Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 2025, 19(2), 304–320
https://doi.org/10.12913/22998624/196264
ISSN 2299-8624, License CC-BY 4.0

Received: 2024.10.04
Accepted: 2024.12.15
Published: 2025.01.01

Application of large language models in diagnostics and maintenance of aircraft propulsion systems

Mikołaj Klekowicki[1], Grzegorz M. Szymański[1], Marek Waligórski[2*], Wojciech Misztal[2]

[1] Institute of Transport, Faculty of Civil and Transport Engineering, Poznan University of Technology, ul. Piotrowo 3, Poznań, Poland
[2] Institute of Powertrains and Aviation, Faculty of Civil and Transport Engineering, Poznan University of Technology, ul. Piotrowo 3, Poznań, Poland
* Corresponding author's email: marek.waligorski@put.poznan.pl

ABSTRACT

This study delves into the application of large language models (LLMs) in the diagnostics and maintenance of aircraft propulsion systems. Rapid advancements in aviation technology make there is an increasing need for sophisticated tools to assist in predicting and preventing equipment failures. LLMs, trained on extensive datasets, offer the potential to analyze telemetry and operational data, providing diagnostic insights and maintenance recommendations. This research explores the capabilities of LLMs in interpreting sensor data, identifying anomalies, and generating maintenance guidelines. The performance and limitations of LLMs are evaluated utilizing synthetic data from NASA to simulate real-world scenarios. Findings indicate that LLMs can enhance the reliability and efficiency of aircraft propulsion system maintenance significantly, despite challenges related to data quality and model limitations.

Keywords: large language models, aircraft propulsion systems diagnostics, artificial intelligence, predictive maintenance, telemetry data analysis.

INTRODUCTION

The aviation industry has experienced significant advancements, leading to highly reliable aircraft designs. Alongside the technological evolution, there has been a substantial growth in the legal and procedural frameworks governing the flight, diagnostics, and maintenance of an aviation machinery. Since World War II, the rapid pace of a technological advancement has often rendered new technologies obsolete shortly after the deployment, posing challenges in keeping regulatory and procedural frameworks up-to-date with innovations.

The human factor remains constant in this evolving field. Human adaptation to technological environments occurs over generations considerably slower than technological advancements, leading to focus on human error as the primary cause of aviation accidents. In response, artificial intelligence (AI) has been explored to support and enhance human decision-making processes in aviation. Advanced machine learning models like LLMs hold significant potential in the diagnostics and maintenance of aircraft propulsion systems. LLMs can contribute to increased safety and operational efficiency in aviation by analyzing vast datasets, including telemetry data and flight reports. LLMs are sophisticated AI tools designed to process and generate a human-like text based on their training data [1]. Although their application in technical fields like aviation is emerging, LLMs have been extensively used in natural language processing tasks [2]. The primary advantage of LLMs lies in their ability to process large volumes of data and extract meaningful patterns and insights that can aid in the predictive maintenance and diagnostics [3].

This research investigates the role of LLMs in aviation, particularly their potential in predicting and preventing failures and optimizing maintenance processes for aircraft propulsion systems. Utilizing synthetic data provided by NASA to simulate realistic scenarios, albeit simplified compared to the full complexity of actual operational data, this study aims to demonstrate the potential and capabilities of LLMs in the aviation diagnostics and maintenance.

REVIEW OF SCIENTIFIC ACHIEVEMENTS IN THE RESEARCH TOPIC AREA

The advancement of LLMs has opened new possibilities in various domains [4–7], including the aviation diagnostics and maintenance. Models like GPT-3.5 [8], GPT-4 [9], Falcon-180B, and Llama 2-7b, trained on extensive text corpora, enable understanding and generating human-like language responses. These models are sophisticated AI tools designed to mimic human language processing capabilities, trained on vast datasets to comprehend context and generate coherent responses. In aviation, LLMs can analyze large sets of operational and diagnostic data to predict and prevent equipment failures, enhancing the safety and efficiency. These models can identify patterns and anomalies indicating potential issues, by leveraging telemetry data and flight reports [10, 11].

Data conversion into a vector format, known as embedding (to be compatible with LLMs), is a significant challenge in utilizing LLMs for aviation diagnostics. This involves transforming raw text data from various formats into numerical vectors that LLMs can process. Previous research has demonstrated the potential of LLMs in various technical fields [12–14], interpreting the sensor data, and providing diagnostic insights without an extensive retraining. For instance, LLMs have been used in the manufacturing to predict equipment failures based on the sensor data and in healthcare to analyze patient records and suggest potential diagnoses [15, 16].

The effectiveness of LLMs heavily depends on the quality and consistency of the input data [6, 17–19]. The high-quality and well-structured data enable more accurate and reliable predictions, while the poor-quality or inconsistent data can lead to erroneous predictions, reducing an overall model effectiveness. Ensuring data quality is critical for implementing LLMs in technical fields, including aviation.

OUTLINING THE RESEARCH PROBLEM

The aviation industry constantly strives to improve the safety and operational efficiency by focusing on the diagnostics and maintenance of aircraft propulsion systems. The development of composite materials used in modern aviation enhances the safety and reliability of aircraft. Thera are plenty of scientific papers including about practical usage of composite materials [20]. Current methodologies for diagnosing and maintaining these systems are often labour-intensive and require a significant human expertise. Despite technological advances, the human factor remains important in the maintenance and operational processes, introducing challenges resulting primarily from potential human error and limitations in processing large volumes of complex data.

Traditional methods of the aircraft propulsion system maintenance rely heavily on scheduled inspections and a reactive maintenance. Scheduled inspections, based on predefined intervals, may not accurately reflect the actual condition of components, leading to unnecessary maintenance actions or delayed interventions, potentially threatening safety. The reactive maintenance addresses issues only after they occur, resulting in unscheduled downtimes and higher costs due to emergency repairs and potential secondary damages. The human error remains a leading cause of maintenance-related incidents in aviation.

The significant challenge in a modern aviation diagnostics is effectively handling and analyzing vast amounts of data generated by aircraft systems. Traditional data processing methods are often insufficient for extracting meaningful insights from this data. Key challenges include the data volume, the data variety and data quality. Efficient processing and analyzing large volumes of data in real time goes beyond the capabilities of traditional methods. Aircraft systems produce diverse data types, including sensor readings, maintenance logs and operational reports, making the integration of these disparate data types into a cohesive analysis framework complex. The accuracy and reliability [21, 22] of insights derived from data analytics heavily depend on the input data quality. The inconsistent or poor-quality data can lead to incorrect diagnostics and maintenance recommendations [23–28].

LLMs offer a promising solution to these challenges. Their ability to process and analyze the large volumes of text data can be leveraged to enhance diagnostics and maintenance processes in aviation. LLMs can aid in a predictive maintenance by analyzing historical data and identifying patterns to predict potential failures before they occur, allowing proactive maintenance actions. They can enhance the diagnostics by comparing current data with historical data and known failure patterns, improving the accuracy and consistency of a diagnostics. Additionally, LLMs can provide the maintenance personnel with insights and recommendations based on a comprehensive data analysis, supporting better decision-making and reducing the reliance only on a human judgment [29].

Despite these potentials, challenges remain in implementing LLMs [30–32], including the data quality assurance, integrating LLMs with existing systems and managing the computational resources required for the LLM processing. Addressing these challenges can lead to more efficient, reliable and safer aviation operations [33, 34].

THEORETICAL BASES IN THE SCIENTIFIC TOPIC

In this section, we will delve into the theoretical foundations of Large Language Models (LLMs), the specific models chosen for this study, and the basics of technical diagnostics and operation in aviation. This section is crucial for understanding how LLMs can be applied effectively to improve the maintenance and diagnostics of aircraft propulsion systems.

