

Harness the Power of AI and CI/CD to Fuel Scientific Discovery

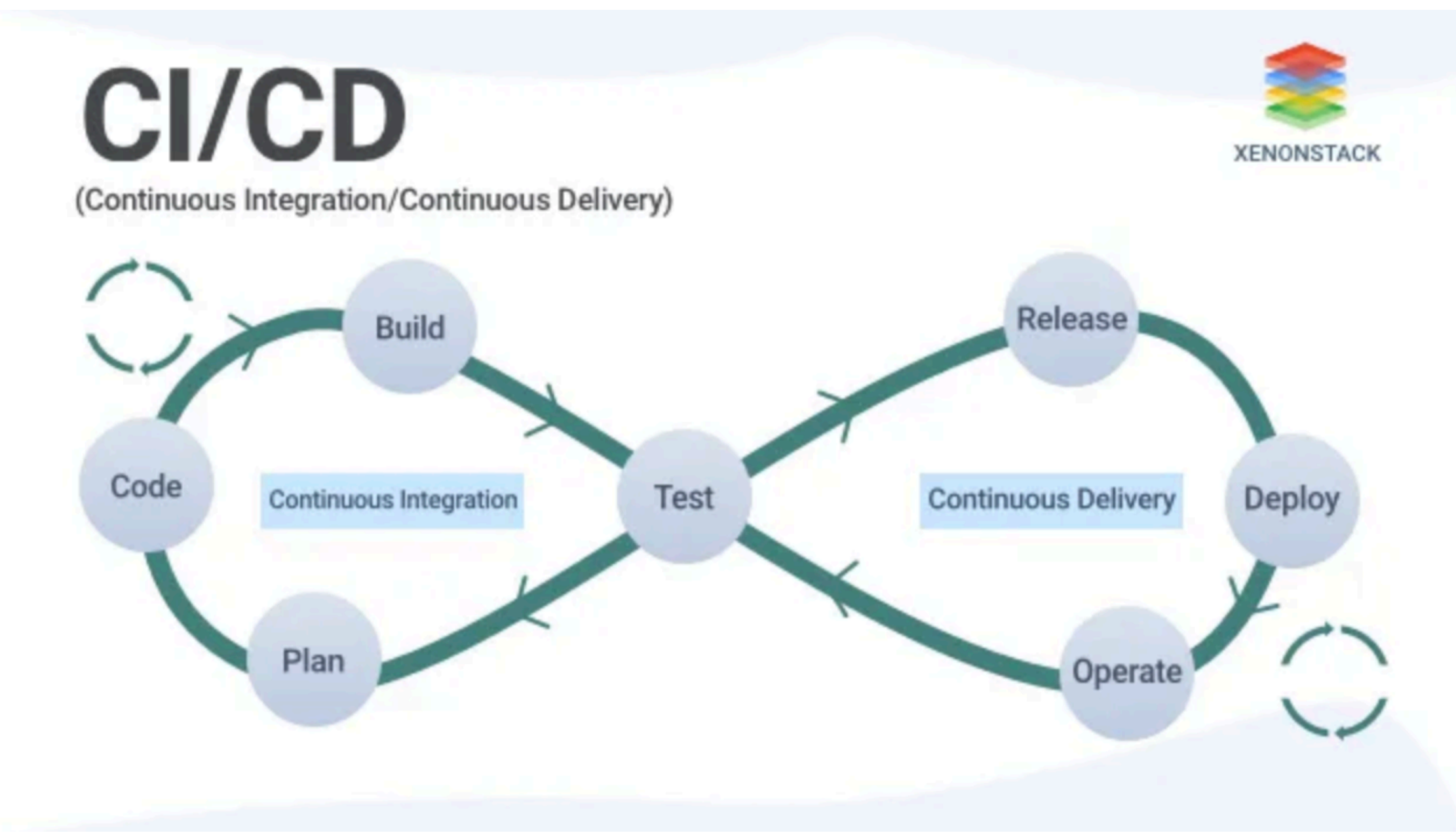
Tea Nurcellari, *New York University*, Under the Mentorship of Namratha Urs and the GlideinWMS Team

Security and Emergency Management Division, Cybersecurity Team, Fermi National Accelerator Laboratory, Batavia, Illinois 60510 FERMILAB-POSTER-24-0159-V

Introduction

The goal of this project is to improve and protect scientific computing systems used in significant experiments like DUNE at Fermilab and CMS at the LHC by leveraging AI and CI/CD technologies. Our goal is to develop and automate solutions that will enhance the functionality of two essential systems in the scientific computing infrastructure: GlideinWMS and HEPCloud.

