# Managing MLOps Projects

Alankrit Nayak
*Master Software Technology*
*Hochschule Fur Technik*
Stuttgart, Germany
alankritnayak99@gmail.com

***Abstract*—Machine Learning Operations (MLOps) is a new field that combines machine learning (ML) with DevOps to make it easier to launch, manage, and keep an eye on ML models. We will take a look at different parts of MLOps, such as its workflow, tools, and methods. We will talk about the difficulties of handling machine learning and big data projects, especially when there are multiple organisations involved, and how agile methods can help with these difficulties. The fairness, explainability, accountability, and sustainability principles can be used in MLOps techniques to make ML deployments more reliable and effective. We will also compares MLOps to regular DevOps and stress on the unique needs for working with ML models. The goal of this paper is to give a full picture of MLOps along with overview of Managing MLOps projects, as well as future directions in the field.**

## I. Overview and Definition of MLOps

The goal of Machine Learning Operations (MLOps) is to make putting machine learning (ML) models into production easier and more automated. Machine Learning Ops includes best practices, a set of ideas, and a way of developing that focuses on making ML products more automated and useful. Even though MLOps is becoming more important, the term is still not very clear. Different meanings lead to misunderstandings and bad communication.

- MLOps takes a complete method to bridging the gap between developing ML models and putting them into production.
- This makes sure that ML systems are scalable, repeatable, and easy to maintain.
- Continuous integration and continuous delivery (CI/CD), versioning, workflow orchestration, and ongoing monitoring are all parts of MLOps.

[3]

## II. MLOps and Project Management

The current corporate environment is characterised by rapid changes, and the widespread adoption of machine learning (ML) and big data analytics has brought about a new period of innovation and potential. Companies in several sectors are utilising data-driven technology to acquire knowledge, enhance operations, and provide benefits to their consumers. Nevertheless, because to the rapid increase in data volumes and the growing intricacy of machine learning algorithms, organisations have distinct issues when it comes to managing initiatives in these fields. Conventional project management approaches, which involve sequential planning and inflexible procedures, frequently have difficulties in adapting to the dynamic and unpredictable nature of machine learning and big data initiatives. Consequently, organisations encounter delays, excessive expenses, and less-than-ideal results, which impede their capacity to take use of the revolutionary capabilities of data-driven technology.

- Agile techniques are a promising solution in this situation, as they provide a flexible and iterative approach to project management that is well-suited to the intricacies of machine learning (ML) and big data projects.

[10].

## III. The Growing Importance of Managing MLOps Projects

Machine Learning Operations (MLOps) is attracting increasing interest due to the widespread availability of software components and middleware that facilitate many elements of ML, ranging from AutoML to specialised ML engineering tasks like explainability and interpretability. The growing interest in MLOps highlights the crucial role it plays in integrating machine learning into regular business operations. Nevertheless, the intricate nature of these systems can render them unviable from a societal, technological, and administrative standpoint.

- Managing these intricacies in a proficient manner, is crucial to guarantee that AI software maintains its efficacy and sustainability in the long run.

[7]

## IV. The Need for Specialized DevOps in ML

Ensuring prompt, dependable, and consistent software delivery is crucial for every organisation, particularly as artificial intelligence (AI) becomes increasingly popular in diverse applications and usage scenarios. Specialised techniques are necessary to ensure quality due to the intricacy of constructing machine learning (ML) models.

- MLOps is the adaptation of traditional DevOps approaches, which concentrate on the continuous integration and continuous delivery (CI/CD) of software, to tackle the distinct issues encountered in ML development and deployment.
- MLOps combines the principles of DevOps with the special requirements of machine learning.
- It helps to guarantee the ability to handle large workloads, reproduce results, and maintain ML systems effectively.

[8]

## V. Explainability, Fairness, Accountability, and Sustainability in MLOps

There are four main principles of sustainable MLOps, which are linked to each other. These principles are:

- Fairness : Being open and honest is important for making sure that everyone is treated fairly and that AI processes produce fair results
- Explainability : The ability of an AI system to give clear and understandable details about how it makes decisions is called ”explainability.”
- Responsibility : means that the AI system and the people who run it are responsible for what it does and decides.
- Sustainability : Maintaining high standards of transparency helps organisations deal with and fix any problems that come up, which supports the AI system's long-term viability.
- These concepts work well together to make a strong base for creating and keeping MLOps practices that are fair, open, and responsible.