Basics of large language models

LLMs are advanced machine learning models designed to process and generate a human-like text based on vast amounts of training data [35, 36]. These models have demonstrated remarkable capabilities in various natural language processing (NLP) tasks, such as the text generation, translation, summarization and question-answering [37, 38]. LLMs like GPT-3.5 and GPT-4 are based on the transformer architecture, which allows them for understanding the context and generating coherent responses. These models are trained on extensive datasets containing diverse textual information, enabling them to capture nuanced language patterns

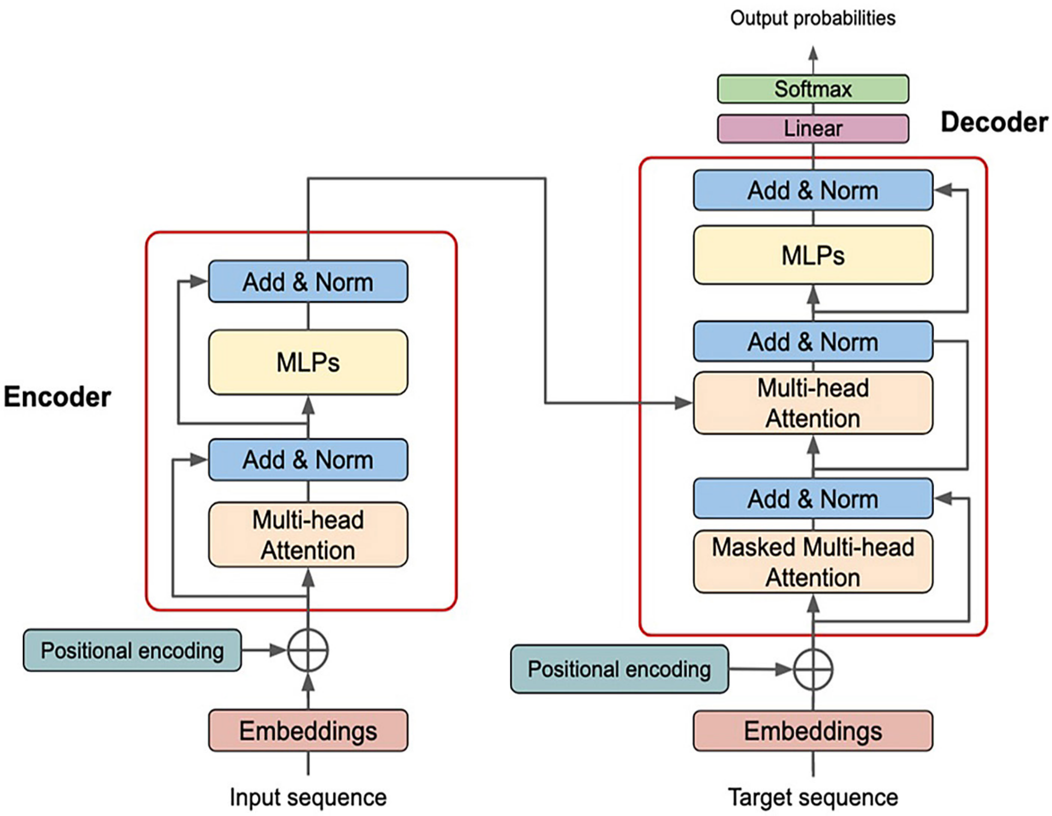

Figure 1. Transformer architecture overview

and contextual meanings (Figure 1). The transformer architecture, introduced by [39], consists of an encoder-decoder structure where the encoder processes the input sequence and the decoder generates the output sequence. The key innovation of this architecture is the self-attention mechanism, which allows the model to weigh the importance of different words in a sentence when a response is generated.

Multidimensional vector spaces

LLMs rely heavily on the concept of embeddings, which transform words or phrases into numerical vectors in a high-dimensional space. These embeddings are critical for the model's ability to understand and generate the human language. Each dimension in this vector space represents a latent feature that captures some aspect of the word's meaning or a usage context.

For example, in a 300-dimensional vector space, the word "aircraft" might be represented by a vector where each dimension corresponds to a specific characteristic of the word, such as its relation to "flight", "mechanics" or "safety".

The process of creating these embeddings is complex and involves several steps:

- Tokenization: Breaking down the text into smaller units (tokens) such as words or subordinate words.
- Vector representation: Assigning a high-dimensional vector to each token based on its context and usage in the training data.
- Contextual adjustment: Using mechanisms like the attention to adjust these vectors based on surrounding words, thereby capturing the context-specific meaning of each token.

The visualization above shows how different words can be mapped in a high-dimensional vector space, where semantically similar words are located closer to each other (Figure 2) [40].

Chosen models

For this study, we selected GPT-4 by OpenAI due to its advanced capabilities and ease of implementation. GPT-4 supports multiple languages, including Polish, making it suitable for analyzing the data in diverse linguistic contexts.

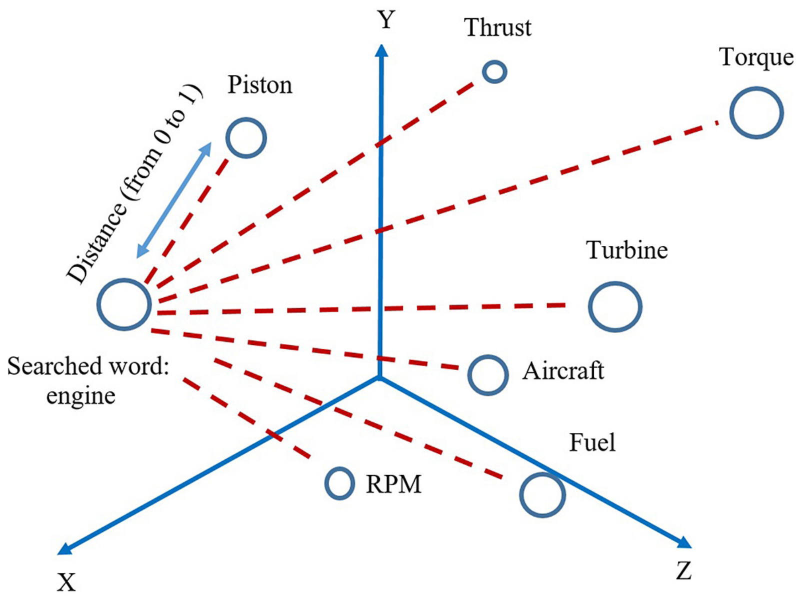

Figure 2. Embedding space visualization

Table 1. Parameters for GPT-4 model configuration

Parameter	Description
Temperature	Controls the randomness of predictions
Maximum tokens	Sets the limit on the response length
Top-p	Nucleus sampling parameter
Repetition penalty	Prevents repetitive text generation

Additionally, GPT-4 provides APIs that facilitate the seamless integration with existing systems.

Example code: Integrating GPT-4 in Python

This code snippet demonstrates how to interact with the GPT-4 model using the OpenAI API. The messages parameter allows you to set the context and user query, and the model responds based on its training data (Figure 3) [41, 42].

BASICS OF TECHNICAL DIAGNOSTICS AND EXPLOITATION

The technical diagnostics in aviation is a multidisciplinary field focused on assessing the condition of various systems and components of an aircraft. The primary goal is to ensure the safe and efficient operation of an aircraft by identifying potential issues before they lead to failures. This section explores the fundamental concepts, methods and standards relevant to the technical diagnostics and operation in aviation [43–48].

Key concepts in technical diagnostics

Technical diagnostics: this involves identifying the current, past and future states of technical systems. The primary objective is to determine the present condition of technical devices to predict their behaviour and performance. The main concepts include:

- Condition: the current state of a system, influenced by its history and necessary for predicting its future behaviour.
- Diagnostic signal: a signal representing changes over time in a specific physical quantity, providing information about the system's condition.
- Diagnostic model: a model used to interpret diagnostic signals and assess the system's condition.
- Diagnostic procedure: the process of using diagnostic models and signals to determine the system's condition.
- Diagnosis: the result of the diagnostic procedure, indicating the current state of the system.

Key characteristics of diagnostic signals

Sensitivity: the ratio of a change in the diagnostic parameter ($\Delta y_n(u)$) to the change in the state parameter ($\Delta x_m(u)$) [49, 50].

$$K = \frac{\Delta y_n(u)}{\Delta x_m(u)} \quad (1)$$

Uniqueness: each state parameter value corresponds to a single diagnostic parameter value.

Stability: the diagnostic parameter should remain stable under consistent diagnostic conditions. Ease of a measurement: the diagnostic signal should be measurable without a significant effort, such as disassembling the system or requiring specialized tools.

