AI Copilot for FEL Experiments

Mahmoud Ajami[1,2,†]

[1]Hamburg University of Technology, Blohmstr. 15, Hamburg, 21079, Germany
[2]European XFEL, Holzkoppel 4, Schenefeld, 22869, Germany

Abstract
This research investigates the development of an AI Copilot to support experimental scientists and staff during operations at the European X-Ray Free Electron Laser facility as part of ongoing efforts to leverage generative AI in scientific facility operations. The work focuses on problems commonly encountered during operations, identified through an initial survey of scientists and technical staff. The proposed Copilot combines anomaly detection methods with a retrieval-augmented generation (RAG) based knowledge assistant integrated into an open-source messaging platform, enabling scientists to interact with the system through a familiar interface while receiving notifications when anomalies are detected. Control system data collected during operational activities is used to build and evaluate the anomaly detection approach. In particular, a Markov chain–based anomaly detection method is evaluated using a synthetic dataset and later real operational control system data. This work lays the foundation for investigating how an AI copilot can best support European XFEL practitioners, how explainable AI methods based on case-based reasoning can explain anomaly detection outcomes, and how a RAG-based architecture can present real-time anomaly alerts and their explanations in an understandable, actionable, and factually grounded manner while minimizing hallucinations.

Keywords
AI Copilot, European XFEL, Anomaly Detection, Generative AI

1. Introduction

European XFEL is an international research facility located near Hamburg (Germany) that provides intense X-ray laser pulses for experiments in fields such as physics, chemistry, biology, and materials science [1]. Due to the complexity and time-sensitive nature of experimental workflows at the European XFEL, maintaining smooth experiment execution is particularly important, as operational issues can lead to increased time and resource consumption during experiments. Given the rapid advancements in the field of AI, the development of an AI Copilot was identified as a suitable approach for supporting experimental scientists and technical staff during operations [2]. As part of broader efforts to investigate the applicability of AI applications, particularly generative AI at the European XFEL, this project was initiated to develop an AI Copilot capable of assisting users during experiments. The system is intended to work alongside users by combining human expertise with intelligent assistant capabilities, providing anomaly detection and AI-assisted guidance throughout the experimental workflow rather than replacing human operators.

A survey I conducted on experimental scientists and technical staff, revealed that issues related to data acquisition, device errors, and device misconfiguration were among the most common problems affecting experimental workflows. Based on these findings, data acquisition was identified as the most appropriate first test case.

Data acquisition at the European XFEL. The European XFEL operates a data acquisition infrastructure that captures and manages detector data, instrument control information, and experimental metadata during operations. These data streams are processed using a distributed self-developed framework for control and data management, allowing synchronization, monitoring, storage, and online analysis for multiple scientific instruments [3, 4].

ICCBR DC'26: Doctoral Consortium at ICCBR2026, August 16, 2026, Bremen, Germany
Mahmoud.ajami@tuhh.de (M. Ajami)
www.linkedin.com/in/mahmoud-ajami (M. Ajami)

Additionally, data acquisition follows a sequential workflow during experiments [5]. In many experiments, each measurement is effectively a single-shot event, as the sample may be affected or destroyed after being subjected to the X-ray beam [6], resulting in unique detector images each time. Since experiments can generate thousands of measurement events per second [7], large volumes of detector data are produced at very high rates. At the same time, metadata, including timing information, instrument settings, and device states, are collected alongside the detector images. Before storage, the control and data acquisition system synchronizes and organizes these data streams into structured data files for later processing and analysis [8].

From this perspective, the experimental workflow can be seen as a step-by-step process that includes experimental measures and actions performed by experts. This creates a continuous stream of operational and experimental data that can be analyzed, modeled, and potentially improved using AI-assisted methods.