. [7]


Fig. 1. Interconnections between Explainability, Fairness, Accountability, and Sustainability in MLOps

[7]

## VI. Understandability of MLOps System Architectures

Understanding MLOps system architectures might be difficult due to the intricacies involved. Informal explanations and illustrations frequently encounter issues such as discrepancies, absence of standardisation, and differing levels of detail, which can impede understanding. The use of semi-formal diagrams, namely those based on UML, greatly improves the comprehensibility of these systems. These diagrams enhance task accuracy among professionals without lengthening work completion time, indicating that they offer clarity without introducing intricacy. Implementing semi-formal representations in MLOps can significantly improve communication and comprehension among stakeholders. [1]

## VII. MLOps Lifecycle and Tools

The lifecycle of an ML model in an MLOps framework involves several stages and utilizes various tools to manage complexities effectively. These stages include:

label=0.

1) **Data Collection and Preprocessing**:
    - Initial step in the MLOps lifecycle.
    - Involves gathering and preparing data for model training.
2) **Feature Engineering**:
    - Process of selecting and transforming variables to improve model performance.
    - Critical for enhancing the accuracy of ML models.
3) **Model Training**:
    - Involves using algorithms to train the model on the prepared data.
    - Utilizes significant computational resources such as GPUs and TPUs.
4) **Model Validation**:
    - Ensures that the trained model performs well on unseen data.
    - Important for verifying the model's accuracy and generalizability.
5) **Model Deployment**:
    - Deploying the validated model into a production environment.
    - Requires robust infrastructure to handle real-time predictions.
6) **Continuous Monitoring and Maintenance**:
    - Monitoring the model's performance in production.
    - Involves retraining and updating the model as needed to maintain accuracy.
7) **Tools and Platforms**:
    - **Kubeflow**:
        - Provides a comprehensive platform for developing, orchestrating, deploying, and running scalable and portable ML workloads.
    - **MLflow**:
        - Offers tools for tracking experiments, packaging code into reproducible runs, and sharing and deploying models.
        - Helps manage the complexities of ML pipelines.

[9]

## VIII. MLOps vs. DevOps

MLOps is commonly perceived as an expansion of DevOps, distinguished by the inclusion of model deployment-specific elements that are exclusive to machine learning. DevOps primarily emphasises the seamless integration and deployment of software through continuous integration and continuous deployment (CI/CD). On the other hand, MLOps encompasses more intricate aspects, like model versioning, continuous training, and model monitoring. The complexity stem from the non-deterministic nature of machine learning models and the need to manage enormous and constantly changing information. ML models necessitate frequent retraining and updating to accommodate new data and uphold correctness, hence introducing additional complexity to the deployment process compared to conventional software. In addition, MLOps must tackle data management challenges such as data quality, privacy, and governance, which are not as prominent in DevOps. Integrating MLOps principles in organisations necessitates a complete approach to continuous integration. This needs collaboration across multiple teams and the installation of strong governance structures to ensure compliance with ethical and legal norms. Although there are difficulties, implementing MLOps principles can greatly improve the dependability and effectiveness of deploying machine learning models in production settings, making it an essential advancement of DevOps for AI-powered applications. [5]

## IX. Integration Challenges in Multi-Organization MLOps

Diverse datasets make MLOps deployment in multi-organization scenarios problematic. Organisations store data in many formats and systems, making data consolidation and quality assurance difficult. Legal and regulatory restrictions complicate data management, especially for healthcare data. These restrictions can hinder cross-organizational data sharing and use. To solve these problems and ensure dataset stability and usability for machine learning model training, efficient data integration strategies are essential.

Strong APIs and uniform data formats are needed to overcome integration challenges. These components provide smooth connectivity across an organization's systems. Integration patterns created for AI/ML applications are needed to secure, preserve integrity, and efficiently transmit data across borders. This entails using data lakes to combine multiple data sources and efficiently handle vast amounts of data while retaining data ownership. Due to privacy concerns, organisations must use innovative methods to study and use their data without direct access. These concepts divide data processing, model creation, and operational deployment to foster MLOps collaboration. [2].