Exploitation, as defined by the Polish standard PN-82/N-04001 [51], encompasses the set of organizational, technical, and economic actions taken with a technical object from its acceptance for use until its disposal. Various strategies are employed

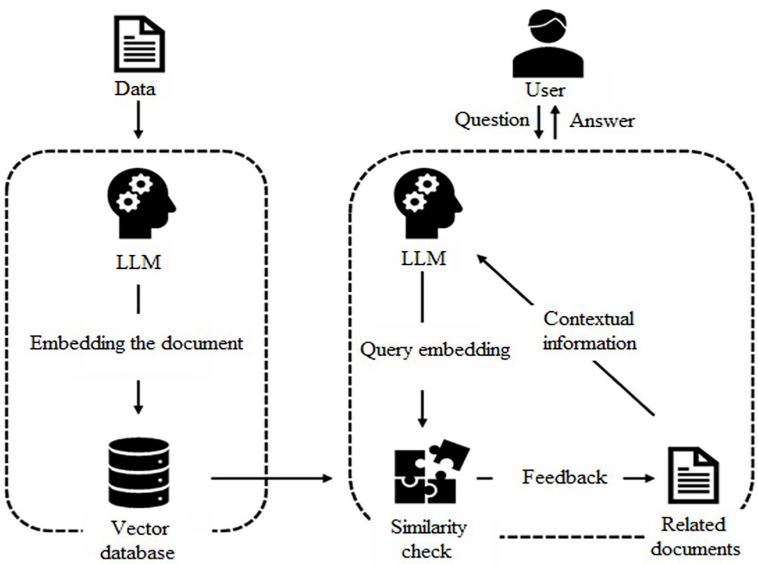

Figure 3. Data flow visualization

to manage the exploitation of technical objects, each offering distinct advantages and challenges. The Run-to-Break Strategy is a common strategy, which involves operating an object until it fails. While this approach is simple, it can result in extended downtimes and high repair costs. The Preventive Maintenance is another approach, which entails regular inspections and maintenance to preserve the object's functionality and to minimize downtime. The condition-based maintenance (CBM) involves monitoring the object's current condition and performing a maintenance based on the real-time data, optimizing maintenance efforts but requiring the detailed knowledge of the object and sophisticated monitoring tools. Additional exploitation strategies include:

- exploitation based on resource (PE), which involves using the object within its designed lifespan and capabilities,
- technical state-based exploitation, where decisions are made based on the current technical state of the object,
- mixed exploitation strategy, which combines various approaches to suit specific needs.
- economic efficiency-based exploitation, focusing on minimizing costs while maximizing benefits.
- reliability-centered maintenance (RCM), a modern strategy with an emphasis on the comprehensive risk analysis and a reliability.

In the field of a technical diagnostics, data acquisition methods are crucial and can be categorized into destructive and non-destructive testing. Destructive Testing involves methods that lead to the damage or destruction of the object, such as tensile tests, compression tests and impact tests. Non-Destructive Testing (NDT) methods assess the object's condition without causing damage and include techniques such as the: ultrasonics, radiographic testing, magnetic particle testing, visual inspections, thermography, dye penetrant inspection, eddy current testing, acoustic emission testing and oil analysis.

Aircraft propulsion systems can be categorized based on how they generate thrust: the direct thrust generation, as seen in jet engines like turbojets and ramjets, and the indirect thrust generation, as in propeller and rotorcraft engines, including piston and turboprop engines. Propulsion systems are subjected to various types of loads, including thermal loads, which can cause the deformation, thermal fatigue and oil degradation, and mechanical loads, which involve forces acting on engine components during operation, leading to the material fatigue, the bearing wear and the shaft damage.

Effective diagnostics and maintenance strategies are essential for ensuring the reliability and the safety of both direct and indirect thrust generation systems. Techniques such as the vibration analysis, the thermography and the oil condition monitoring help identify potential issues early and prevent catastrophic failures. By understanding and applying these theoretical foundations, one can enhance the reliability, the efficiency and the safety of aircraft propulsion systems through the advanced diagnostics and exploitation strategies [47, 52–54]. The flowchart above outlines the predictive maintenance process, highlighting the steps from the data acquisition to the maintenance decision-making (Figure 4).

MATHEMATICAL MODELS AND TECHNIQUES

In this subsection, the authors of this scientific paper explore the mathematical models and techniques employed to analyze data and derive meaningful insights during the study. The application of mathematical principles is crucial for processing diagnostic signals, predicting failures and optimizing maintenance schedules.

Standard Deviation and 2-Sigma Rule: the standard deviation (σ) and the 2-sigma rule ($\pm 2\sigma$) is one of the key statistical techniques used in this

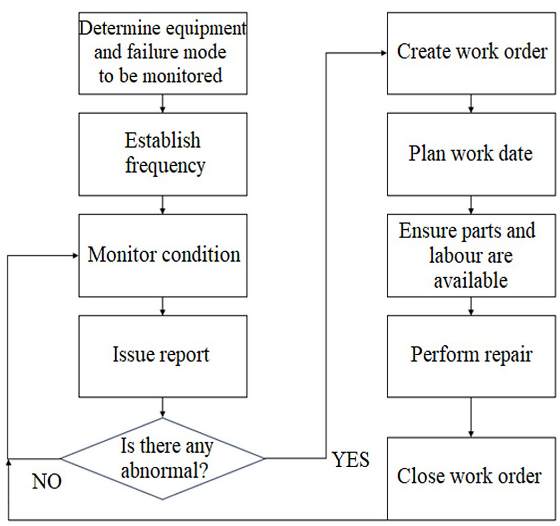

Figure 4. Flowchart of the predictive maintenance process

Table 2. Common diagnostic parameters for aircraft engines

Parameter	Unit	Description
T_{02}	°R	Stagnation temperature at a compressor inlet
T_{03}	°R	Stagnation temperature at a compressor outlet
P_{02}	MPa	Stagnation pressure at a compressor inlet
P_{03}	MPa	Stagnation pressure at a compressor outlet
RMS	m/s²	Root Mean Square of a compressor bearing vibration

study. This method helps to identify anomalies and to assess the reliability of the diagnostic data.

The standard deviation is calculated as:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2} \tag{2}$$

where: N is the number of observations, x_i represents each observation, μ is the mean of the observations.

Using the 2-sigma rule, data points that fall within ±2σ from the mean (μ) are considered to be within the normal range. This rule helps in the detection of outliers or anomalies, which could indicate potential issues in the system.

Root mean square (RMS) calculation: RMS is a statistical measure used to quantify the magnitude of varying quantities, often applied in the vibration analysis of mechanical systems. For a set of values x_1,x_2,...,x_n, the RMS is given by:

$$'S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2} \tag{3}$$

This formula helps to assess the overall energy of vibrations, which can indicate the technical state of components like bearings and shafts.

Temperature and pressure ratios: In thermodynamics, evaluating the efficiency of a compressor or a turbine involves analyzing the temperature and pressure ratios. The isentropic efficiency η_c of a compressor is calculated as:

$$\eta_C = \frac{\left(\frac{T_{03}}{T_{02}}\right)-1}{\left(\frac{P_{03}}{P_{02}}\right)^{\frac{\gamma-1}{\gamma}}-1} \tag{4}$$

where: T_{02} and T_{03} are the stagnation temperatures at the compressor inlet and outlet respectively, P_{02} and P_{03} are the stagnation pressures at the compressor inlet and outlet respectively, γ is the specific heat ratio.

This formula is essential for diagnosing the performance and efficiency of propulsion systems under different operating conditions (Figure 5) [24, 55, 56].

EXPERIMENT DESIGN AND EXECUTION

In this chapter, the authors of this paper discuss the design and execution of the experiment, including the methodologies used, the data

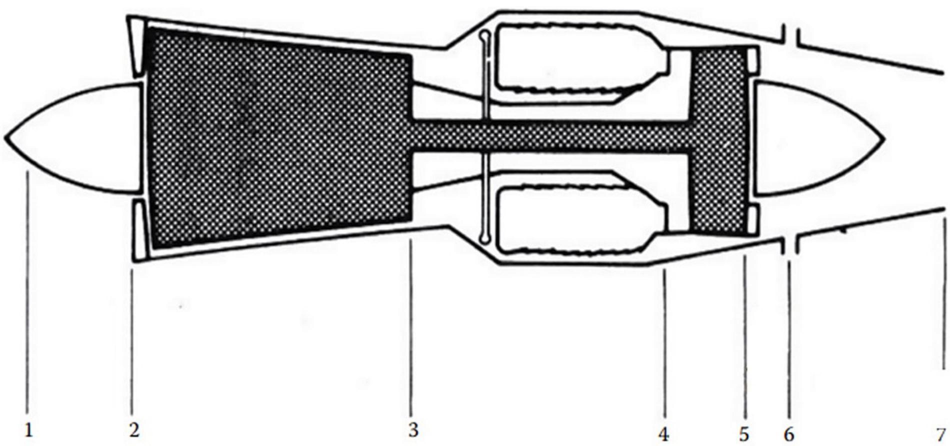

Figure 5. Diagram of the examined engine: 1–2 – inlet, 2–3 – compressor, 3–4 – combustion chamber, 4–5 – turbine, 5–6 – exhaust nozzle, 6–7 – exhaust

collection process and the analysis techniques applied. This structured approach ensures the reliability and validity of the research findings.

