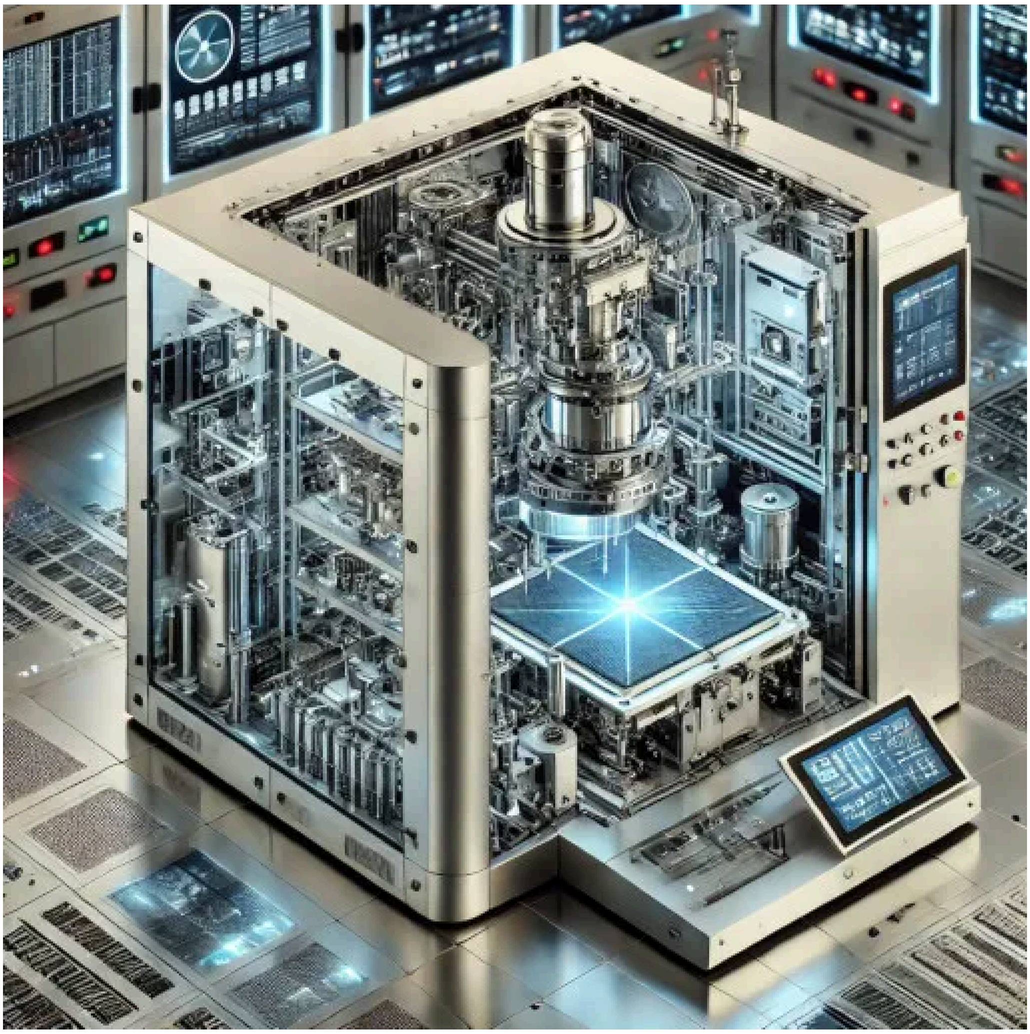
AI-Driven EUV Equipment Predictive Maintenance: Smarter Strategies for High-Tech Manufacturing

Kirill | AI Solutions for High-Tech
14 min read · Mar 3, 2025

The aim of this research is to demonstrate how modern analytical approaches and Industry 4.0 technologies can transform the maintenance methods of complex systems like advanced EUV (Extreme Ultraviolet) equipment, minimizing failure risks and increasing overall productivity. EUV equipment are critical in semiconductor manufacturing, where even minor disruptions can result in significant financial losses. This material provides technical specialists with practical and analytical tools to address the challenges of high-complexity equipment and ensure uninterrupted production processes.



Introduction to Industrial Development and Industry 4.0

Industrial development has gone through several key stages, each fundamentally transforming production and management methods:

> ***First Industrial Revolution:*** *Introduction of mechanization using water and steam power.*
>
> ***Second Industrial Revolution:*** *Mass automation driven by electricity and assembly line technologies.*
>
> ***Third Industrial Revolution:*** *Transition to digital technologies and computerization.*

Today, the world is experiencing the **Fourth Industrial Revolution, or Industry 4.0,** which is transforming production processes through the integration of advanced technologies:

> ***Internet of Things (IoT)***
>
> ***Big Data and Analytics***

> *Artificial Intelligence (AI)*
>
> *Cloud Computing*

These technologies enable companies to automate equipment management, analyze production data, and optimize processes in real time.

Why Transition to Industry 4.0 Is Inevitable

The shift to Industry 4.0 is driven by several factors:

> ***Increased Competition:*** *Globalization demands greater efficiency and production flexibility.*
>
> ***Technological Advancements:*** *New technologies create opportunities for optimization and productivity improvements.*
>
> ***Changing Customer Needs:*** *Rising demand for customized products and rapid responses to market changes.*
>
> ***Economic Efficiency:*** *Automation and digitalization reduce costs and enhance productivity.*

Benefits of Industry 4.0: How Technologies Transform Manufacturing

Industry 4.0 technologies are revolutionizing industrial enterprises, turning them into smart ecosystems where efficiency, flexibility, and innovation reach new heights. These advancements not only accelerate production processes but also lay a sustainable foundation for long-term growth.