## X. Model Management and Monitoring

Successful MLOps require efficient model management, especially in multi-organizational settings. To comply with data privacy rules, models must be carefully created, verified, and implemented in protected settings. The model should be designed to easily update and retrain with new data while adhering to laws. After deployment, machine learning models must be monitored for issues like model drift and biases that could affect prediction accuracy. Monitoring should include ML-specific parameters like model accuracy, data drift, and bias detection in addition to software metrics.

Stakeholders need transparent and accountable governance to understand data use and provide model input. In a complex organisation with various entities and diverse roles and responsibilities, clear and exact processes for model transfer, version control, and validation are needed. To simplify validation and ensure consistency across settings, models should produce deterministic results. Although data privacy restrictions often limit data sharing across businesses, models and insights can enable cooperative decision-making without compromising data security. For AI systems to be trusted and valuable while adhering to ethical and legal standards, rigorous monitoring and governance are essential. [2].

## XI. Organizational Challenges in Implementing MLOps

A major obstacle to integrating MLOps in organisations is the difficulty associated with human resources and skillsets. The swift progress in machine learning and its incorporation into business operations necessitate a workforce that possesses not just a comprehension of ML but also expertise in the operational aspects of installing and sustaining these models. Nevertheless, there is frequently a discrepancy between the education received in academia and the specific skills and knowledge demanded by the business, resulting in a workforce that requires significant orientation and instruction. The market's high demand for experienced workers worsens the shortage of skilled workers, making it difficult for organizations to find and keep the necessary staff. Moreover, current teams may face resistance when it comes to embracing new tools and methods, as the process of learning and adapting to new MLOps tools can be challenging and intimidating. Organisational culture is essential in overcoming opposition, and cultures that prioritise data are better at seamlessly incorporating these new practices. [5]

## XII. Technical Challenges in MLOps

Integrating MLOps in an organisation presents numerous technical hurdles arising from the intricacy of machine learning models, data administration, and the amalgamation of diverse tools and technologies. Establishing a dependable infrastructure is essential for expanding machine learning models in a production environment. However, numerous organisations face challenges when it comes to scaling resources, overseeing dispersed systems, and guaranteeing data quality. Efficient data management is crucial, as the calibre of data directly influences the efficacy of machine learning models. The absence of uniformity and comprehensive documentation for MLOps technologies exacerbates the complexity of the integration process. Testing ML models has distinct issues in comparison to traditional software due to the inherent non-deterministic nature of ML models. This necessitates the use of specialised testing methods. Deciding on the most suitable tools from a wide range of choices and successfully incorporating them into the current IT infrastructure can be overwhelming and costly. [5]

## XIII. Operational and Business Challenges

The operational difficulties in MLOps frequently pertain to the intricacy of continuous integration and deployment procedures. MLOps, in contrast to traditional DevOps, necessitates the management of substantial amounts of data and the ongoing refinement of models, resulting in increased intricacy. Delays and inefficiencies might arise from dependencies on other teams for deployment pipelines and API hosting providers. Organisations are faced with the challenge of balancing cost, forecast speed, and model accuracy, which often results in the development of overly complex solutions that squander resources. Business issues encompass the task of providing a compelling rationale for the investment in MLOps to senior executives, who could encounter difficulties in perceiving the immediate financial benefits. Furthermore, there is frequently a disparity between the lofty aspirations for what AI/ML may accomplish and the actual constraints it faces, necessitating the skillful handling of stakeholder expectations. The adoption and scalability of MLOps procedures are made more complicated by budget constraints and the sunk cost fallacy. [5]

## XIV. Agile Methodologies in ML and Big Data Projects

Agile methodologies offer a flexible and iterative approach to managing the complexity of machine learning (ML) and big data projects. These methodologies emphasize collaboration, adaptability, and continuous delivery, which are crucial for navigating the dynamic and uncertain nature of these projects. Unlike traditional project management methodologies that follow linear planning and rigid processes, agile methodologies allow for rapid adjustments to changing requirements and market conditions. This flexibility helps organizations avoid delays, reduce cost overruns, and achieve better project outcomes. By integrating agile principles into ML and big data projects, organizations can foster a culture of continuous improvement and innovation, ultimately enhancing their ability to harness the transformative potential of data-driven technologies. Agile frameworks such as Scrum, Kanban, and Extreme Programming (XP) have been successfully applied to ML and big data projects, each offering unique benefits tailored to specific project needs [6].