The experiment was designed to evaluate the effectiveness of LLMs in diagnosing and predicting maintenance needs for aircraft propulsion systems. The primary objective of the experiment was to determine whether LLMs can accurately identify anomalies and predict maintenance needs based on the historical and real-time data from aircraft propulsion systems. The tested hypothesis was that LLMs, when trained on large datasets of the operational data, can achieve the high accuracy in diagnosing faults and predicting maintenance requirements. Key variables considered in the experiment included types of the data input (e.g. the temperature, the pressure, the vibration data), the training dataset size, and LLM configurations (independent variables), and the accuracy of a fault diagnosis, the precision of maintenance predictions, and model performance metrics (dependent variables).

The data collection was a critical component of the experiment, ensuring that the models had the sufficient and relevant information to learn from and make accurate predictions. The data sources included the historical maintenance data, the real-time sensor data, and the synthetic data. The historical maintenance data, collected from aircraft maintenance logs, provided a record of previous faults, maintenance actions taken and the outcomes, offering a historical perspective essential for training the LLMs to recognize patterns associated with various faults. The real-time sensor data from an aircraft, including the temperature, pressure, and vibration readings, were streamed and recorded, crucial for testing the models' ability to make real-time predictions and diagnose ongoing issues. Additionally, the synthetic data generated by NASA was used to simulate various operational scenarios, augmenting the training dataset and providing more diverse examples for the models to learn from.

The experiment was executed in several phases to test the hypothesis and to evaluate the performance of the LLMs systematically. The first phase involved data pre-processing to ensure the consistency and quality. This included cleaning the data, handling missing values, normalizing the variables, and converting the raw data into suitable formats for model training. Following pre-processing, the LLMs were trained using the pre-processed data. The training process involved feeding the models with input data, adjusting their parameters and optimizing their performance through iterative learning. Hyper parameters such as the learning rate, the batch size, and the number of epochs were fine-tuned to achieve optimal results.

The data was split into training, validation and test sets to evaluate the models' performance. The validation set was used to tune the models during training, while the test set was used for the final evaluation. Performance metrics such as the accuracy, the precision, the recall, and the F1-score were calculated to assess the models' effectiveness. The trained models were then used to detect anomalies in the real-time sensor data and predict maintenance needs. These predictions were compared against actual maintenance records to determine the accuracy and reliability of the models.

Several analysis techniques were employed to interpret the results and draw meaningful conclusions from the experiment. Statistical methods, including descriptive statistics and hypothesis testing, were used to analyze the data distributions and test the significance of the results. This helped to understand the underlying patterns and to validate the experiment's findings. Data visualization tools were utilized to represent the data and model predictions graphically. Visualizations such as histograms, scatter plots and time-series graphs provided insights into the data trends and the model performance. Additionally, an error analysis was conducted to identify the types and sources of errors in the models' predictions. This involved examining false positives, false negatives and other misclassifications to understand the limitations and potential areas for the improvement.

The aim of the research was to provide the robust and reliable evidence on the efficacy of LLMs in the aircraft propulsion system diagnostics and maintenance through a meticulous design and execution of the experiment. The insights gained from this study can inform future developments and applications of AI in the aviation maintenance [57–59].

Experiment procedure

Step 1: Data import – The experiment begins with importing and verifying the consistency of the data provided by the user. These data include various formats such as Excel, PDF, Word, and TXT.

Step 2: Data analysis – using tools like Python, Pandas, and NumPy, the data are analyzed for consistency and readiness for further processing. This process includes segregating the data according to defined experimental scenarios.

Step 3: Creating characteristics with Matplotlib, characteristics for each scenario are created. These characteristics illustrate the relationships between various operational parameters, considering the boundaries for states "Operationally Fit", "Fit but Not Operational" and "Not Fit and Not Operational".

Step 4: Creating and analyzing correlations – the next step involves creating correlations between parameters, excluding cycles and time. These correlations aim to identify patterns and dependencies that may be crucial for assessing the technical condition of the propulsion systems.

Step 5: Technical condition assessment – in this step, the highest parameter values from each scenario are compared to the established threshold values for the states "Fit but Not Operational" and "Not Fit and Not Operational". If any parameter exceeds the threshold for "Fit but Not Operational" and/or "Not Fit and Not Operational", the entire object is automatically classified according to the highest exceeded threshold. This analysis allows for assessing the overall technical condition of the propulsion systems.

Step 6: Verdict – a verdict regarding the technical condition of the propulsion systems is formulated based on the collected data and conducted analyses. This verdict includes the classification of each parameter and the overall assessment of the propulsion system's condition.

Step 7: Conclusions – conclusions from the analysis are formulated finally. These conclusions summarize the key observations and recommendations that can be used for further engineering actions and operational decisions.

By following this detailed procedure, the experiment provides the comprehensive and systematic approach to assessing the effectiveness of LLMs in diagnosing and predicting maintenance needs for aircraft propulsion systems. The results provide valuable insights and practical recommendations for improving maintenance strategies in aviation.

Datasets and verification process

In this chapter, the authors detail the datasets used in the experiment and the process of verifying their consistency and suitability for the analysis. Ensuring the quality and reliability of data is critical for the validity of the experiment's outcomes. The datasets used in this experiment were diverse and comprehensive, encompassing the historical maintenance data, the real-time sensor data and the synthetic data. These datasets provided a rich foundation for training and testing the LLMs. The historical maintenance data were sourced from aircraft maintenance logs. These logs included records of previous faults, maintenance actions taken and the outcomes of those actions. The historical data provided a valuable context for the LLMs to learn from past incidents and to recognize patterns associated with various faults.

The real-time data were collected from various sensors installed on an aircraft. These sensors measured critical parameters such as the temperature, pressure and vibration levels. The real-time data were essential for testing the models' ability to make immediate predictions and to diagnose ongoing issues accurately.

The synthetic data generated by NASA were used to complement the real-world data. These data simulated a wide range of operational scenarios, helping to broaden the training dataset and to introduce more variability. The synthetic data were particularly useful in creating scenarios that were not encountered in the historical data frequently.

The verification process was critical to ensure that the data used in the experiment were consistent, accurate and suitable for the analysis. This process involved several steps to validate the integrity and readiness of the data for further processing. The experiment began with importing and verifying the consistency of the data provided by the user. These data included various formats

Table 3. Data summary

Data Type	Source	Description
Historical maintenance	Maintenance logs	Records of past faults, maintenance actions and outcomes
Real-time sensor data	Aircraft sensors	Measurements of the temperature, pressure and vibration levels collected in real-time
Synthetic data	NASA simulations	Simulated operational scenarios to introduce variability and broaden the training dataset

such as Excel, PDF, Word, and TXT. The verification process involved checking for the completeness, removing duplicates and ensuring that all necessary parameters were present. Inconsistencies and missing values were addressed using appropriate data cleaning techniques.

The data were analyzed for consistency and readiness for further processing using tools like Python, Pandas and NumPy. This process included segregating the data according to defined experimental scenarios. Each scenario was carefully constructed to represent different operational conditions and potential faults.

Characteristics for each scenario were created with Matplotlib. These characteristics illustrated the relationships between various operational parameters, considering the boundaries for states "Operationally Fit", "Fit but Not Operational", and "Not Fit and Not Operational". This step was essential for visualizing the data and identifying any patterns or anomalies.

The next step involved creating correlations between parameters, excluding cycles and time. These correlations aimed to identify patterns and dependencies that might be crucial for assessing the technical condition of the propulsion systems. By analyzing these correlations, the study aimed to uncover insights that could improve the diagnostic accuracy.

In assessing the technical condition, the highest parameter values from each scenario were compared to the established threshold values for the states "Fit but Not Operational" and "Not Fit and Not Operational". If any parameter exceeded the threshold for "Fit but Not Operational" and/or "Not Fit and Not Operational", the entire object was automatically classified according to the highest exceeded threshold. This analysis allowed for assessing the overall technical condition of the propulsion systems.

A verdict regarding the technical condition of the propulsion systems was formulated based on the collected data and conducted analyses. This verdict included the classification of each parameter and the overall assessment of the propulsion system's condition. Finally, conclusions from the analysis were formulated, summarizing the key observations and recommendations that could be used for further engineering actions and operational decisions.

By following this detailed procedure, the experiment ensured a comprehensive and systematic approach to evaluate the effectiveness of LLMs in diagnosing and predicting maintenance needs for aircraft propulsion systems. The results provided valuable insights and practical recommendations for improving maintenance strategies in aviation.