Let's take a closer look at how this transformation unfolds.

Increased Efficiency: Less Downtime, More Results

One of the key achievements of Industry 4.0 is the use of digital twins — virtual replicas of real production processes. With their help, companies can test scenarios and address issues before they occur in reality. This reduces downtime by 20–30%. ***Example:*** *At Bosch factories, the implementation of IoT devices and an analytical platform increased the productivity of assembly lines by 15%. Thanks to precise data tracking and prompt responses to changes, each production cycle has become more efficient [1].*

Flexibility: Adapting to New Conditions in Hours

In the face of unstable demand, businesses need to adapt quickly. Automation of production lines allows companies to switch between different types of products within a matter of hours. This is especially important for industries where product ranges are frequently updated.

Example: *At Siemens facilities, automated systems enable rapid reconfiguration of equipment to produce new products, significantly reducing downtime [2].*

Improved Quality: Smart Algorithms Ensuring Standards

One of the standout achievements of Industry 4.0 is the application of machine learning for quality control. Real-time data analysis systems can detect defects at early stages of production, preventing their spread.

Example: *Tesla has implemented intelligent algorithms that monitor product quality at every stage. The result is a 25% reduction in defects, which improves customer satisfaction and lowers costs associated with defect corrections [3].*

Cost Reduction: Forecasting Instead of Reacting

One of the main values of Industry 4.0 is the use of predictive analytics. Instead of addressing the aftermath of failures, companies can predict potential equipment breakdowns and prevent them.

Example: *At General Electric facilities, predictive analytics has enabled savings of up to 12% on equipment maintenance. Smart systems analyze sensor data, predict component wear, and indicate precisely when maintenance is needed [4].*

Innovation: New Products and Technologies

The combination of AI and Big Data drives companies to develop new products and improve existing ones. Big Data analysis accelerates design processes and helps identify innovative solutions.

Example: *Intel leverages Big Data for chip design. By implementing analytical platforms, the company has reduced the time required to develop new models, providing a competitive edge in the high-tech sector [5].*

Industry 4.0 is more than just technologies; it's a revolution reshaping the very principles of production — from enhancing precision to creating innovative products. Companies that have already adopted these approaches not only boost their competitiveness but also lay the groundwork for sustainable development.

The Connection Between Industry 4.0, Predictive Maintenance, and Root Cause Analysis

The principles of Industry 4.0 enable a shift from addressing the aftermath of failures to preventing them through Predictive Maintenance (PdM) and Root Cause

Analysis (RCA). These approaches are critically important for high-precision systems, such as innovative EUV equipment, which demand maximum accuracy and minimal downtime.

Industry 4.0: Predictive Maintenance

Predicting Failures with Machine Learning and Big Data Offers the Following Key Benefits:

> ***Financial Efficiency:*** *Downtime of EUV equipment in a factory can cost tens of thousands of USD per hour or more. Preventing failures reduces production losses by millions of dollars annually.*
>
> ***Supply Quality****: Reliable equipment operation ensures timely deliveries and high-quality finished products.*
>
> ***Reduced Maintenance Costs:*** *Condition-based component replacement minimizes excessive repairs, reduces spare parts inventory, and enhances overall efficiency.*

Industry 4.0: Root Cause Analysis

When failures do occur, Root Cause Analysis enables the following:

> ***Problem Identification:*** *Quickly determine the causes of instability or breakdowns.*
>
> ***Process Optimization****: Improve design, configurations, and maintenance intervals to minimize future risks.*

These technologies not only save costs but also help maintain leadership in the global market.

Problem Statement and Experiment Design

Modern EUV (Extreme Ultraviolet) light sources are high-tech equipment used in the production of advanced semiconductors. They utilize extreme ultraviolet light with a wavelength of 13.5 nanometers to create ultra-dense topologies on silicon wafers.

These sources are critical to the lithographic process, ensuring high precision and transistor density, which is essential for the development of powerful and energy-efficient processors, memory, and other electronic components.

The immense complexity of these systems requires precise calibration and the processing of vast amounts of data. The volume of data increases significantly when the system is integrated into a factory's production cycle.

To ensure effective predictive maintenance of the production line, the following aspects must be considered:

Autonomy of the Solution:

> ***Development of an IT system*** *that operates without internet connectivity or reliance on external public services.*

Analysis of Large Telemetry Data Volumes:

> ***Complex data structure:*** *Data volumes can reach 1 GB every 20 seconds, capturing* 1^{05} *parameters from various sensors at frequencies of several dozen Hertz.*

Handling Different Types of Errors:

> ***Degradation changes:*** *Gradual changes in the system that lead to reduced performance or failures.*
>
> ***Unique errors:*** *Combined or rare errors that occur outside typical operational scenarios.*

Limited Data Availability:

> *The solution must perform effectively even in the absence of large amounts of training and classification data.*

No Constant Human Oversight:

> *Autonomous system operation without the need for constant adjustment, monitoring, or intervention by a specialist to maintain high accuracy.*

Working with Unstructured Data:

> *Processing and analyzing data that has not been pre-labeled or structured, including raw text logs or log files.*

Real-Time Operation:

> *Analyzing equipment telemetry data in real time to detect issues and forecast equipment failures.*

Throughout the article, when we refer to equipment or logs, we are specifically referring to EUV equipment.

