ISBN: 978-3-95450-219-6 ISSN: 2226-0366 doi: 10.18429/JACoW-LINAC2024-THPB040

LLM INTEGRATION INTO EPICS

E. Adams, A. Sobhani, Oak Ridge National Laboratory, TN, USA

Abstract

The last two years have seen a drastic rise in the usage of Large Language Models (commonly known as LLMs) including ChatGPT, across multiple industries and applications. This has firmly established them as broadly applicable tools, proving their flexibility and power to understand and generate human-like text. In this paper, we explore the integration of an LLM into the EPICS system. The integration will focus on using the LLM for state-of-the-art image processing and spatial analysis of the images from the beamlines. We aim to increase the accuracy and efficiency of image interpretation, enabling more precise data analysis and decision-making within EPICS through the services offered by this LLM. This integration best showcases the various potentials of LLMs in science and industry, setting the stage for future advancements in automated control systems.

INTRODUCTION

EPICS is a software system that has grown out of scientific-industrial cooperation, aiming to satisfy the emerging needs for the development of distributed control systems. Generally, EPICS has been designed to cope with the complex control requirements of large experimental setups, such as particle accelerators, telescopes, and other highly sophisticated scientific instruments. Its robust architecture and scalability have made it the natural choice for such applications. Moreover, it provides flexibility and efficiency in managing the large sets of data produced in experiments.

The core building blocks of the EPICS control system are Input/Output Controllers (IOCs). IOCs interface directly with the hardware and address tasks like data acquisition, device control, and real-time processing. They are responsible for running the devices under their control, executing commands, and gathering data from sensors and instruments. This decentralized approach offers better scalability and fault tolerance: each IOC operates independently while communicating with other IOCs and higher-level applications through a well-defined protocol.

The recent breakthrough in large language models, such as ChatGPT, has enabled numerous possibilities for enhancing control systems. Large Language Models (LLMs) have tremendously succeeded in understanding human-like text and handling complex analysis tasks, and they can easily automate many different tasks. Clearly, integrating LLMs with the EPICS system would provide real benefits in image processing. These functionalities are key to understanding the huge volumes of visual data that come from today's beamlines, where exact and efficient analysis forms the foundation of the scientific approach.

This paper discusses the integration of an LLM into the EPICS framework, specifically for image processing of beamline images at the VULCAN beamline at SNS. Building on the advanced computational capabilities enabled by LLMs, we plan to accomplish various improvements to the full ease of use of the control system.

SYSTEM DESIGN

The system is entirely written in Python, forming a robust yet flexible core for integration into EPICS. It is implemented as a PCASpy [1] IOC, which allows for seamless integration with the existing EPICS infrastructure. The input images are obtained using PyEpics [2], processed through the LLM, and the output images are then hosted with PCASpy, making them accessible to other components of the control system.

This setup includes the implementation of a Control System Studio (CSS) page to facilitate user interaction with the software's Process Variables (PVs), providing an easy way of controlling and monitoring the system (Fig. 1).

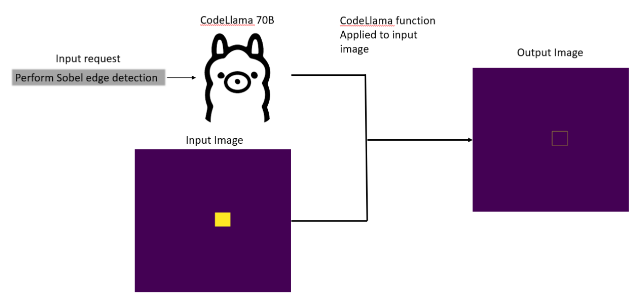

Figure 1: System Architecture Diagram showing the interaction between EPICS, CSS, and LLMs.

Upon entering the CSS page, users are presented with two major functions: image processing and spatial analysis. These options cover most of the possible analytical needs, ensuring comprehensive capabilities in data interpretation and manipulation.