RESULTS AND DISCUSSION

After applying the LLM model to the provided data, the model presented the following results: parameters Measured – Compressor #3.xlsx After importing, the files will undergo consistency checks and then the analysis. The data has been successfully imported. The files have been analyzed correctly – based on them, I am ready to generate the necessary characteristics:

Table 4. Parameters table

Parameter	Unit	Unfit	Faulty	Parameter description
T_{02}	°R	540.0	648.0	Temperature at a compressor inlet
T_{03}	°R	720.0	864.0	Temperature at a compressor outlet
P_{02}	MPa	0.860	1.20	Pressure at a compressor inlet
P_{03}	MPa	2.236	3.12	Pressure at a compressor outlet
RMS	m/s²	20.0	22.0	RMS of a compressor bearing

Table 5. Cycles data

Cycles	T_{02}	T_{03}	P_{02}	P_{03}	RMS	v	Compression
1	493.110	732.381581	0.50	1.996400	19.992801	270.000000	3.992801
2	502.972200	747.119358	0.49	1.957299	19.994487	270.000000	3.994487
3	493.110	728.935759	0.52	2.042266	19.927435	267.975000	3.927435
4	478.316700	709.108968	0.51	2.023304	19.967263	269.984813	3.967263
5	488.326833	730.195094	0.48	1.962413	20.088360	271.334737	4.088360

- Parameter T_{02} vs. Cycles
- Parameter T_{03} vs. Cycles
- Parameter P_{02} vs. Cycles
- Parameter P_{03} vs. Cycles
- Parameter RMS vs. Cycles
- Parameter v vs. Cycles
- Parameter Compression vs. Cycles

Explanation

- Parameters table:
  - T_{02} and T_{03} represent the temperatures at the inlet and outlet of the compressor, respectively, measured in degrees Rankine (°R).
  - P_{02} and P_{03} represent the pressures at the inlet and outlet of the compressor, respectively, measured in Megapascals (MPa).
  - RMS is the root mean square value of the vibrations of the compressor bearing, measured in meters per second squared (m/s^2) (Table 4).
- Cycles data:
  - Each row represents a measurement cycle.
  - For each cycle, parameters such as the temperature (T_{02}, T_{03}), the pressure (P_{02}, P_{03}), RMS, the velocity (v), and the compression ratio (Compression) are recorded (Table 5).

Process

- Data import and consistency check: the data is imported and checked for the consistency.
- Data analysis: after successful import and consistency checks, the data is analyzed.
- Characteristics Generation: Based on the analyzed data, the following characteristics will be generated:
  - T_{02} vs. Cycles
  - T_{03} vs. Cycles
  - P_{02} vs. Cycles
  - P_{03} vs. Cycles
  - RMS vs. Cycles
  - v vs. Cycles
  - Compression vs. Cycles (Figures 6–13)

Proceeding to create the characteristics.

Comparison of parameter values to reference values:

- T_{02} > Operational
- T_{03} > Unfit
- P_{02} > Operational
- P_{03} > Faulty
- RMS > Faulty
- v > Operational
- Compression > Faulty

The object is considered Faulty because the parameters T_{03}, P_{03}, RMS, and Compression exceed the values for the Faulty state.

Observed anomalies:

No anomalies observed. Proceeding to formulate conclusions.

Conclusion 1: High values for the parameters T_{03}, P_{03}, RMS, and Compression indicate serious issues with the compressor's operation, requiring the immediate intervention.

Conclusion 2: A strong correlation between the temperature parameters (T_{02} and T_{03}) and pressure parameters (P_{02} and P_{03}) may indicate that