From Simulation to Prediction: How We Built an Intelligent Data Analysis System

When we began working on a task that seemed solvable only in theory, every step became a challenge. There were no real data sets or ready-made solutions capable of handling such a high-load data stream. However, this task opened up a space for creativity and discovery.

In this article, we'll share how, step by step, we are developing a technology capable of analyzing massive amounts of data for predictive analytics.

The First Hurdle: Creating Data from Nothing

It all began with a simple yet seemingly unsolvable dilemma: how can we test algorithms if the necessary data doesn't exist in a volume sufficient for testing?

The solution came in the form of a program that simulates equipment operation. We created an emulator capable of generating logs for a specialized EUV equipment, based on a specific module example. Naturally, real log messages from an actual EUV equipment differ slightly from those produced by the simulator. However, the message generation principles are realistic.

As a result, we developed more than just a tool for generating random numbers — the emulator became an almost independent model that:

> *Enabled the generation of data from 10,000 sensors at frequencies of up to 100 Hz.*
>
> *Simulated signals from any sensor through gradual degradation or instantaneous spikes.*
>
> *Allowed the creation of various time periods for anomaly occurrences.*

The system we developed didn't just start working — it brought the parameters to life, turning abstract numbers into a real story of the equipment's operation.

Below is a small example of an unstructured log where data from various system sensors is randomly recorded:

```
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.6701707504984 units.
[2024-11-21 00:00:00] INFO: Laser_Wavelength: Current value is 577.791132484859 units.
[2024-11-21 00:00:00] INFO: Beam_Quality_M2: Current value is 449.9305288407662 units.
[2024-11-21 00:00:00] INFO: Pulse_Frequency: Current value is 96.9203246928512 units.
[2024-11-21 00:00:00] INFO: Cooling_Flow: Current value is 151.08195956926008 units.
[2024-11-21 00:00:00] INFO: Spectral_Resolution: Current value is 218.8908997569555 units.
[2024-11-21 00:00:00] INFO: Beam_Steering_Angle: Current value is 816.6643672428085 units.
[2024-11-21 00:00:00] INFO: Beam_Shaping_param_1: Current value is 970.1981406396637 units.
[2024-11-21 00:00:00] INFO: Thermal_Management_param_1: Current value is 955.2371110620372 units.
[2024-11-21 00:00:00] INFO: Feedback_Control_param_1: Current value is 257.42958007339837 units.
[2024-11-21 00:00:00] INFO: Focus_Control_param_1: Current value is 554.161227013597 units.
[2024-11-21 00:00:00] INFO: Acousto-Optic_Modulators_param_1: Current value is 177.087947719098 units.
[2024-11-21 00:00:00] INFO: Spectroscopy_param_1: Current value is 706.3649219146524 units.
[2024-11-21 00:00:00] INFO: Wavelength_Control_param_1: Current value is 739.5724291287613 units.
[2024-11-21 00:00:00] INFO: Pulse_Generation_param_1: Current value is 80.79899959222257 units.
[2024-11-21 00:00:00] INFO: Frequency_Doubling_param_1: Current value is 704.4345560309972 units.
[2024-11-21 00:00:00] INFO: Focus_Control_param_2: Current value is 319.6334325407725 units.
[2024-11-21 00:00:00] INFO: Acousto-Optic_Modulators_param_2: Current value is 813.9442257229196 units.
[2024-11-21 00:00:00] INFO: Optics_Alignment_param_1: Current value is 407.78849480496376 units.
[2024-11-21 00:00:00] INFO: Acousto-Optic_Modulators_param_3: Current value is 819.7137030569038 units.
[2024-11-21 00:00:00] INFO: Beam_Control_param_1: Current value is 235.23255155651387 units.
[2024-11-21 00:00:00] INFO: Wavelength_Control_param_2: Current value is 196.40844019994432 units.
[2024-11-21 00:00:00] INFO: Spectroscopy_param_2: Current value is 632.7866171618999 units.
[2024-11-21 00:00:00] INFO: Safety_Systems_param_1: Current value is 927.5481374215436 units.
[2024-11-21 00:00:00] INFO: Polarization_Control_param_1: Current value is 841.6157154971984 units.
[2024-11-21 00:00:00] INFO: Laser_Cavities_param_1: Current value is 164.89039681646472 units.
[2024-11-21 00:00:00] INFO: Laser_Cavities_param_2: Current value is 208.7589753912862 units.
[2024-11-21 00:00:00] INFO: Laser_Cavities_param_3: Current value is 517.1734316374778 units.
[2024-11-21 00:00:00] INFO: Acousto-Optic_Modulators_param_4: Current value is 21.6957913576791 units.
[2024-11-21 00:00:00] INFO: Beam_Quality_param_1: Current value is 961.7281804628589 units.
[2024-11-21 00:00:00] INFO: Acousto-Optic_Modulators_param_5: Current value is 388.21722697820394 units.
[2024-11-21 00:00:00] INFO: Beam_Control_param_2: Current value is 60.27973666393552 units.
[2024-11-21 00:00:00] INFO: Wavelength_Control_param_3: Current value is 981.648645159968 units.
[2024-11-21 00:00:00] INFO: Frequency_Doubling_param_2: Current value is 153.03523845885306 units.
[2024-11-21 00:00:00] INFO: Optics_Alignment_param_2: Current value is 758.6095318259114 units.
[2024-11-21 00:00:00] INFO: Acousto-Optic_Modulators_param_6: Current value is 813.0076854699417 units.
[2024-11-21 00:00:00] INFO: Modulation_System_param_1: Current value is 118.56253244267091 units.
```

