

AI AND ML INTEGRATION FOR BEAMLINE OPTIMIZATION AND VIRTUAL ASSISTANCE AT THE SOLARIS SYNCHROTRON

M. Szczepanik[†], M. Piekarski, M. Fałowski

National Synchrotron Radiation Centre SOLARIS, Kraków, Poland

Abstract

This paper presents two complementary initiatives currently under consideration at the SOLARIS synchrotron facility, aimed at enhancing beamline operations and user support through artificial intelligence (AI) and machine learning (ML). The first initiative explores the potential of ML-based mirror alignment and beam stabilization, while the second investigates the development of a virtual assistant powered by large language models (LLMs). Both approaches are in the conceptual phase and leverage existing control infrastructure and modern AI frameworks to improve experimental efficiency, reproducibility, and accessibility.

INTRODUCTION

The SOLARIS synchrotron, located in Kraków, Poland, is a national research infrastructure dedicated to advanced photon science. As synchrotron facilities evolve toward higher brilliance and more complex experimental configurations, the need for intelligent, adaptive control systems becomes increasingly critical. This paper presents two technical solutions for integrating AI and ML into SOLARIS operations: automated beamline optimization and a virtual assistant for scientific and operational support.

MACHINE LEARNING-BASED BEAM OPTIMIZATION

The beamlines at SOLARIS are equipped with multiple mirrors of varying geometries and monochromators, forming a complex optical path that requires precise alignment. The control system is based on the Tango Controls framework, with Sardana used for high-level experiment orchestration. All control software runs on AlmaLinux 9, ensuring compatibility with modern scientific computing libraries and ML frameworks. The following ML-based solution is currently being evaluated and has not yet been implemented at SOLARIS.

The ML-based solution addresses two distinct but interrelated optimization goals. The first is the enhancement of beam quality, defined by maximum intensity, tight spatial focus, and optimal energy resolution for a given photon energy. This is achieved by analyzing signals from a soft X-ray photodiode, which serves as a proxy for beam intensity and spectral purity. The second goal is the stabilization of the beam position on the sample, ensuring that the beam consistently hits the same location regardless of energy changes. This is monitored using the induced current on the exit slit blades, which reflects the beam's spatial alignment.

Due to the physical constraints of the beamline, simultaneous acquisition of photodiode and slit current signals is

not feasible. Therefore, the system will employ two separate ML models. The first model will be trained on data collected during photodiode-based measurements and will predict mirror and monochromator adjustments to maximize beam quality. The second model will use slit current data to maintain beam position, compensating for energy-dependent shifts in the optical path. Both models will be updated online using data collected during routine measurements, allowing them to adapt to changing beamline conditions without manual intervention.

Recent research supports the feasibility and effectiveness of such approaches. Morris et al. [1] demonstrated a general Bayesian optimization framework for autonomous beamline alignment, capable of handling high-dimensional, nonlinear parameter spaces with minimal prior information. Similarly, Morris et al. [2] applied online learning techniques to optimize beamline parameters at NSLS-II, showing that real-time ML can outperform traditional alignment methods in both speed and precision. Rebuffi et al. [3] developed an AI-driven system for autonomous alignment of nanofocusing mirror systems, highlighting the role of digital twins and wavefront sensing in achieving sub-micron precision. Xie et al. [4] proposed a deep learning method for estimating alignment errors in Kirkpatrick-Baez mirror systems, demonstrating that convolutional neural networks can extract meaningful features from speckle-modulated X-ray images. Finally, Yazak [5] reported on the integration of AI agents at NSLS-II for real-time anomaly detection and experiment monitoring, underscoring the broader trend toward autonomous synchrotron operations.

The implementation of machine learning for beam optimization at the SOLARIS synchrotron is supposed to integrate seamlessly with the existing control infrastructure, which is based on Tango Controls and Sardana, running on AlmaLinux 9. The system will be developed in Python 3.9, ensuring compatibility with both the control framework and modern machine learning libraries such as scikit-learn, PyTorch, and NumPy.

The proposed architecture consists of two independent machine learning models, each addressing a distinct aspect of beam optimization. The first model is responsible for maximizing beam quality, defined by intensity, focus, and energy resolution. It operates on data acquired from a soft X-ray photodiode and is trained to predict optimal mirror and monochromator settings for a given photon energy. The second model ensures spatial stability of the beam on the sample, compensating for energy-dependent shifts in the optical path. It uses data from the induced current on the exit slit blades to maintain consistent beam positioning.

