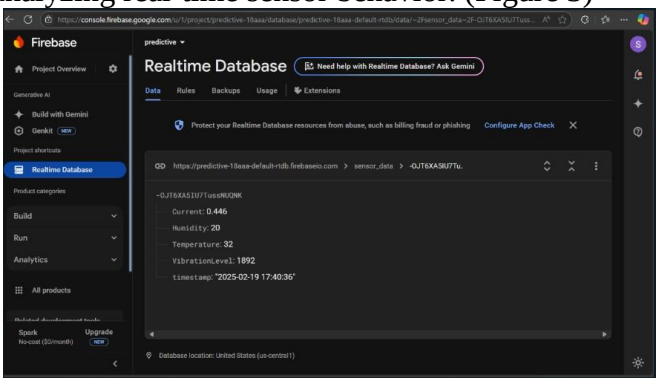

Figure 3 Sample Data Stored in Firebase

6. Comparison OF Llms for Predictive Maintance

Large Language Models (LLMs) have significantly enhanced predictive maintenance by analyzing sensor data and historical patterns to anticipate failures. In this study, we evaluated five LLMs—OpenAI GPT-4o, Mistral, DeepSeek, Gemma-3, and Llama-3—to determine their efficiency in terms of response time and accuracy. The models were tested using predefined queries related to predictive maintenance, sensor data interpretation, and fault diagnosis.

6.1.Performance Metrics and Comparison

To ensure a fair evaluation, each model was tested twice using structured test cases. The minimum response time and highest accuracy for each model were considered for comparison.

- **Response Time (s):** Measures how fast the LLM generates a response.
- **Accuracy (%):** Computed by comparing generated responses with expected outputs, based on semantic similarity and keyword match.

6.2.LLM Decision-Making Process

Each LLM utilizes different mechanisms to determine its predictions:

- **GPT-4o:** Uses deep contextual understanding and vast training data for high reasoning accuracy.
- **Mistral:** Prioritizes efficiency and low latency, making it ideal for real-time monitoring.
- **DeepSeek:** Excels in technical knowledge extraction, useful for industry-specific

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

datasets.

- **Gemma-3:** A balance between accuracy and interpretability, useful for structured failure analysis.
- **Llama-3:** Performs well in multimodal sensor data processing, ensuring comprehensive fault detection.

7. Results and Analysis

Table 1 Presents the Best Response Times and Accuracy Scores for Each Model

LLM Model	Min Response Time (s)	Max Accuracy (%)
OpenAI GPT-4o	6.03	80.0
Mistral	3.12	66.67
DeepSeek	14.93	80.0
Gemma-3	42.53	8.0
Llama-3	13.47	93.33

From the results, Llama-3 demonstrated the highest accuracy at 93.33%, while Mistral provided the fastest response time at 3.12 seconds. The choice of LLM should be guided by whether speed or accuracy is the primary requirement in predictive maintenance tasks.

8. Comparison of Llm Outputs

To analyze the performance of different Large Language Models (LLMs), the following outputs were generated. Each model processed the same input data and provided its respective response. (Figure 4,5,6,7,8)