My research. My research focuses on exploring the use of AI-assisted approaches for supporting experimental workflows at the European XFEL, starting with data acquisition as the initial use case. The work explores how sequential operational and experimental data can be utilized to support anomaly detection and operational decision-making during experiments, potentially contributing to further improvements in operational efficiency and additional optimization of time and resource utilization. In particular, I aim to provide answers to the following research questions:

1. What support do European XFEL practitioners need from an AI copilot during experimental workflows, and how can these needs guide the design of a useful, trustworthy, and adoptable AI assistant?
2. How can explainable AI methods based on case-based reasoning, including counterfactual explanations, be tailored to explain anomaly detection outcomes in European XFEL experimental workflows?
3. How can a RAG-based architecture be designed to interface an online anomaly detection system with a user interface, so that real-time anomaly alerts and their explanations are presented to scientists in an understandable, actionable, and factually grounded manner while minimizing hallucinations?

2. Research Plan

In this section, the research objectives and the methodology followed during the PhD project are presented. This includes the development of the AI Copilot concept, the current anomaly detection approach, and challenges related to explainability and RAG.

2.1. Research Objectives

The main objective of this research is to develop an AI Copilot that supports experimental scientists and technical staff during operations at the European XFEL. The research focuses on improving operational support, starting with data acquisition as the first test case. The work further aims to investigate the use of machine learning to detect abnormal behavior in control system data, which includes more than 10,000 components such as motors, detectors, cameras, and other operational devices [9, 3], and to assist users in troubleshooting operational issues.

To achieve this, the proposed system consists of three modules: (1) a device integrated within the control system to collect useful operational data; (2) an AI Copilot, which constitutes the primary focus of this PhD project, combining anomaly detection techniques with retrieval-augmented generation (RAG) to detect and summarize possible anomalies in the incoming data streams; and (3) a communication module that forwards relevant information to a communication platform that has already been proven to be effective for AI assistant applications at the facility.

2.2. Approach / Methodology

Assessing the needs To ensure that the proposed AI Copilot addresses real operational needs, I designed a survey and conducted it among experimental scientists and technical staff at the European XFEL. The survey included 65 participants and focused on identifying the main challenges encountered during beamtime operations. Three major issues were identified: data acquisition, beam stability, and device misconfiguration or errors. The survey results indicated that participants generally viewed AI as an opportunity rather than a threat. In addition, respondents reported that their trust in the Copilot would increase if the system provided a confidence score within its notifications, showing how much the anomaly detection model is confident in its own response, allowing users to estimate the reliability of the generated output. Given the results of the survey, the decision was made to focus initially on the data acquisition case, considered the most realistic and technically feasible starting point for the project.

Data collection During the first year of the PhD project, maintenance activities at the facility temporarily paused experimental operations. This limited access to real experimental data that would be later used in the project. However, during the deployment of a new version of the control system, it was possible to collect representative data generated from device restarts, initialization procedures, and calibration processes. Although not originating from active experiments, the collected data follows the same structure as the real operational data expected during operational activities. This step provided valuable insight into the type of data that would be accessible to the AI Copilot and enabled initial exploratory data analysis. It also enabled the development of a data analysis pipeline prepared for future integration with real experimental data. Following the analysis of the different sources of data flowing through our control system device, the focus shifted toward analyzing the data generated from the interaction between users, in this case, experimental scientists and technical staff, and the control system through the Graphical User Interface.

Anomaly detection The primary problem is the detection of abnormal observations in streams of operational and experimental data. Such anomalies may be caused either by deviations in the practitioner's execution of the procedure, or by issues affecting the integrity of the collected data.

A practical example can be found in motors. Many motors at the European XFEL are equipped with end switches that define the safe range of movement and stop the motion if activated. Since the switches represent the two boundaries of the movement range, they cannot be activated simultaneously. However, wiring faults or hardware failures might occasionally produce this physically impossible state, constituting a *data integrity anomaly*. Another common source of motor anomalies, related to a deviation in the practitioner's execution of the procedure, is an incorrect sequence of movement commands. The correct sequence of operations consists in: (1) setting a new target position, (2) issuing the move command, (3) verifying that the motor transitions to the moving state, (4) waiting for the motor to exit the moving state, and (5) verifying that the actual motor position is within a small tolerance of the target position. Anomalies can occur if any of these steps are omitted or performed in the wrong order, for instance, if the move command is issued before the target position is updated.