The ability to build containerized CI/CD pipelines, use AI to enhance code quality and coverage, and automate security checks will all be taught to interns. The project will also give participants practical experience in configuring, debugging, and optimizing distributed computing systems utilizing Grid, Cloud, and HPC resources. Working with professionals and other interns will provide a special chance to actively support open-source development and scientific achievements.



GlideinWMS

GlideinWMS (Glidein Workflow Management System) is a pilot-based workload management system used extensively in large-scale scientific experiments like CMS at LHC. It facilitates the efficient distribution of computing jobs across diverse resources, ensuring optimal use of available computing power. By working on this project, you will delve into creating containerized CI/CD pipelines to streamline the deployment and management of GlideinWMS components. This includes leveraging AI to improve code coverage and quality, identifying and fixing potential issues before they impact production, and automating security verifications to safeguard the system against vulnerabilities.



HEPCloud

HEPCloud (High Energy Physics Cloud) extends the traditional computing resources of high-energy physics experiments into the cloud, providing a flexible and scalable environment for data processing. You will learn to enhance HEPCloud's operations by integrating advanced CI/CD practices, ensuring seamless software releases with integrated feedback loops. This involves using AI tools to analyze code performance and detect areas for improvement, automating the process of code validation, and implementing security checks to protect the integrity of the data and applications. Strengthen code security through automated testing and continuous integration practices. Develop proficiency in collaborative development tools such as Git, Jenkins, and GitLab, which are essential for modern software development and deployment. By mastering these technologies and methodologies, you will play a crucial role in enhancing the efficiency, reliability, and security of scientific computing infrastructures that drive significant scientific discoveries.