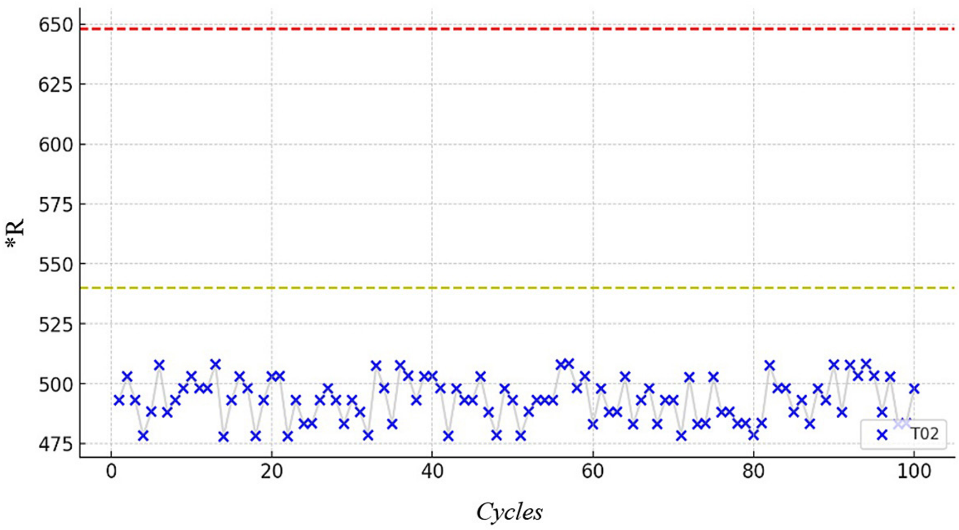

Figure 6. Characteristic of T_{02} vs. Cycles generated by the LLM model

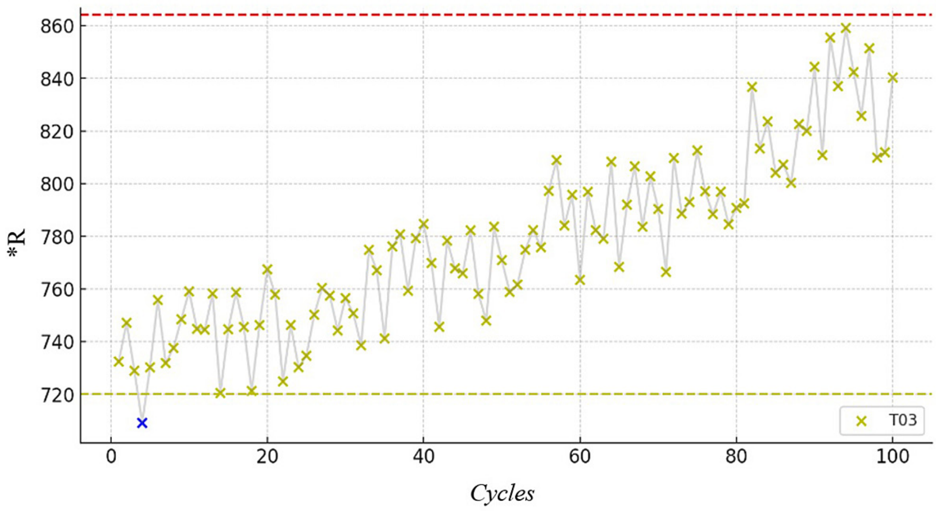

Figure 7. Characteristic of T_{03} vs. Cycles generated by the LLM model

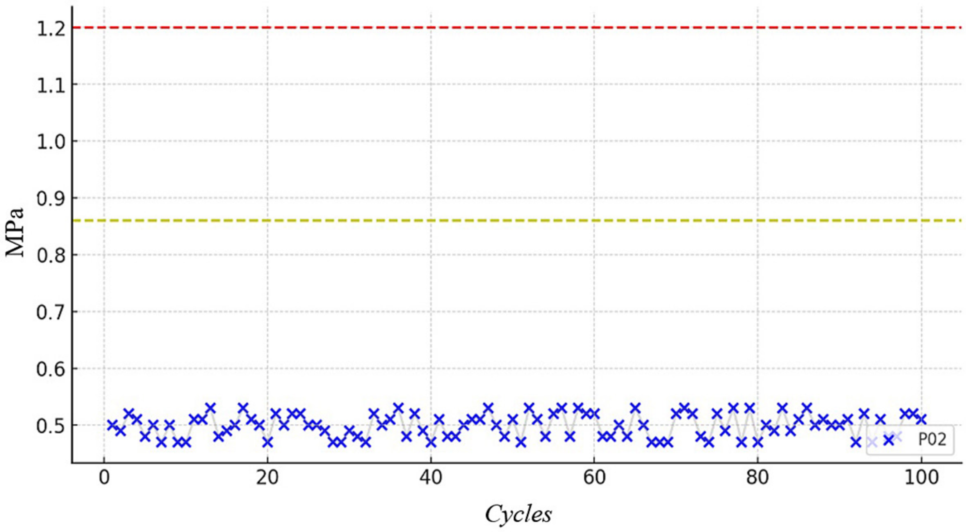

Figure 8. Characteristic of P_{02} vs. Cycles generated by the LLM model

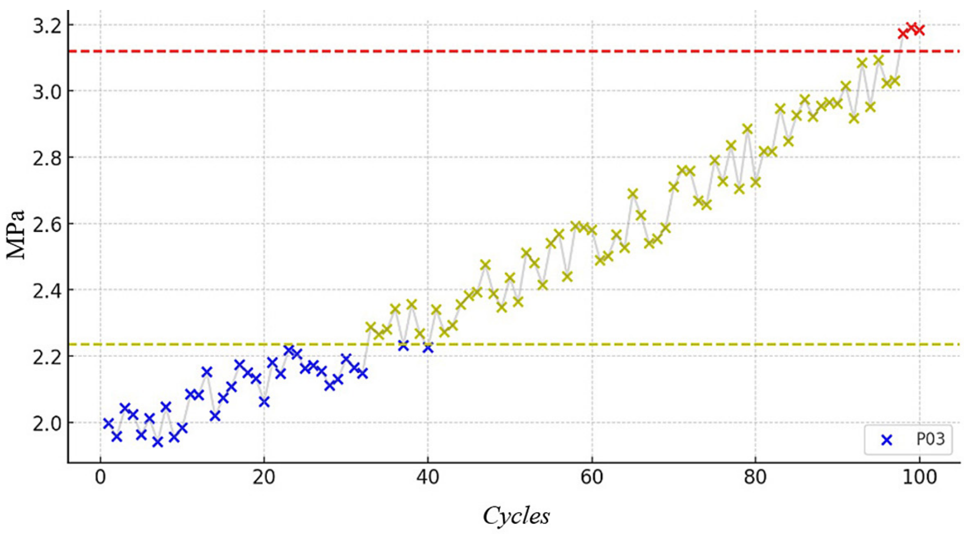

Figure 9. Characteristic of P_{03} vs. Cycles generated by the LLM model

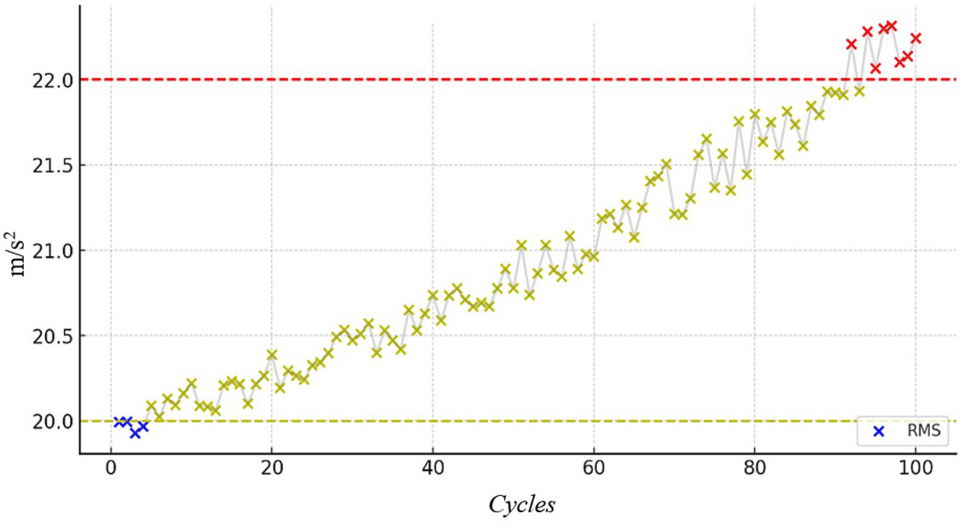

Figure 10. Characteristic of RMS vs. Cycles generated by the LLM model

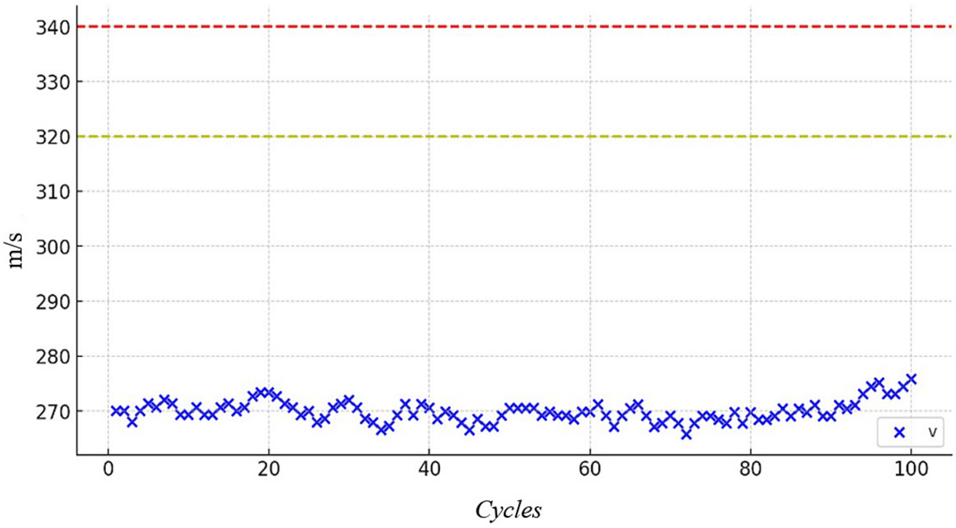

Figure 11. Characteristic of v vs. Cycles generated by the LLM model

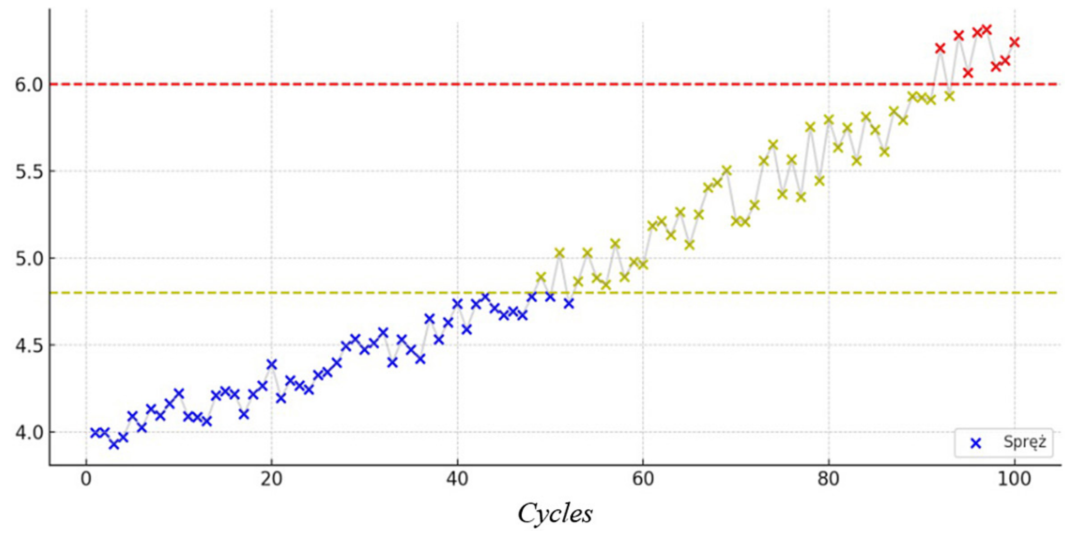

Figure 12. Characteristic of Compression vs. Cycles generated by the LLM model

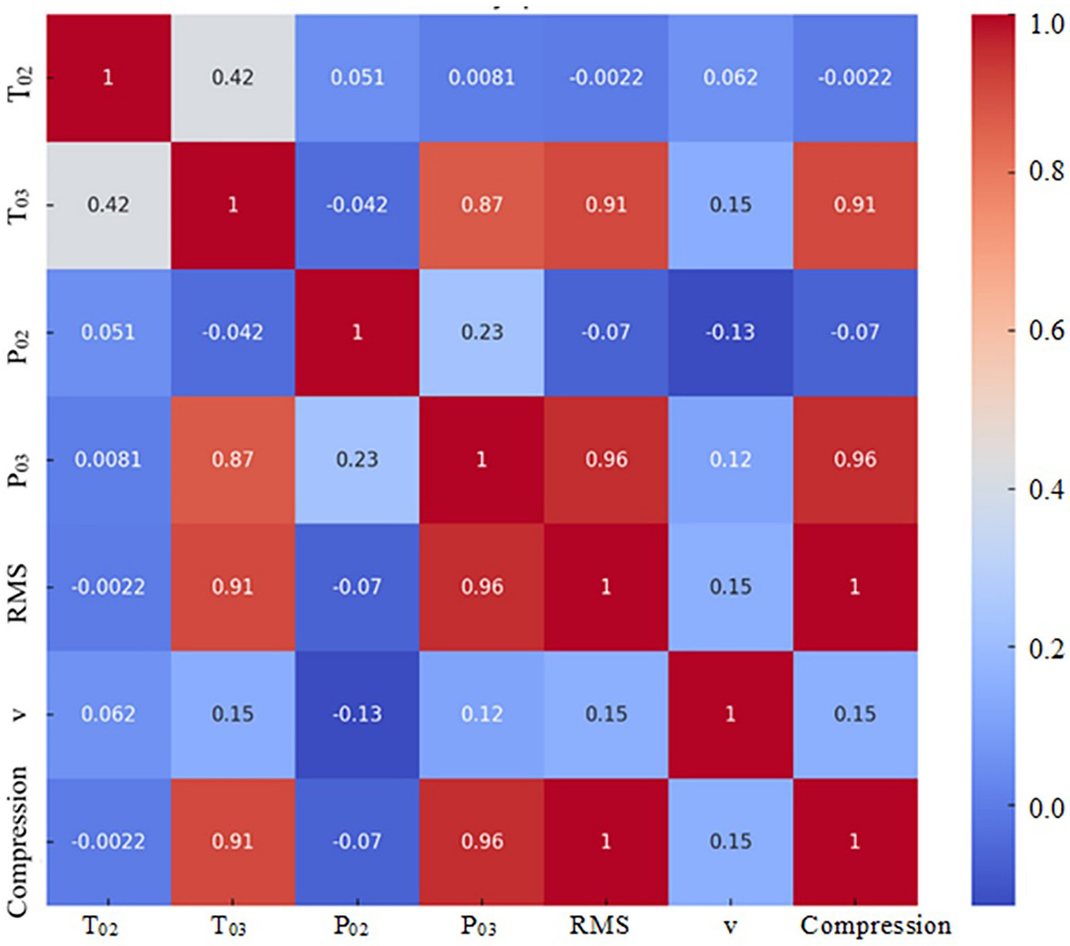

Figure 13. Characteristic of T_{02} vs. Cycles generated by the LLM model

these parameters depend on common operational or structural factors of the compressor.

Conclusion 3: Stable values for the parameters T_{02} and P_{02} suggest that the problems do not occur at the compressor's inlet but may be related to processes occurring within its interior or at the outlet.

Conclusion 4: A more thorough diagnostic of the compressor is necessary, particularly in the areas indicated by the parameters T_{03}, P_{03}, RMS, and Compression, to identify the causes of the faults and to plan appropriate corrective actions.

The experiment utilized synthetic diagnostic data prepared by NASA. Such data offers a range of advantages as well as certain limitations. The primary advantage lies in the ease of access and verification of the data, along with the ability to simulate specific conditions required for testing. This allows researchers to conduct controlled experiments and analyze performance under ideal or hypothetical scenarios without the constraints of real-world variability.

However, one of the significant drawbacks of using synthetic data is the inability, or substantial difficulty, to replicate the noise and anomalies that naturally occur during the actual operation of the equipment. These real-world irregularities, often caused by wear, environmental factors, or unpredictable operational conditions, are critical for comprehensive diagnostics. Consequently, synthetic data may not fully capture the complexity and variability of real-world systems, which can limit the accuracy of predictions and the robustness of diagnostic models when applied to live scenarios.

CONCLUSIONS

After analyzing the results provided by the model, it is evident that LLMs demonstrate significant capabilities in handling and processing the complex matrix data. The model transformed data into graphical representations efficiently and derived linguistically and substantively accurate conclusions. Although the drawn conclusions were not at an advanced technical level, it is important to note that the model operated without additional domain-specific information related to the turbine engine diagnostics and operations, relying solely on its embedded knowledge base.

The conclusions indicate that LLMs can process and analyze data effectively even in specialized fields where they have not been specifically trained. The model's ability to generate correct interpretations based on a foundational database suggests a strong potential for broader applications in various technical domains. Moreover, the model demonstrated an impressive capacity for self-optimization and error correction during the experiment. This adaptive behaviour reduced the time required for research and analysis significantly. The model identified and rectified its own errors in real-time, showcasing its utility in streamlining and enhancing the efficiency of analytical processes.