The image processing function allows users to apply various transformations to input images. This includes operations such as thresholding, edge detection, image arithmetic, and Gaussian blurring. These transformations are executed by the LLM, and the processed images are then made available as PVs for other EPICS components to access.

The spatial analysis function extracts useful information directly from the input images. For example, it can determine the coordinates of centroids or other features of interest. This data is stored in PVs, enabling other parts of the EPICS system to leverage the insights provided by the LLM-powered analysis.

Additionally, the system offers an online/offline mode option. This allows users to choose how the system processes requests based on their operational context and available resources.

ISBN: 978-3-95450-219-6 ISSN: 2226-0366 doi: 10.18429/JACoW-LINAC2024-THPB040

In online mode, the system connects to ChatGPT via the OpenAI API [3]. This configuration allows the required code to be generated automatically and returned as a function. This ensures high-quality and efficient code generation, meeting all user requirements with the advanced capabilities available through ChatGPT.

In offline mode, the system uses a local instance of the large language model CodeLlama [4]. This model runs on a VM equipped with an NVIDIA Tesla V100S 32GB GPU to avoid any performance degradation of the beamline computer. By offloading the computational load to this powerful GPU, the system is optimized in its performance and responsiveness, thereby maintaining the efficiency of beamline operations.

Some techniques tested within the system include image thresholding, edge detection, image subtraction and addition, image multiplication, centroid detection, and Gaussian blur. All these techniques have been thoroughly evaluated to ensure their effectiveness and compatibility with CodeLlama, offering a comprehensive solution for advanced image analysis.

INTEGRATION AT VULCAN AND FUTURE APPLICATIONS

The system has been successfully been integrated at the VULCAN beamline at the Spallation Neutron Source (SNS). This integration demonstrates the practical applicability of LLMs in a real-world scientific environment (see Fig. 2). The current implementation focuses on image processing tasks, but the flexibility of the system allows for potential expansion into other areas of data analysis.

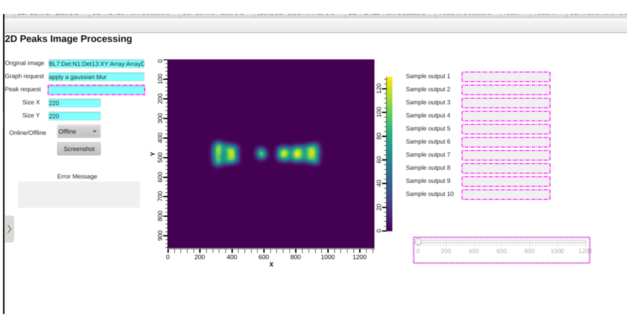

Figure 2: CSS User Interface at VULCAN showing a Gaussian blur.

Looking towards the future, the system could be fine-tuned towards applications such as fitting of d-space peaks, rietveld refinement, and fourier transform analysis in Nuclear Magnetic Resonance.

CONCLUSION

The integration of large language models into the EPICS framework through this Python-based PCASpy IOC represents a significant advancement in control systems technology, particularly as demonstrated at the VULCAN beamline at SNS. The combination of user-friendly interfaces, flexible processing modes, and a robust set of image processing techniques underscores the potential for enhancing data interpretation and decision-making in scientific and industrial applications. As we continue to explore and expand the capabilities of this system, we anticipate even greater improvements in the efficiency and accuracy of beamline operations and data analysis.

REFERENCES

[1] Xiaoqiang Wang, "PCASpy: Portable Channel Access Server in Python", `https://github.com/paulscherrerinstitute/pcaspy`

[2] Matthew Newville, "PyEpics: EPICS Channel Access for Python", `https://github.com/pyepics/pyepics`

[3] OpenAI, "OpenAI API Manual", `https://platform.openai.com/docs/api-reference/introduction`

[4] Meta Llama CodeLlama, `https://github.com/meta-llama/codellama`