As a first approach, a Markov chain–based model was developed to learn the expected transitions between operational states from historical data. Each state represents a user interaction with the Graphical User Interface (GUI), such as pressing a button associated with a specific device. During normal operations, users perform sequences of interactions across multiple devices to achieve a particular device configuration required for a given operational procedure, where each procedure consists of multiple sequential steps. Transitions with very low probabilities are then flagged as potential anomalies, enabling the identification of atypical operator behavior compared to normal operational patterns. In the GUI, each operational state consists of a combination of multiple features. The diversity of the GUI data then results in a very large number of unique states, which causes many transitions to be underrepresented, reducing the efficiency of the Markov chain model. This issue can be alleviated by filtering out highly specific and less relevant state features and effectively grouping many similar states under the same operational pattern.

An initial evaluation of the method was performed on synthetic data, offering control on the rate of anomalies. This preliminary study allowed an initial estimation of the anomaly detection performance and helped verify the overall correctness and sanity of the implemented system. The dataset simulates user navigation behavior in an e-commerce website using predefined normal and anomalous state transitions. Both training and test datasets contain mixtures of normal and anomalous transitions. The Markov-chain-based anomaly detection model showed promising preliminary results when applied to the synthetic dataset. However, several aspects still require further refinement, including probability threshold calibration in longer sequences. These challenges are mainly related to the implementation and evaluation methodology rather than to the underlying Markov chain concept itself.

To ensure competitive performance in real-world operational environments, state-of-the-art anomaly detection methods will be implemented [10, 11]. Our Markov chain-based model will be used as a surrogate model [12], which provides an interpretable representation of the underlying decision process and facilitates the explanation of detected anomalies.

Explainability of anomalies. The output of the anomaly detection model may consist of operational and experimental data that are not readily interpretable by users. However, it is essential that scientists retain control over the experimental process and understand the rationale behind anomaly predictions. Moreover, the AI Copilot is intended to assist experimental scientists and facility staff during experiments, rather than burden them with low-level system information that requires additional time and effort to interpret, particularly in time-critical scenarios. Furthermore, enhancing the explainability of the system may also improve user trust and usability by making the model outputs more transparent and easier to interpret.

Among the different explanation paradigms proposed in the XAI literature, counterfactual explanations are particularly well suited to the considered problem. Rather than describing why a sequence is anomalous in terms of abstract feature importance, counterfactual explanations identify the minimal changes required for the sequence to be considered normal [13]. Such explanations provide users with helpful guidance by detecting and explaining the missing transition required to restore the expected operational workflow, and naturally translate into actionable procedural guidance. For example, instead of merely indicating that a sequence is anomalous, the system can explain that inserting the missing determine motor position step between select motor and move motor would result in a valid operational sequence. This form of explanation is intuitive for users, closely aligns with their procedural reasoning, and provides direct guidance for correcting or preventing anomalies.

Case-Based Reasoning (CBR) provides a natural framework for generating such counterfactual explanations. Rather than relying solely on the internal representation of the anomaly detector, a CBR system can retrieve previously observed operational traces that are similar to the current anomalous sequence and use them to construct explanations by analogy. Counterfactual explanations can be obtained by identifying the nearest valid execution trace and highlighting the minimal modifications required to transform the anomalous sequence into a normal one. Recent work has demonstrated the potential of coupling opaque machine learning models with CBR systems to provide factual and counterfactual explanations through transparent retrieval and comparison of representative cases [14, 15]. Building on these ideas, future work will investigate the integration of a CBR explanation layer with the proposed anomaly detection framework, where historical operational sequences will serve as explanatory cases to generate actionable, process-oriented counterfactual explanations for users [16].