In summary, LLMs are very promising in the field of the data analysis, capable in processing of complex datasets, in generating visual representations, and in providing accurate conclusions. Their potential to adapt and optimize their functioning autonomously underscores their value in both academic and practical applications. Future studies and researches could involve integrating more specialized knowledge into LLMs to enhance their capability for the advanced technical analysis and diagnostics.

Acknowledgements

This work was partly supported by Polish Ministry of Education and Science fund for Statutory Activities of: the Institute of Transport, PUT (PL) 0416/SBAD/0006 and the Institute of Powertrains and Aviation, PUT (PL) 0415/SBAD/0342.

REFERENCES

1. Naveed H., Khan A.U., Qiu S. et. al. A comprehensive overview of Large Language Models. Computer Science, Computation and Language, 2024. arXiv preprint arXiv:2307.06435. https://doi.org/10.48550/arXiv.2307.06435
2. Mahowald K., Ivanova A.A., Blank I.A. et al. Dissociating language and thought in large language models. Trends in Cognitive Sciences. Elsevier, Feature review, 2024, 28(6), 517–540. https://doi.org/10.1016/j.tics.2024.01.011
3. Birhane A., Kasirzadeh A., Leslie D., Wachter S. Science in the age of large language models. Nature Reviews Physics, 2023, 5, 277–280. https://doi.org/10.1038/s42254-023-00581-4
4. Collins K.M., Jiang A.Q., Frieder S. et al. Evaluating language models for mathematics through interactions. Computer Science, Machine Learning, 2023. arXiv preprint arXiv:2306.01694. https://doi.org/10.48550/arXiv.2306.01694
5. Gao I., Ilharco G., Lundberg S., Ribeiro M.T. Adaptive testing of computer vision models. Computer Science, Computer Vision and Pattern Recognition, 2022. arXiv preprint arXiv:2212.02774. https://doi.org/10.48550/arXiv.2212.02774
6. Monigatti L. From prototype to production: Vector databases in generative AI applications. 2023. (Online). Available: https://stackoverflow.blog/2023/10/09/from-prototype-to-production-vector-databases-in-generative-ai-applications/.
7. Shanahan M., McDonell K., Reynolds L. Role play with large language models. Nature, 2023, 623, 493–498. https://doi.org/10.1038/s41586-023-06647-8
8. Bodroza B., Dinic B.M., Bojic L. Personality testing of GPT-3: Limited temporal reliability, but highlighted social desirability of GPT-3's personality instruments results. Computer Science, Machine Learning, 2023. arXiv preprint arXiv:2306.04308. https://doi.org/10.48550/arXiv.2306.04308
9. Bubeck S., Chandrasekaran V., Eldan R. et al. Sparks of artificial general intelligence: Early experiments with GPT-4. Computer Science, Machine Learning, 2023. arXiv preprint arXiv:2303.12712. https://doi.org/10.48550/arXiv.2303.12712
10. Bartling M. Decídalo insights: About LLMs and chatbots. 2023. [Online]. Available: https://www.data-assessment.com/en/llms-and-chatbots/
11. Velychko D., Osukhivska H., Palaniza Y., Lutsyk N., Sobaszek Ł. Artificial intelligence based emergency identification computer system. Advances in Science and Technology Research Journal, 2024; 18(2): 296–304. https://doi.org/10.12913/22998624/184343
12. Fan W., Fan W., Li J. et. al. Recommender systems in the era of Large Language Models (LLMs). Computer Science, Information Retrieval, 2023. arxiv:cs.IR/2307.02046. https://doi.org/10.48550/arXiv.2307.02046
13. Hou X., ZhaoY., Liu Y. et. al. Large language models for software engineering: A systematic literature review. Computer Science, Software Engineering, 2024. arXiv preprint arXiv:2308.10620. https://doi.org/10.48550/arXiv.2308.10620
14. Imani S., Du L., Shrivastava H. MathPrompter: Mathematical Reasoning using Large Language Models. Computer Science, Computation and Language, 2023. arXiv preprint arXiv:2303.05398. https://doi.org/10.48550/arXiv.2303.05398
15. Nvidia. Large Language Models. 2023. [Online].