Example of a Log File: Data from 1,000 sensors is received in random order simultaneously at a frequency of 20 Hz. Each sensor has its own individual settings.

Multitasking: The Parallel World of Data Generation

Generating data turned out to be more complex than it might seem. Simulating the operation of the system is a time-consuming process. To speed up the process to a reasonable time frame (for example, generating a simulation of one day of system operation within an hour), we had to implement parallel computing in memory. Up to 200 processes ran simultaneously, each modeling a portion of the time series for specific sensors.

Multithreading and Synchronization: Managing Chaos

One of the key challenges in creating the simulator was organizing multithreaded computations in memory. Data from different processes, each generating its own stream, arrived out of order and at different times, making the task of assembling everything in memory the first puzzle to solve.

To address this challenge, we needed not only to ensure high speed but also to properly synchronize the results. The following became critical:

> ***Memory Overflow Control:*** *Efficient resource allocation to ensure that, even under peak loads from the data generator, the system remains stable and no data is lost.*

> ***Continuous Data Streaming to Storage:*** *Generated results are transmitted to the database without delays, maintaining both the speed of generation and reliability. Post-processing of the generated data allowed for the construction of a correct time series.*

This approach enabled us not only to process massive amounts of information but also to do so reliably, with a high degree of accuracy, preserving the integrity and synchronization of the data.

The result? We developed a module that synchronizes and consolidates data with remarkable precision, avoiding system overload. In just 15 minutes, our generator could create a data volume comparable to a full day's operation of an EUV equipment.

For instance, most of our calculations were performed on data simulating 7 hours of equipment operation with 1,000 parameters recorded at a frequency of 20 Hz. The resulting data volume from such a simulation reached approximately 60 GB.

Below is an example of a configuration file used for data emulation:

```
▼ Laser_Subsystem:
    module_name:                          "Laser_Subsystem_V0_0005_1D001_1D001_1D001"
    sampling_rate:                        20
    duration:                             100
    start_time:                           "2024-11-21T00:00:00Z"
  ▼ degradation_parameters:
    ▼ 0:
        parameter:                        "Main_Laser_Power"
        degradation_percent:              0.01
        duration_sec:                     250
        number_of_degradations:           10
        degradation_type:                 "gradual"
  ▼ parameters:
    ▼ Main_Laser_Power:
        default:                          142.66445398248715
        variation_prc:                    0.00227155534218
        unit:                             "units"
    ▼ Laser_Wavelength:
        default:                          577.7745107366123
        variation_prc:                    0.00332801577913
        unit:                             "units"
    ▼ Beam_Quality_M2:
        default:                          449.9396316757038
        variation_prc:                    0.00396961631193
        unit:                             "units"
    ▶ Pulse_Frequency:                    {…}
    ▶ Cooling_Flow:                       {…}
    ▶ Spectral_Resolution:                {…}
    ▶ Beam_Steering_Angle:                {…}
    ▶ Beam_Shaping_param_1:               {…}
    ▶ Thermal_Management_param_1:         {…}
    ▶ Feedback_Control_param_1:           {…}
    ▶ Focus_Control_param_1:              {…}
    ▶ Acousto-Optic_Modulators_param_1:   {…}
    ▶ Spectroscopy_param_1:               {…}
    ▶ Wavelength_Control_param_1:         {…}
    ▶ Pulse_Generation_param_1:           {…}
    ▶ Frequency_Doubling_param_1:         {…}
    ▶ Focus_Control_param_2:              {…}
    ▶ Acousto-Optic_Modulators_param_2:   {…}
    ▶ Optics_Alignment_param_1:           {…}
    ▶ Acousto-Optic_Modulators_param_3:   {…}
    ▶ Beam_Control_param_1:               {…}
    ▶ Wavelength_Control_param_2:         {…}
    ▶ Spectroscopy_param_2:               {…}
    ▶ Safety_Systems_param_1:             {…}
    ▶ Polarization_Control_param_1:       {…}
    ▶ Laser_Cavities_param_1:             {…}
    ▶ Laser_Cavities_param_2:             {…}
```


Example of a Small Section of the Configuration File for the Data Emulation Program. The file lists all parameters, their permissible values, and degradation parameters.

Classic or Innovation: Choosing the Analysis Method

It might seem that analyzing data stored in text logs is a straightforward task. However, standard methods like regular expressions (regexp) proved inadequate. We quickly realized that such a complex problem required more than just pattern matching in strings.

This marked the beginning of an era of experimentation with artificial intelligence. Public models, such as BERT and ChatGPT, demonstrated their potential utility, but their high computational demands for autonomous deployment or dependence on external services limited their functionality for real-time local operations. This pushed us to forge our own path.