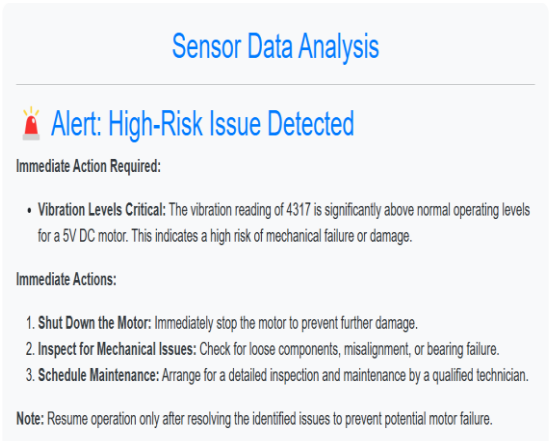

Figure 4 Output from GPT-4o model of the Open AI API

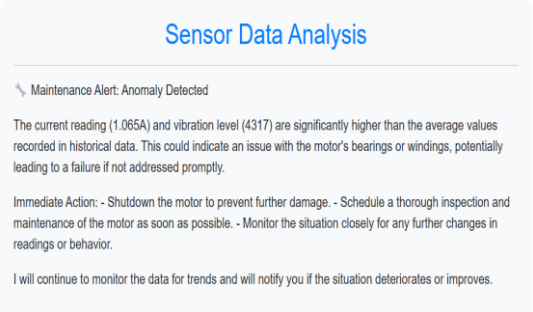

Figure 5 Output from Mistral

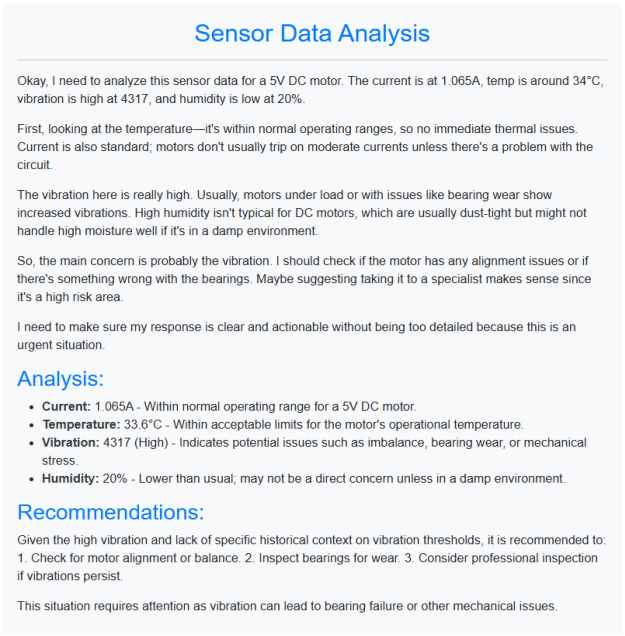

Figure 6 Output from Deepseek-r1 Based on Llama and Qwen

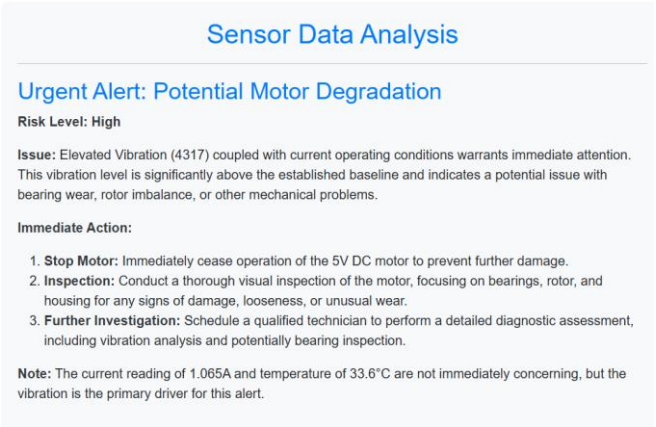

Figure 7 Output from Gemma-3 from Google Built on Gemini

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

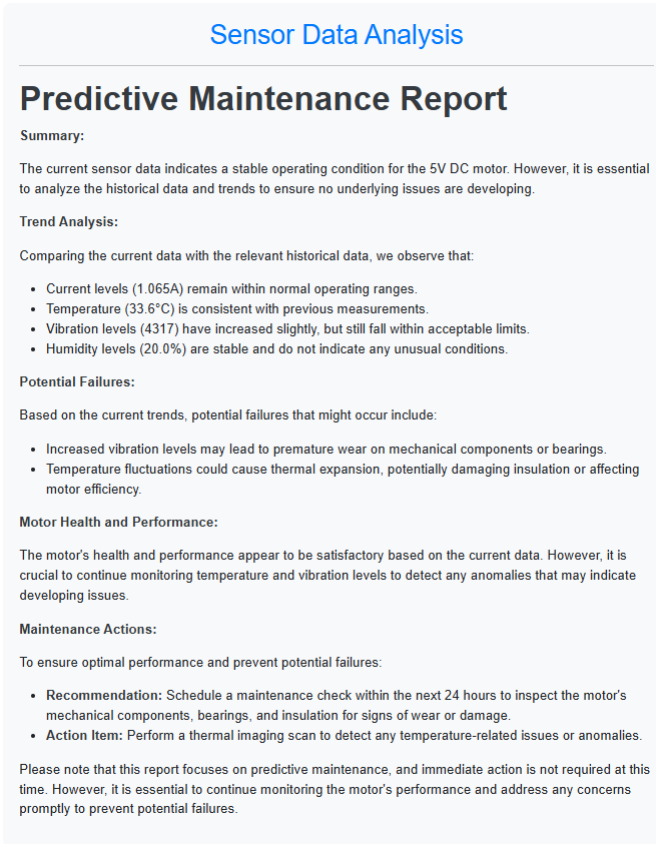

Figure 8 Output from Llama-3 from Meta

9. Result and Discussion

The proposed system was evaluated based on real-time sensor data analysis and its ability to detect machine anomalies. Real-time Sensor Data Processing

Raspberry Pi Pico W successfully transmitted data to Firebase, achieving a data update frequency of 10 minutes. The FAISS search system retrieved historical cases within 0.05 seconds, enabling real-time comparisons.