## XV. CD4ML Pipeline and Task Allocation

The diagram below shows a streamlined Continuous Delivery for Machine Learning (CD4ML) workflow. It shows how data engineers, data scientists, and developers usually divide up tasks. Data engineers are in charge of labelled data and making sure it is ready to be processed further. Then, using the ready data and training code, data scientists build and train machine learning models. Once developers have made a model that works well, they put it into a production setting along with the necessary application code and data.

- This organised splitting of work speeds up the ML process by making sure that each stage is handled by experts with the right skills.
- The CD4ML method focuses on small, safe steps that can be safely repeated and released. This makes ML deployments more efficient and reliable.

[4].

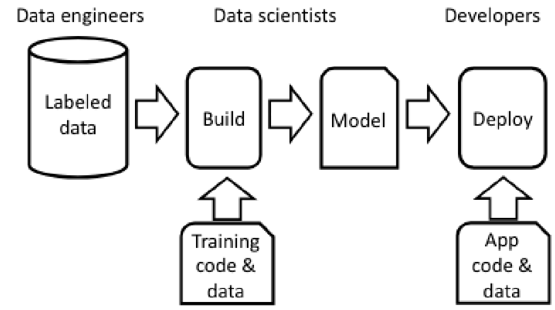


Fig. 2. A simplified CD4ML pipeline and its artifacts, with conventional task allocation for data engineers, data scientists, and application developers [4]. [4]

## XVI. Conclusion

For machine learning and big data technologies to reach their full potential, MLOps must be fully integrated into business processes. We talked about the main parts and problems of MLOps, focusing on the need for DevOps techniques that are specifically designed to meet the needs of machine learning, as well as Managing MLOps projects effectively, for example: Agile methods look like a good way to handle the changing and complicated nature of machine learning projects, making it easier to keep improving and coming up with new ideas. Organisations can make their machine learning systems more scalable, repeatable, and easy to manage by using advanced tools such as Kubeflow and MLflow. For MLOps to work, it is important to deal with the organisational, technical, and operational problems that we brought up. To keep MLOps practices working well, future research and development should focus on automating more of the ML lifecycle processes and combining new technologies.

## References

[1] Thomas Burr et al. On the understandability of mlops system architectures. *IEEE Transactions on Software Engineering*, 49(3):1234–1245, 2023.

[2] Tuomas Granlund, Aleksi Kopponen, Vlad Stirbu, Lalli Myllyaho, and Tommi Mikkonen. Mlops challenges in multi-organization setup: Experiences from two real-world cases. In *2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN)*, pages 82–88. IEEE, 2021.

[3] Dominik Kreuzberger, Niklas Kühl, and Sebastian Hirschl. Machine learning operations (mlops): Overview, definition, and architecture. *IEEE Access*, 11:31866–31878, 2023.

[4] Sasu Mäkinen, Henrik Skogström, Eero Laaksonen, and Tommi Mikkonen. Who needs mlops: What data scientists seek to accomplish and how can mlops help? In *2021 IEEE/ACM 1st Workshop on AI Engineering - Software Engineering for AI (WAIN)*, pages 109–112. IEEE, 2021.

[5] Ashwini Kolar Narayanappa and Chintan Amrit. An analysis of the barriers preventing the implementation of mlops. In *Proceedings of the 2024 International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2024.

[6] Justin Scott and Jack Nathan. Agile methodologies for managing complexity in machine learning and big data projects for business markets. *Journal of Business and Technology*, 3(1), 2024.

[7] Damian A. Tamburri. Sustainable mlops: Trends and challenges. In *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 17–23. IEEE, 2020.

[8] Andres Felipe Varon Maya. The state of mlops. Proyecto de Grado 202626174, Universidad de los Andes, Depto. de Ing. de Sistemas y Computación, Jan 2021.

[9] Samar Wazir, Gautam Siddharth Kashyap, and Parag Saxena. Mlops: A review. *arXiv preprint arXiv:2308.10908*, 2023.

[10] Yizhen Zhao. Machine learning in production: A literature review. *Technical Report*, March 2021.