Supplementary Information for

An Intelligent Maintenance System for Wind Turbine Gearboxes with Integrated Deep Learning and Large Language Models

Author：Xueyi Li, Guangsheng Ran, Tianyang Wang, Guangyao Zhang, Qi Li, Ming Li, Jianxin Tan, Cheng Shi, Haochen Qi, Hua Li, Fangjie Lu, Yining Xie, Fulei Chu

Supplementary Dataset Construction

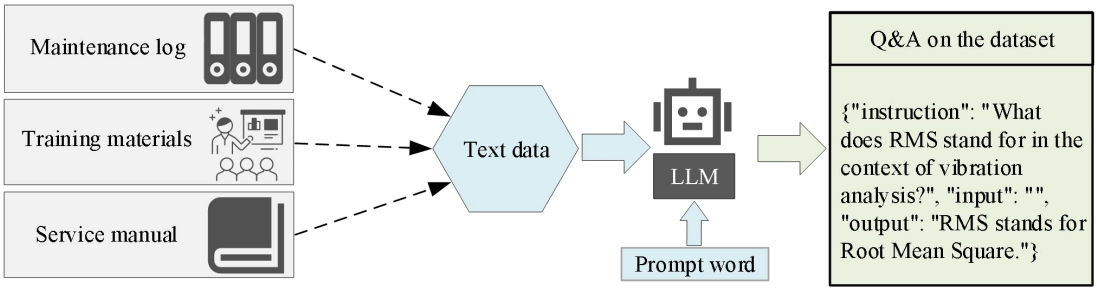
In order to develop a specialized large language model and train the Signal Transformer, this research has compiled an extensive dataset that merges real-world maintenance logs and training materials from wind farms, along with vibration signal data pertaining to wind turbine gearbox failures.

Supplementary Dataset for Fine-tuning

The research gathered maintenance logs from several wind farms, which meticulously documented a range of fault occurrences, repair procedures, tools and techniques employed, and the outcomes of repairs for wind turbine gearboxes. Through a process of selecting and preprocessing the content of these logs, the research isolated information directly associated with gearbox fault repairs, resulting in the generation of 13,426 question-answer pairs based on fault issues and their corrective actions.

Furthermore, the research compiled maintenance training resources, such as repair manuals, fault diagnosis handbooks, and case studies. Initially, the research extracted textual information that detailed the gearbox's construction, operational principles, and standardized procedures for fault diagnosis and repair. Subsequently, this content was paired with prompt phrases and systematically fed into the large language model to transform it into a question-answer pair format, ultimately expanding the dataset to encompass 25,479 entries.

The methodology for constructing the dataset is illustrated in Figure 5. It involves extracting text data from the curated materials, combining it with prompt phrases, and inputting it into the large language model, which is then guided to convert the information into a question-answer pair format.



Supplementary Fig. 5: Dataset Conversion Process

Supplementary Signal Transformer Dataset

Moreover, this research employed a dataset of vibration signals from wind turbine gearbox faults to train the Signal Transformer. This dataset encompasses vibration signals from the gearbox under varying operational conditions, including both normal operations and signals from different fault scenarios. The research utilized a model 608A11 vibration sensor to gather vibration data from the planetary gearbox along the x and y axes, with a sampling rate of 5120 Hz. Two distinct operational conditions were established for the experiment: the speed and load were set at either 20 Hz - 0 V or 30 Hz - 2 V. Under each condition, vibration data were collected for four different bearing fault conditions and one condition of normal health, with each condition assigned a unique label, as detailed in Table 1:

Supplementary Tab. 1 Bearing fault information

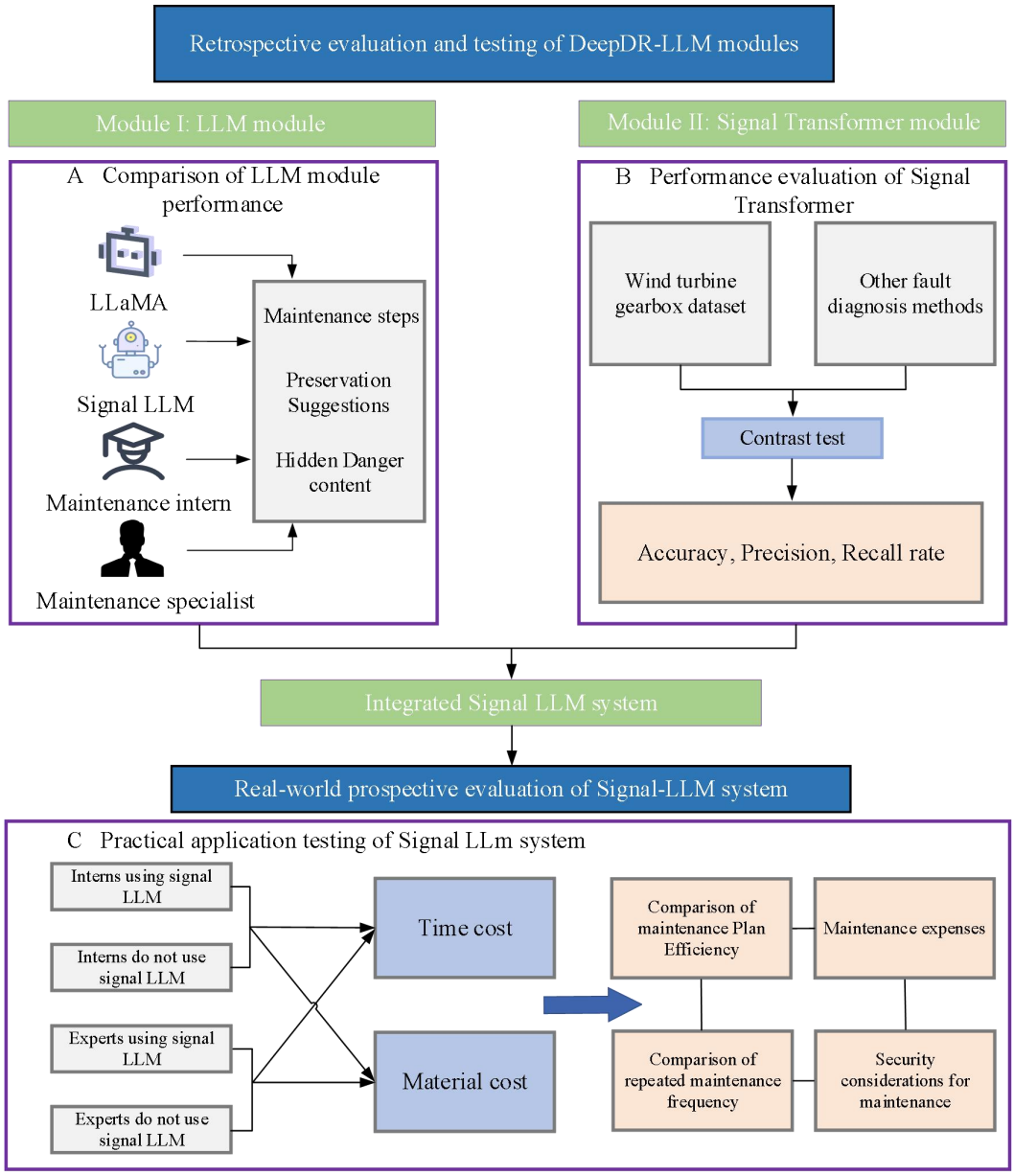
Fault Type	Description	Label
Ball Fault	Cracks present in the ball element	Class 1
Complex Fault	Cracks in both the inner and outer races	Class 2
Healthy State	Free of faults	Class 3
Inner Ring Fault	Cracks in the inner race	Class 4
Outer Ring Fault	Cracks in the outer race	Class 5

Through the analysis and processing of these signals, the Signal Transformer is capable of learning the correlation between the gearbox's operational state and its vibration features, which serves as a foundation for fault diagnosis.

Supplementary experimental design

This research consists of three experimental evaluations to assess the system's performance, as illustrated in Figure 6. Experiment one is intended to compare the accuracy of the LLM module in our system in offering maintenance procedures and follow-up maintenance advice against existing models and the performance of maintenance experts. Experiment two aims to evaluate the accuracy of the Signal-Transformer module in diagnosing gearbox faults. Experiment three involves

testing the system's impact on maintenance personnel with varying levels of technical expertise in a real-world setting to confirm its practical effectiveness and economic viability.



Supplementary Fig. 6: Performance Evaluation Experiment