We decided to build our own LLM (Large Language Model) specifically tailored for analyzing equipment logs. It was akin to constructing a skyscraper from scratch — a labor-intensive process requiring vast machine time and complex computations, but one that gave us complete control over the process.

New Horizons of Analysis: Metrics and Patterns

During the analysis process, we developed a set of metrics consisting of more than 50 indicators that help not only to understand the current data but also to predict future changes.

Below are some of the metrics grouped by their main objectives:

Statistical Metrics

> *1. **Mean** and **Median**: These metrics provide an understanding of the typical value in the data.*
>
> *2. **Minimum** and **Maximum** values, as well as the **Range**: These help to understand how much values vary within the analyzed segment.*

Stability Metrics

> *1. **Mean Absolute Deviation (MAD)**: Indicates how much the values deviate from the mean.*
>
> *2. **Median Absolute Deviation**: Evaluates the robustness of the data and is less affected by sharp outliers.*

Spectral Analysis

> *1. **Dominant Frequency:** Indicates how often a specific pattern repeats in the data.*
>
> *2. **Zero-Crossing Rate:** The number of times values switch from positive to negative or vice versa, helping to detect rapid signal changes.*
>
> *3. **Spectral Entropy:** Measures how evenly the signal's energy is distributed across frequencies.*
>
> *4. **Spectral Bandwidth:** Represents the range of frequencies covering the majority of the signal's energy.*

Peak Detection

> *1. **Number of Peaks:** Indicates how many significant spikes occurred in the signal.*

Long-Term and Short-Term Trends

> *1. **Short-Term Average:** Reflects the behavior of data over the most recent time periods.*
>
> *2. **Long-Term Average:** Shows the overall trend.*

Speech Analysis Metrics

> *1. **MFCC Coefficients:** Used for analyzing and recognizing speech signals by identifying unique acoustic characteristics of the signal.*
>
> *2. **Amplitude Modulation:** Helps to determine changes in the signal's volume over time.*
>
> *3. **Frequency Modulation:** Indicates frequency changes, which are useful for identifying oscillatory patterns.*

Various metrics help transform complex data into comprehensible signals that are easy to interpret. For instance, changes in spectral entropy and an increase in frequency peaks may indicate system instability, while deviations in MFCC coefficients can signal a shift in the system's behavioral pattern.

Next, we will demonstrate how these metrics allow us to analyze the system's state.

To provide greater clarity, we will perform a series of operations that are not required for the final result but will give readers of this article a "peek under the hood" to see how everything works.

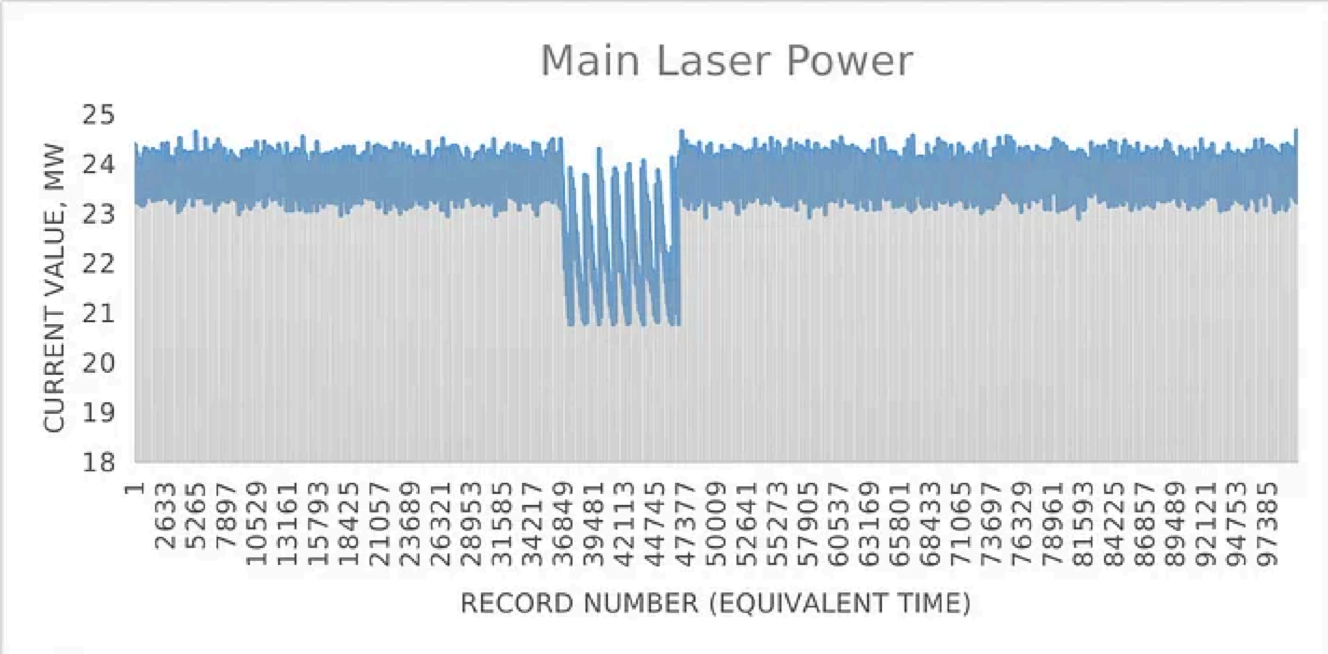
For the analysis, we used emulated system operation data in the form of a log file, an example of which was provided earlier. When generating this file, we specified that the parameter **Main Laser Power** would undergo nine periods of degradation.