Due to the physical constraints of the beamline, simultaneous acquisition of photodiode and slit current signals is

not possible. Therefore, the models will be trained and executed independently, each within its own Tango device server. These servers expose attributes and commands that allow Sardana macros to interact with the models during measurement cycles. Each macro is responsible for collecting relevant diagnostic data, invoking the appropriate model, applying the predicted corrections to the optical elements, and logging the results for future training.

The training process is hybrid, combining offline initialization with online incremental updates. Historical data from beamline operations is used to bootstrap the models, while new data collected during routine measurements is used to refine them. For the beam quality model, training data includes photon energy, mirror positions, monochromator settings, and photodiode signal. For the beam position model, the dataset includes energy, mirror positions, and slit current values. Data is stored in structured formats such as CSV or HDF5, and managed using Python-based data pipelines.

Model inference is performed in real time during measurement sequences. Sardana macros are configured to alternate between beam quality and beam position optimization, depending on the measurement context. This modular approach allows the system to adapt to different experimental requirements without interfering with the underlying control logic. The ML models are encapsulated in Python classes that implement standardized interfaces for prediction and training, facilitating integration with Tango device servers.

To ensure robustness and maintainability, the system includes logging, exception handling, and diagnostic tools. Logs are written in a format compatible with existing Tango archiving solutions, and diagnostic plots can be generated using Matplotlib to visualize model performance over time. The system is designed to be extensible, allowing future incorporation of additional diagnostics such as wavefront sensors or camera-based feedback.

This implementation represents a significant advancement in the automation of beamline operations at SOLARIS. By leveraging machine learning within the existing control framework, the system enhances beam stability and quality, reduces manual intervention, and enables more complex experimental protocols. Future work will focus on validating the models across multiple beamlines, optimizing training strategies, and exploring reinforcement learning approaches for long-term adaptive control.

VIRTUAL ASSISTANT FOR SCIENTIFIC AND OPERATIONAL SUPPORT

Modern synchrotron facilities face increasing complexity in both instrumentation and data management. Beamline scientists and users must navigate diverse software environments, including control systems like Sardana and Tango Controls, and proprietary applications for detectors and analysis. At SOLARIS, the need for intuitive, intelligent support systems has led to the exploration of virtual

assistants based on LLMs. Recent prototypes such as VISION at Brookhaven National Laboratory [6] demonstrate the feasibility of natural language interfaces for beamline control and experiment guidance. These systems aim to reduce the cognitive load on users and staff, improve reproducibility, and democratize access to advanced instrumentation. The virtual assistant described below is a proposed concept and has not yet been deployed at SOLARIS.

SOLARIS seeks to develop a virtual assistant that will assist scientists in configuring and running experiments on beamlines, provide technical staff with quick access to manuals, procedures, and troubleshooting guides, support administrative personnel with procurement regulations, safety protocols, and internal documentation, and enable natural language interaction with control systems and documentation repositories.

SOLARIS uses Sardana, a SCADA system built on Tango Controls, for beamline supervision and data acquisition. These systems are widely adopted in facilities such as ALBA, DESY, SOLEIL, and MAX IV, forming a robust ecosystem for scientific instrumentation. Sardana provides a Python-based macro environment and a CLI (Spock), which can be interfaced with external agents. This makes it suitable for integration with LLM-based assistants via frameworks like LangChain or FastAPI. Machine learning is increasingly used for beamline optimization, autonomous experimentation, and data analysis. Facilities such as SLAC and CHESS have demonstrated AI-driven workflows for high-throughput experiments and accelerator tuning [7].