Available: https://www.nvidia.com/en-us/glossary/large-language-models/

16. Sourceforge. Best vector databases for hugging face. 2024. [Online]. Available: https://sourceforge.net/software/vector-databases/integrates-with-hugging-face/
17. Chang Y., Wang X., Wang J. et. al. A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 2023, 15(3), 39, 1–45. https://doi.org/10.1145/3641289
18. Fu Y., Ou L., Chen M. et. al. Chain-of-thought hub: A continuous effort to measure large language models' reasoning performance. Computer Science, Computation and Language, 2023. arXiv preprint arXiv:2305.17306. https://doi.org/10.48550/arXiv.2305.17306
19. Wei J., Tay Y., Bommasani R. et. al. Emergent abilities of Large Language Models. Computer Science, Computation and Language, 2022. arXiv preprint arXiv:2206.07682. https://doi.org/10.48550/arXiv.2206.07682
20. Seltek W., Kowalik R., Lusiak T. Practical use of composite materials used in military aircraft, Materials 2021, 14(17), 4812; https://doi.org/10.3390/ma14174812
21. Turygin Y., Bozek P., Abaramov I., Nikitin Y. Reliability Determination and Diagnostics of a Mechatronic System. Advances in Science and Technology Research Journal, 2018; 12(2): 274–290. https://doi.org/10.12913/22998624/92298
22. Kieszek R., Kozakiewicz A., Rogólski R. Optimization of a Jet Engine Compressor Disc with Application of Artificial Neural Networks for Calculations Related to Time and Mass Criteria. Advances in Science and Technology Research Journal, 2021, 15(2): 208–218. https://doi.org/10.12913/22998624/134061
23. Danilecki S. Eksploatowanie samolotów. (2nd ed.). Wojskowa Akademia Techniczna, Warszawa, 2016.
24. El-Sayed A.F. Aircraft propulsion and gas turbine engines. (2nd ed.). Boca Raton. Taylor & Francis Group, LLC, 2017, 205. https://doi.org/10.1201/9781315156743
25. Lewitowicz J. Podstawy eksploatacji statków powietrznych – statek powietrzny i elementy teorii. Wydawnictwo Instytutu Technicznego Wojsk Lotniczych, Warszawa, 2001.
26. Lewitowicz J. Problems of failures of airplanes and helicopters. Journal of KONES, Warszawa, 2007, 14(3), 339-346. [Online]. Available: https://bibliotekanauki.pl/articles/246444
27. Lewitowicz J. Eksploatacyjne własności i właściwości samolotów i śmigłowców. Wydawnictwo Instytutu Technicznego Wojsk Lotniczych. Warszawa, 2015.
28. Szrama S. Lotnicze napędy turbinowe. Wydawnictwo Politechniki Poznańskiej, Poznań, 2023.
29. Sutton G.P., Biblarz O. Rocket propulsion elements. (9th ed.). Wiley, 2016.
30. Bian N., Han X., Sun L. et. al. ChatGPT is a knowledgeable but inexperienced solver: An investigation of commonsense problem in large language models. Computer Science, Machine Learning, 2023. arXiv preprint arXiv:2303.16421. https://doi.org/10.48550/arXiv.2303.16421
31. Ferrara E. Should ChatGPT be biased? Challenges and risks of bias in large language models. Computer Science, Computers and Society, 2023. arXiv preprint arXiv:2304.03738. https://doi.org/10.48550/arXiv.2304.03738
32. Frieder S., Pinchetti L., Chevalier A. et. al. Mathematical capabilities of ChatGPT. Computer Science, Machine Learning, 2023. arXiv preprint arXiv:2301.13867. https://doi.org/10.48550/arXiv.2301.13867
33. Schwaber-Cohen R. What is a vector database & how does it work? Use Cases + Examples. 2023. [Online]. Available: https://www.pinecone.io/learn/vector-database/
34. Wisłocki K. Napędy lotnicze: podstawy, systematyka, historia, konstrukcja. Wydawnictwo Politechniki Poznańskiej, Poznań, 2023.
35. Chen M., Tworek J, Jun H. et al. Evaluating large language models trained on code. Computer Science, Machine Learning, 2021. arXiv preprint arXiv:2107.03374. https://doi.org/10.48550/arXiv.2107.03374
36. Mialon G., Dessì R., Lomeli M. et. al. Augmented language models: a survey. Computer Science, Computation and Language, 2023, arXiv preprint arXiv:2302.07842. https://doi.org/10.48550/arXiv.2302.07842
37. Xu F.F., Alon U., Neubig G., Hellendoorn V.J. A systematic evaluation of large language models of code. MAPS 2022: Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming, 2022, 1-10. https://doi.org/10.1145/3520312.3534862
38. Zhao H. et. al. Explainability for Large Language Models: A Survey. ACM Transactions on Intelligent Systems and Technology, 2024, 15(2), 20, 1–38. https://doi.org/10.1145/3639372
39. Vaswani A., Chen H., Yang F. et al. Attention is all you need. Advances in Neural Information Processing Systems, 30, 2017, 5998–6008. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
40. All M. The top 5 vector databases, a comprehensive guide to the best vector databases. Master high-dimensional data storage, decipher unstructured

information, and leverage vector embeddings for AI applications. 2023. [Online]. Available: https://www.datacamp.com/blog/the-top-5-vector-databases.

41. OpenAI. API do korzystania z bibliotek OpenAI. OpenAI. 2023. [Online]. Available: https://openai.com/product#made-for-developers.
42. OpenAI. platform.openai.com. OpenAI. 2023. [Online]. Available: https://platform.openai.com/docs/guides/embeddings.
43. Legutko S. Eksploatacja maszyn. Wydawnictwo Politechniki Poznańskiej, Poznań,2007.
44. Mix P.E. Introduction to nondestructive testing. A training guide. (2nd ed.). John Wiley & Sons, Inc., 1987.
45. Nosal S. Tribologia. Wprowadzenie do zagadnień tarcia, zużywania i smarowania. Wydawnictwo Politechniki Poznańskiej. Poznań, 2012.
46. Ostwald M. Wytrzymałość materiałów i konstrukcji. Zbiór zadań. (3rd ed.). Wydawnictwo Politechniki Poznańskiej, Poznań, 2018.
47. Randall R.B. Vibration-based condition monitoring: industrial, aerospace, and automotive applications. Chichester. John Wiley & Sons, Ltd., 2011. https://doi.org/10.1002/9780470977668
48. Pavlikova Ľ., Hricová B., Lumnitzer E. Acoustic camera as a fool for identifying machinery and equipment failures. Advances in Science and Technology Research Journal, 2018; 12(1): 322–328. https://doi.org/10.12913/22998624/87110
49. Michaelson S., Bishop R.E.D., Gladwell G.L.E. Macierzowa analiza drgań. Państwowe Wydawnictwo Naukowe, Warszawa, 1972.
50. PWN. sjp.pwn.pl. 2024. [Online]. Available: https://sjp.pwn.pl/sjp/stan;2523760.html
51. Polski Komitet Normalizacyjny. PN-N-04001:1982: Eksploatacja obiektów technicznych. Terminologia ogólna. Norma. Polski Komitet Normalizacyjny, Warszawa, 1982.
52. Niziński S., Michalski R. Diagnostyka obiektów technicznych. Biblioteka Problemów Eksploatacji. Wydawnictwo i Zakład Poligrafii Instytutu Technologii Eksploatacji, Radom, 2002, 15-20.
53. Niziński S., Żółtowski B. Informatyczne systemy zarządzania eksploatacją obiektów technicznych. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego (Olsztyn), Wydawnictwo Akademii Techniczno-Rolniczej (Bydgoszcz). Olsztyn – Bydgoszcz, 2001.
54. Pszczółkowski J.S., Goliasz T. Istota strategii eksploatacji wg kryteriów niezawodności. Autobusy, Technika, Eksploatacja, Systemy Transportowe, Instytut Naukowo-Wydawniczy "Spatium", 2019, 24(6), 161–167. https://doi.org/10.24136/atest.2019.144
55. Szczepanik R., Witos M. Monitorowanie stanu technicznego turbinowych silników lotniczych. Research Works of Air Force Institute of Technology, Warszawa, 2000, 2000(10), 119–129. [Online]. Available: https://www.researchgate.net/publication/262298082_Monitorowanie_stanu_technicznego_turbinowych_silnikow_lotniczych_w_oparciu_o_metode_dyskretno-fazowa_i_metode_portretow_fazowych
56. Wikipedia.org. Wikipedia: Isentropic process. 2024. [Online]. Available: https://en.wikipedia.org/wiki/Isentropic_process.
57. Alto V. Modern generative AI with ChatGPT and OpenAI models: Leverage the capabilities of OpenAI's LLM for productivity and innovation with GPT3 and GPT4. 2023.
58. Boegner M.. Knowledge retrieval architecture for LLM's. 2023. [Online]. Available: https://mattboegner.com/knowledge-retrieval-architecture-for-llms/
59. Ozdemir S. Quick start guide to Large Language Models: Strategies and best practices for using ChatGPT and other LLMs. Addison-Wesley Data & Analytics Series. Addison-Wesley Professional, 2023.