Below, for illustrative purposes, we specifically filtered the values of the **Main Laser Power** parameter from the general log. The image shows the values this parameter takes. The data recording frequency is 20 Hz.

```
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.6701707504984 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:00] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:01] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:01] INFO: Main_Laser_Power: Current value is 142.66657216559352 units.
[2024-11-21 00:00:01] INFO: Main_Laser_Power: Current value is 142.66971800179783 units.
[2024-11-21 00:00:01] INFO: Main_Laser_Power: Current value is 142.6624714622381 units.
[2024-11-21 00:00:01] INFO: Main_Laser_Power: Current value is 142.6627423352196 units.
[2024-11-21 00:00:01] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:01] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:01] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
[2024-11-21 00:00:01] WARNING: Main_Laser_Power: Current degraded value is 142.65473187640964 units.
```

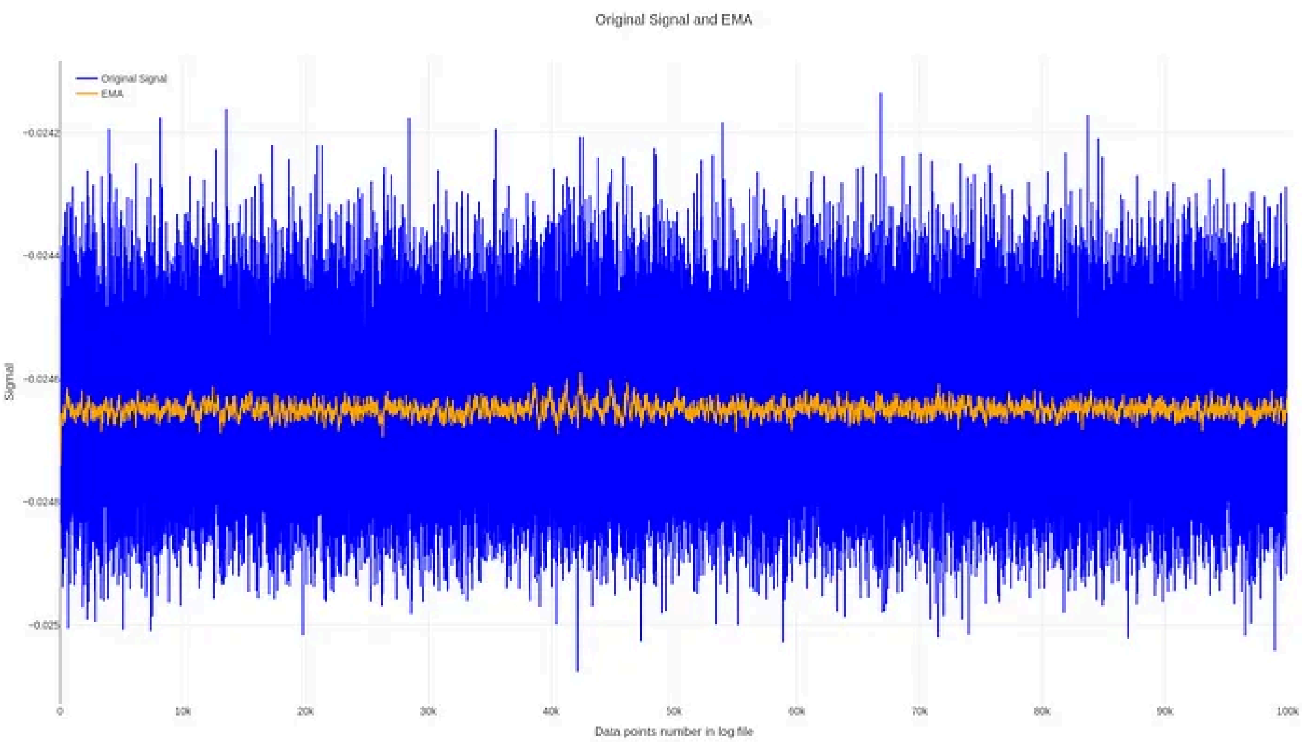
Example of filtered rows for the “Main Laser Power” parameter from the general log.

The next step is to further demonstrate how the parameter Main Laser Power changed over time using a graph.