PLLuM (Polish Large Language Universal Model) [8, 9] is a family of open-source Polish language models developed by a consortium of leading scientific institutions in Poland, including the Wrocław University of Science and Technology, NASK, the Institute of Computer Science of the Polish Academy of Sciences, and others. The model was trained on a massive corpus of over 150 billion Polish-language tokens, including administrative, scientific, and technical texts. It supports parameter sizes ranging from 8 to 70 billion, offering scalability and precision in Polish-language generation. Technologically, PLLuM is built on transformer architectures such as Llama 3.1 and Mixtral, and refined through instruction tuning, preference learning, and alignment techniques. It includes a manually curated set of 40,000 Polish prompt-response pairs and a preference corpus evaluated by human annotators to ensure factual correctness, linguistic quality, and safety. PLLuM supports Retrieval-Augmented Generation (RAG), allowing it to integrate with external knowledge bases and document repositories. This is particularly relevant for SOLARIS, where the assistant must access manuals, regulations, and experimental logs. The model has demonstrated state-of-the-art performance in Polish-language tasks, especially in public administration and legal domains.

To make PLLuM effective in the SOLARIS environment, several steps must be undertaken. First, the model must be fine-tuned on domain-specific data, including experimental procedures, device manuals, Tango/Sardana macros, and internal documentation. This requires curating

a high-quality corpus of texts relevant to synchrotron science and operations. Second, a RAG pipeline must be implemented using tools like LlamaIndex or Haystack. This will allow the assistant to retrieve relevant documents and answer queries based on up-to-date information. The document base should include safety protocols, procurement regulations, and technical specifications. Third, integration with control systems must be established. Using PyTango and Sardana's macro server, the assistant can translate user queries into executable commands. Finally, the assistant must be deployed in a secure, user-friendly environment. JupyterLab can serve as an interface for scientists, while a web-based chatbot can support administrative staff. Authentication via LDAP and role-based access control will ensure data security and compliance.

SOLARIS plans to develop a prototype assistant for one beamline, evaluate its performance in real-world scenarios, and expand its capabilities based on user feedback. Collaboration with the PLLuM consortium and other synchrotron facilities will be essential to share best practices and accelerate development.

CONCLUSION

The integration of machine learning and large language models at the SOLARIS synchrotron marks a significant step toward intelligent, autonomous beamline operations and user support. By combining real-time optimization with virtual assistance, the facility enhances experimental precision, accessibility, and operational efficiency. These initiatives not only align with global trends in synchrotron automation but also position SOLARIS as a forward-thinking center for advanced photon science. Future developments will focus on scaling these solutions across beamlines and refining AI models through continuous feedback and collaboration.

REFERENCES

- [1] T. W. Morris *et al.*, "A general Bayesian algorithm for the autonomous alignment of beamlines," *J Synchrotron Radiat*, vol. 31, no. 6, pp. 1446–1456, Nov. 2024.
doi: 10.1107/S1600577524008993
- [2] T. Morris *et al.*, "On-the-fly optimization of synchrotron beamlines using machine learning," in *Optical System Alignment, Tolerancing, and Verification XIV*, J. Sasián and R. N. Youngworth, Eds., SPIE, Oct. 2022, p. 22.
doi: 10.1117/12.2644996
- [3] L. Rebuffi *et al.*, "AutoFocus: AI-driven alignment of nanofocusing X-ray mirror systems," *Opt Express*, vol. 31, no. 24, p. 39514, Nov. 2023.
doi:10.1364/OE.505289
- [4] J.-N. Xie *et al.*, "Deep learning for estimation of Kirkpatrick–Baez mirror alignment errors," *Nucl. Sci. Tech.*, vol. 34, no. 8, p. 122, Aug. 2023.
doi:10.1007/s41365-023-01282-4
- [5] D. Yazak, "AI/ML's bright future at NSLS-II," BNL Newsroom, Brookhaven National Laboratory, <https://www.bnl.gov/newsroom/news.php?a=222328>
- [6] S. Mathur, N. van der Vleuten, K. G. Yager, and E. H. R. Tsai, "VISION: a modular AI assistant for natural human-instrument interaction at scientific user facilities," *Mach. Learn.: Sci. Technol.*, vol. 6, no. 2, p. 025051, Jun. 2025.
doi:10.1088/2632-2153/add9e4
- [7] A. Mishra *et al.*, "A start to end machine learning approach to maximize scientific throughput from the LCLS-II-HE," May 2025, arXiv:2505.23858.
doi:10.48550/arXiv.2505.23858
- [8] PLLuM: Polish large language model – artificial intelligence, <https://primotly.com/article/pllum-polish-large-language-model-artificial-intelligence>
- [9] PLLuM: A family of Polish large language models," <https://huggingface.co/CYFRAGOVPL/PLLuM-12B-base>