9.1.LLM-Based Predictive Maintenance

- Analysis Lang Chain's LLM-generated insights included:
- Alerts for high vibration and abnormal current levels.
- Recommendations for predictive maintenance scheduling.
- Context-based reasoning from historical failures.

9.2.Alert System Efficiency

The Brevo API successfully triggered critical alerts, notifying workers via email when anomalies were detected. Overall, the system demonstrated low-latency real-time monitoring, effectively reducing unexpected machine failures. A sample image is as shown below in the (Figure 9)

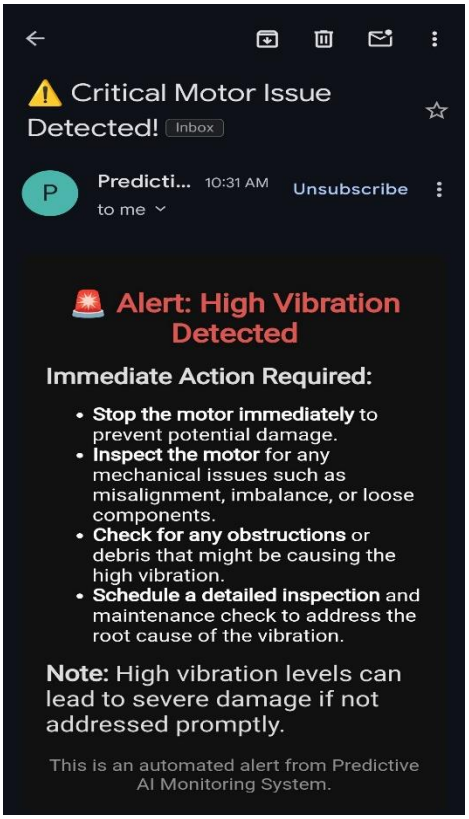

Figure 9 The Email that has Been Sent to a Mobile Phone Via Brevo API

Conclusion and Future Work

This paper presents an AI-driven predictive maintenance system integrating IoT sensors, FAISS-based anomaly detection, and Lang Chain-powered LLM analysis. The system effectively monitors industrial equipment, identifies anomalies, and generates maintenance recommendations, reducing machine downtime.

Future improvements include:

- Extending sensor integration to industrial-grade IoT de vices.
- Implementing reinforcement learning for adaptive maintenance scheduling.
- Exploring voice-based AI assistants for real-time worker interaction.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 05 May 2025
Page No: 2063-2069
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0301

- By leveraging real-time monitoring and AI-driven models, this system represents a scalable, intelligent solution for industrial predictive maintenance.

References

[1]. Mohapatra, "Generative ai for predictive maintenance: Predicting equipment failures and optimizing maintenance schedules using ai," International Journal of Scientific Research and Management (IJSRM), vol. 12, no. 11, pp. 1648–1672, 2024.

[2]. K. M. Alsaif, A. A. Albeshri, M. A. Khemakhem, and F. E. Eassa, "Multimodal large language model-based fault detection and diagnosis in the context of industry 4.0," Electronics, vol. 13, no. 4912, 2024.

[3]. Y. Liu, W. Zhang, Z. Bao, X. Chai, M. Gu, W. Jiang, Z. Zhang, Y. Tian, and F.-Y. Wang, "Brain-like cognition-driven model factory for iiot fault diagnosis by combining llms with small models," IEEE Internet of Things Journal, pp. 1–1, 2024.

[4]. [4] H. Wang and Y.-F. Li, "Large language model empowered by domain specific knowledge base for industrial equipment operation and mainte nance," IEEE Transactions on Industrial Informatics, 2024.

[5]. L. Ma, N. Thakurdesai, J. Chen, J. Xu, E. Korpeoglu, S. Kumar, and K. Achan, "Llms with user-defined prompts as generic data operators for reliable data processing," Proceedings of the 2023 IEEE International Conference on Big Data (BigData), pp. 3144–3148, Dec 2023