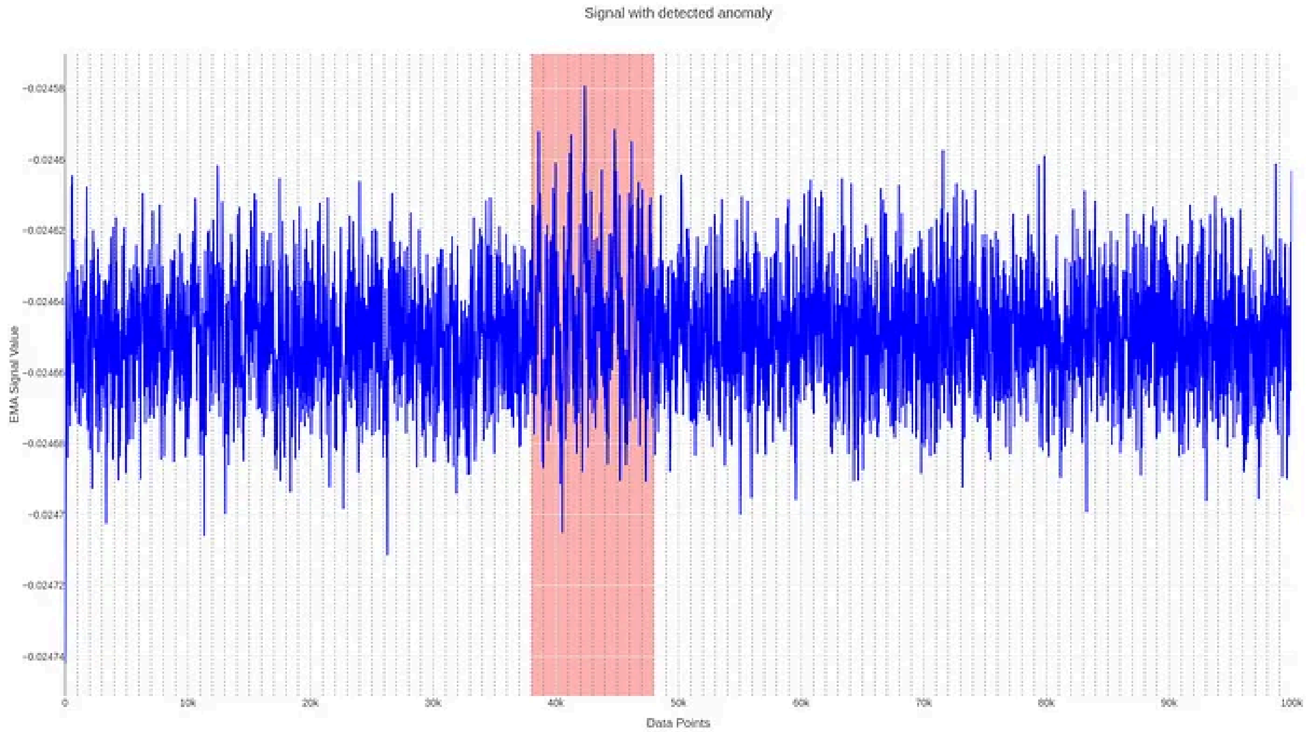
Example of changes in the "Main Laser Power" parameter over time index. The graph clearly shows a period when the parameter values exceed the normal operating range.

The next step is to digitize the log file using the previously created LLM model. The result of the digitization can be seen in the graph below.

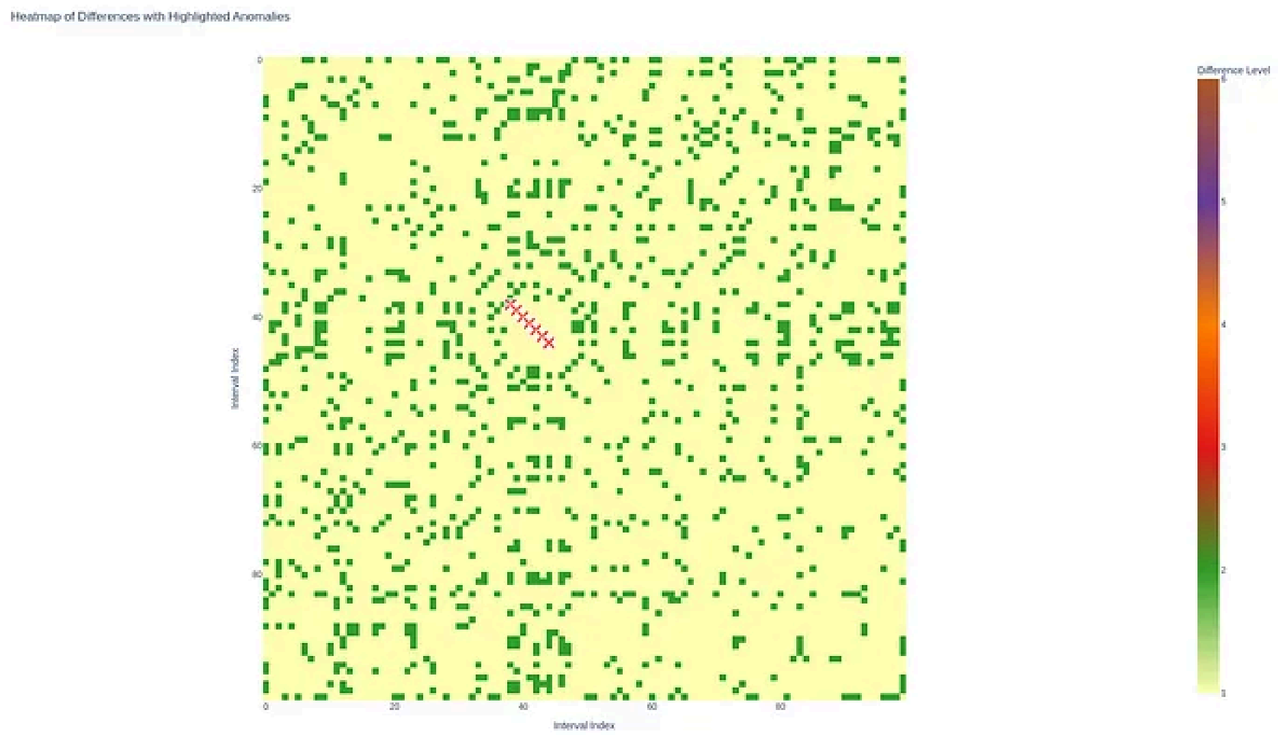


Example of a Digitized Log File Using LLM Algorithms. On the x-axis: data time index. On the y-axis: log file representation as a function. The blue line represents the original signal. The orange line represents the averaged value calculated using an exponential moving average.