Artificial Intelligence (AI) Tools Used

ChatGPT -

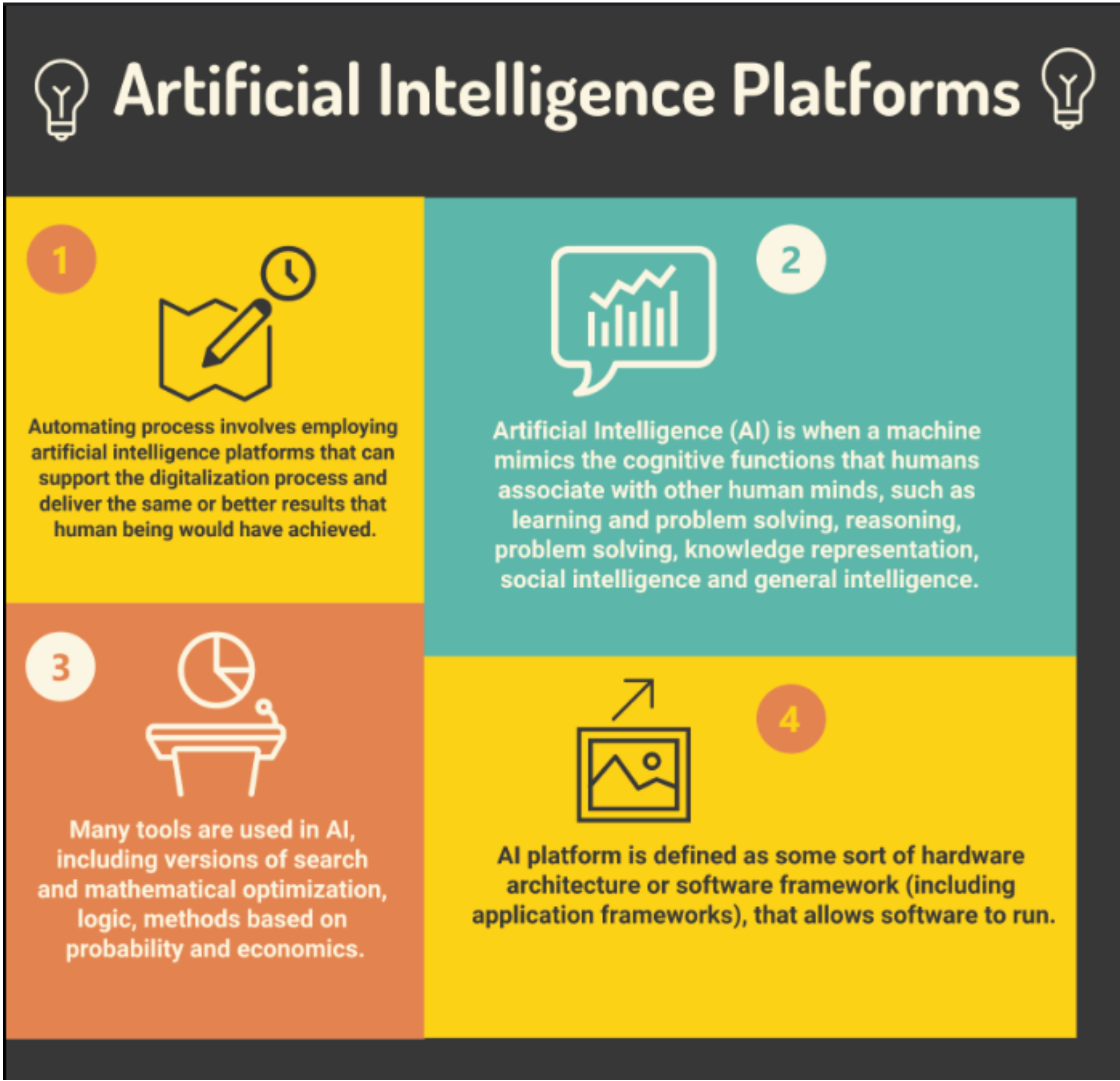
ChatGPT offers extensive support for our project by providing step-by-step guidance on setting up and optimizing CI/CD pipelines and managing containerization with Docker and Kubernetes. It explains specific components and best practices for GlideinWMS and HEPCloud, helping enhance and automate these systems. ChatGPT recommends AI tools for improving code coverage and quality, suggests code refactoring for better performance and security, and guides automating security checks with tools for code analysis and vulnerability scanning. It aids in effective use of version control with Git, setting up collaborative development tools, and maintaining comprehensive documentation. Additionally, it offers resources for learning container orchestration, CI/CD, AI for software development, and secure coding practices. ChatGPT also helps organize and prioritize tasks using agile methodologies and suggests integrating feedback into the development workflow for continuous improvement.

Amazon Code Whisperer -

Amazon CodeWhisperer offers real-time code suggestions and completions, helping you write code faster and with fewer errors. It generates boilerplate code for setting up CI/CD pipelines, container orchestration, and other routine tasks, reducing manual effort. CodeWhisperer provides AI-driven recommendations to improve code quality, including better practices for efficiency, readability, and maintainability, and offers suggestions for refactoring code to enhance performance and security. It suggests security best practices while coding and integrates automated security checks into your development workflow. CodeWhisperer provides contextual code examples and documentation to help you learn new libraries, frameworks, and best practices, as well as real-time assistance with debugging and error resolution. It ensures consistent code style and standards across your team and enhances pair programming sessions with instant code suggestions. Finally, CodeWhisperer seamlessly integrates with your existing development environment and enhances CI/CD pipelines by automating code checks and deployments for more efficient and reliable processes.

Conclusion

The "Harness the Power of AI and CI/CD to Fuel Scientific Discovery" project offers a comprehensive and hands-on experience in advanced scientific computing. By enhancing and automating critical systems like GlideinWMS and HEPCloud, participants gain valuable skills in container orchestration, AI-driven code improvement, and security automation. This project not only equips you with practical technical expertise but also allows you to contribute meaningfully to groundbreaking scientific experiments. Through collaboration with industry experts and fellow interns, you will play a vital role in driving forward the capabilities of scientific research and open-source development.



This research was supported in part by the U.S. Department of Energy (DOE), Omni Technology Alliance Internship Program. The program is championed by the DOE's Office of Chief Information Officer (OCIO) and represents a partnership with the leadership of the Office of Economic Impact and Diversity, the Office of Science, the Office of Nuclear Energy, and the National Nuclear Security Agency. The program is administered by the Oak Ridge Institute for Science and Education

This work was produced by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy. Publisher acknowledges the U.S. Government license to provide public access under the DOE Public Access Plan DOE Public Access Plan