Development of RAG The AI Copilot will adopt a case-based Retrieval-Augmented Generation (CBR-RAG) [17] architecture, where historical support tickets maintained by the Data Operations Center constitute the case base. Over the years, a large collection of support tickets has been accumulated, documenting operational problems together with their diagnosis and resolution, representing a valuable case base. When an anomaly is detected, the system retrieves the most similar historical cases based on the current operational context and provides them to the LLM as evidence for generating explanations and recommendations. This approach combines the retrieval and similarity mechanisms of Case-Based

Reasoning with the generative capabilities of LLMs, enabling explanations that are grounded in previous operational experience rather than relying solely on the model's parametric knowledge.

The development of this RAG system may face several challenges. The first concerns the latency of anomaly notifications. Based on the conducted survey, several respondents indicated that alerts delayed by more than a few seconds or minutes may already be ineffective in certain experimental situations, as their impact may have already occurred, potentially requiring parts of the experiment to be repeated. Therefore, maintaining low-latency responses is essential. The second challenge is hallucination and lack of factual faithfulness, where generated responses may not accurately reflect the actual interface [18]. In operational environments, such inaccuracies could confuse experimental scientists and facility staff, potentially leading to incorrect decisions, operational risks, or reduced trust in the AI Copilot over time.

3. Progress Summary

So far, the work has focused on designing and conducting a survey to better understand the main challenges faced by experimental scientists and technical staff during operations at the European XFEL. Based on the survey results, the data acquisition case was selected as the first case of the research.

After that, a device was developed and integrated into the control system to collect operational data from multiple sources. The work then focused on the Graphical User Interface (GUI) data source, where a Markov chain-based model was developed to distinguish between typical and atypical operational behavior. Initial experiments were carried out using synthetic data and produced promising early results. However, additional work is still required to further improve both the methodology and implementation of the proposed approach, as well as to evaluate the potential of case-based reasoning methods, particularly counterfactual reasoning, for enhancing the explainability of the anomaly detection model.

4. Conclusion and Future Work

In conclusion, this research investigates the development of an AI Copilot to support experimental scientists and technical staff during operations at the European XFEL, with data acquisition serving as the initial use case. The work explores how AI-assisted methods, including anomaly detection and retrieval-augmented generation (RAG), can help identify operational issues and support decision-making in complex experimental environments. Initial results from the Markov chain-based anomaly detection approach are promising, although further improvements and testing on actual experimental data are still required, particularly in areas such as explainability and latency. Future work will investigate counterfactual reasoning approaches to improve the explainability of the anomaly detection model outputs, while also continuing the development and evaluation of the anomaly detection model, the RAG system, and the user interface. Overall, the project represents an additional step toward integrating AI-based operational support into scientific workflows at the European XFEL.

Declaration on Generative AI

Generative AI tools were used as language assistants in the preparation of this manuscript, primarily to correct grammatical errors and improve the quality of the English.

References

[1] T. Tschentscher, C. Bressler, J. Grünert, A. Madsen, A. P. Mancuso, M. Meyer, A. Scherz, H. Sinn, U. Zastrau, Photon beam transport and scientific instruments at the European XFEL, Applied Sciences 7 (2017) 592.

[2] T. Hellert, D. Abramov, M. Ajami, E. S. Barnard, W. Bhimji, G. Bhardwaj, T. Chavez, H. Daoud, T. J. Dunn, K. Fagnan, et al., Agentic AI for User Facilities: Workshop Report, 2026.

[3] D. Göries, W. Ehsan, G. Flucke, N. Anakkappalla, V. Bondar, R. Costa, S. Esenov, G. Giovanetti, D. Hickin, I. Karpics, et al., The Karabo SCADA system at the European XFEL, Synchrotron Radiation News 36 (2023) 40–46.

[4] S. Hauf, B. Heisen, S. Aplin, M. Beg, M. Bergemann, V. Bondar, D. Boukhelef, C. Danilevsky, W. Ehsan, S. Essenov, et al., The Karabo distributed control system, Synchrotron Radiation 26 (2019) 1448–1461.