The resulting function is effectively analyzed using the existing mathematical framework with the metrics outlined earlier. The outcome of the analysis is the identification of anomaly zones.



Example of a Detected Anomaly (15% Degradation of the "Main Laser Power" parameter). On the x-axis: data time index. On the y-axis: log file representation as a function. The anomaly zone detected automatically is highlighted in pink.



Example of a Detected Anomaly: An Alternative Representation on a Heatmap.

Choosing Tools for Working with LLM: Advantages of a Custom Solution

When selecting tools for working with LLM models, we explored various options, including LangChain and NVIDIA technologies such as RAPIDS and TensorRT. However, at the initial stage of development, it was important for us to create a custom solution that would allow for a deeper understanding of the internal structure and logic of using LLM models, as well as provide maximum control over the process.

We view NVIDIA solutions as promising and consider their integration in later stages when our technology reaches a higher level of maturity. This approach will enable us to leverage the advantages of ready-made tools for further process optimization, improve computational speed, and maintain the flexibility and uniqueness of our solution.

Key Advantages of a Custom Solution:

> ***Offline Operation:*** *Works without internet access.*
>
> ***Full Adaptation to Requirements:*** *By creating our own code, we were able to fine-tune anomaly detection algorithms to meet the specific requirements of the task and data type, which is not possible with universal libraries.*
>
> ***Performance Optimization:*** *The ability to optimize the code for specific types of signals and data models enabled us to achieve better performance and accuracy.*
>
> ***Flexibility and Scalability:*** *The custom solution is easy to modify and expand as requirements grow and change, allowing for the addition of new metrics or adjustments to existing algorithms without the rigid constraints of third-party libraries.*
>
> ***Minimization of External Dependencies:*** *Reducing the number of third-party dependencies simplifies code management and lowers the risk of future compatibility issues, as well as the risk of losing support from library providers.*
>
> ***No Licensing Restrictions:*** *Using custom code eliminates the need to comply with licensing terms of third-party libraries.*

Innovation Under the Hood: Unique Technical Approaches

Developing any new technology is a field for experimentation, where standard solutions often need to be adapted or replaced with entirely new approaches. We took an unconventional path, stepping beyond traditional boundaries and implementing unique technologies and methods to create an efficient and scalable system for analyzing data from equipment.

What's Next? A Look into the Future

At this stage, we are actively working on the concept of anomaly *fingerprints* — a unique "signature" for each failure. This will enable not only the detection of issues but also an understanding of their nature, dynamics, classification, and, most importantly, the creation of temporal dependencies based on anomaly fingerprints. These dependencies will be essential for predicting how changes in the data could evolve into critical system states.

Another equally important task is enhancing sensitivity and optimizing the approach to properly handle noise and reduce false positives.

Conclusions and Results

Our work has been not just a technical challenge but also a reminder that unconventional approaches and perseverance always lead to results — even in situations where the task initially seems too complex to solve.

During our work, we:

> *Developed a robust approach for analyzing equipment telemetry in the form of log files using LLM models, which can be repeatedly applied to various data types.*
>
> *Created mechanisms for high-load computations, enabling real-time data processing.*
>
> *Formulated effective methods for anomaly detection and identifying slow system degradation.*
>
> *Laid the foundation for a methodology to analyze complex systems while considering cause-and-effect relationships. This approach will help in the future to more accurately identify the root causes of failures and plan preventive measures. Refining and implementing this methodology is our next step.*

The experience gained opens up new opportunities for applying the developed solutions across various industries that demand high reliability and efficiency. Our approach to analyzing equipment telemetry in the form of log files and processing large data sets can be adapted to the specific requirements of different production systems, ensuring their stable operation and preventing failures. Thanks to the mechanisms for high-load computations and methods for analyzing cause-and-effect relationships that we have developed, various processes can be optimized, and the overall performance of systems can be improved.

These achievements can serve as a foundation for further development of monitoring systems based on LLM technologies for predictive maintenance in various industrial sectors, fostering sustainable growth and innovation.

Technology Stack

Python, Apache Kafka, Apache Ignite, PostgreSQL

With a proven approach to complex data analysis and system optimization, we continue to push the boundaries of what AI can achieve in industrial environments, driving progress where it matters most.

Predictive Maintenance · Euv Lithography · Industrial Automation · Artificial Intelligence · Semiconductor Industry

Edit profile

Written by Kirill | AI Solutions for High-Tech

0 Followers · 1 Following

Enterprise Architect and AI Solutions Designer, Driving AI predictive analytics and intelligent automation for Industry 4.0