[5] C. Carinan, D. Hammer, M. Bregenholt Jakobsen, P. Schmidt, G. Flucke, D. Göries, I. Bermudez, T. Michelat, E. Sobolev, J. Wrigley, et al., Real-time experiment steering at European XFEL within Karabo, Proceedings of ICALEPCS2025 (2025) 1701–1705.

[6] S. Perrett, A. Fadini, C. D. Hutchison, S. Bhattacharya, C. Morrison, O. Turkot, M. B. Jakobsen, M. Größler, J. Licón-Saláiz, F. Griese, et al., Kilohertz droplet-on-demand serial femtosecond crystallography at the European XFEL station FXE, Structural dynamics 11 (2024).

[7] W. Decking, S. Abeghyan, P. Abramian, A. Abramsky, A. Aguirre, C. Albrecht, P. Alou, M. Altarelli, P. Altmann, K. Amyan, et al., A MHz-repetition-rate hard X-ray free-electron laser driven by a superconducting linear accelerator, Nature photonics 14 (2020) 391–397.

[8] S. Hauf, A. Silenzi, A. Klimovskaia, D. Göries, D. Boukhelef, G. Flucke, J. L. Vázquez-García, K. Wrona, N. Al-Qudami, P. Schmidt, et al., Lessons learned refactoring the EuXFEL's central data acquisition system, Proceedings of ICALEPCS2025 (2025) 879–884.

[9] M. Smith, G. Giovanetti, S. Hauf, I. Karpics, A. Parenti, A. Samadli, L. Samoylova, A. Silenzi, F. Sohn, P. Zalden, et al., Applying Model Predictive Control to Regulate Thermal Stability of a Hard X-ray Monochromator Using the Karabo SCADA Framework (2022).

[10] T. Lu, L. Wang, X. Zhao, Review of anomaly detection algorithms for data streams, Applied Sciences 13 (2023) 6353.

[11] M. Salehi, L. Rashidi, A survey on anomaly detection in evolving data: [with application to forest fire risk prediction], ACM SIGKDD Explorations Newsletter 20 (2018) 13–23.

[12] P. Saves, P. S. Palar, M. D. Robani, N. Verstaevel, M. Garouani, J. Aligon, B. Gaudou, K. Shimoyama, J. Morlier, Surrogate modeling and explainable artificial intelligence for complex systems: A workflow for automated simulation exploration, arXiv preprint arXiv:2510.16742 (2025).

[13] M. T. Keane, E. M. Kenny, E. Delaney, B. Smyth, If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual XAI techniques, arXiv preprint arXiv:2103.01035 (2021).

[14] M. T. Keane, E. M. Kenny, How case-based reasoning explains neural networks: A theoretical analysis of xai using post-hoc explanation-by-example from a survey of ann-cbr twin-systems, in: International conference on case-based reasoning, Springer, 2019, pp. 155–171.

[15] M. T. Keane, E. M. Kenny, M. Temraz, D. Greene, B. Smyth, Twin systems for deepcbr: A menagerie of deep learning and case-based reasoning pairings for explanation and data augmentation, arXiv preprint arXiv:2104.14461 (2021).

[16] A. Buliga, C. Di Francescomarino, C. Ghidini, F. M. Maggi, Counterfactuals and ways to build them: Evaluating approaches in predictive process monitoring, in: International Conference on Advanced Information Systems Engineering, Springer, 2023, pp. 558–574.

[17] N. Wiratunga, R. Abeyratne, L. Jayawardena, K. Martin, S. Massie, I. Nkisi-Orji, R. Weerasinghe, A. Liret, B. Fleisch, CBR-RAG: case-based reasoning for retrieval augmented generation in LLMs for legal question answering, in: International Conference on Case-Based Reasoning, Springer, 2024, pp. 445–460.

[18] N. Wiratunga, V. A. Wijayasekara, I. Nkisi-Orji, P. Salimi, K. Martin, C. Bolaños, iCARE: Ontology-Guided Intent Routing for Multi-Agent LLM-Based Dialogue Systems, context (including current conversation goals) 5 (2025) 6.