Uppsala University

Department of Informatics and Media

MLOps paradigm - a game changer in Machine Learning Engineering?

Francois Regis Dusengimana

Master's thesis project, 30 credits

VT 2023

22 May 2023

Abstract

In the last 5+ years, researchers and the industry have been working hard to adopt MLOps (Machine Learning Operations) to maximize production. The current literature on MLOps is still mostly disconnected and sporadic (Testi et al., 2022). This study conducts mixed-method research, including a literature review, survey questionnaires, and expert interviews to address this gap. The researcher provides an aggregated overview of the necessary principles, components, roles, and the associated architecture and workflows resulting from these investigations. Furthermore, this research furnishes a definition of MLOps and addresses open challenges in the field. Finally, this work proposes a MLOps pipeline to implement product recommendations on the e-commerce platform to guide ML researchers and practitioners who want to automate and operate their ML products.

Keywords: MLOps, ML, AI, Automation, ML in production.

MLOps paradigm - a game changer in Machine Learning Engineering?

Research summery

Machine Learning is growing into a critical technique for maximizing the value of data and enabling companies to be more innovative, efficient, and sustainable. Machine learning has become essential for companies seeking to maximize data value and improve their operations (Chen et al., 2014). As companies become more data-driven, ML is facing continuous advancement to leverage all its advantages.

Machine Learning in Operations (MLOps) comes in with principles that help to produce ML products. The principles promoted by MLOps include; CI/CD automation, Workflow orchestration, Reproducibility, Versioning of data, model & code, Collaboration, Continuous ML training & evaluation, ML metadata tracking and logging, Continuous monitoring, and Feedback loops (Kreuzberger et al., 2022). MLOps aims to bring the principles of DevOps to the world of machine learning, making the deployment of ML models more reliable, scalable, and efficient.

In this study, the mixed research methods, quantitative and qualitative, are put together to evaluate whether MLOps is transforming the ML engineering industry or not. The study also addresses the challenges companies face when implementing MLOps. The summarization of challenges and solutions is in the table 1, 2, and 3.

Study findings show that MLOps is a game changer and it has transformed machine learning engineering by allowing for the systematic and automated management of the entire machine learning lifecycle, from development to deployment and maintenance. MLOps has reduced the manual work required using automated tools and processes, allowing teams to focus on higher-level tasks such as feature engineering and model optimization. MLOps has a bright future as more organizations recognize the benefits of using this approach to machine learning engineering.

Based on the survey and interviews, figure 12 illustrates step-by-step the MLOps pipeline to implement product recommendations on e-commerce platforms. This is another contribution of the study to show a pipeline to follow during the implementation of MLOps.

Acknowledgment

Countless people helped me with this thesis. First, without the Swedish Institute (SI), my master's studies in Sweden would not have been possible. I want to express my gratitude to the SI team for awarding me the Swedish Institute Scholarship for Global Professionals (SISGP).

I am also grateful to the interviewees, their companies, and the respondents who shared their knowledge on the subject matter.

Special thanks to my Uppsala University supervisor, Steve McKeever, and examiner PG Holmlöv who provided support and valuable feedback throughout the project and helped me overcome roadblocks.

Finally, I am eternally grateful to my family: my wife and our children for consistently encouraging me to keep digging, my mother for teaching me how to be stubborn in this life, and my siblings for their encouraging words. Also, thank my dear friends and classmates for your unwavering support throughout my studies.

Table of Contents

Acronyms

AI: Artificial Intelligence

AWS: Amazon Web Services

CCPA: California Consumer Privacy Act

CD: Continuous Delivery

CI: Continuous Integration

CPU: Central Processing Unit

GCP: Google Cloud Platform

GDPR: General Data Protection Regulation

ML: Machine Learning

MLOps: Machine Learning Operations

GPU: Graphics Processing Unit

QPS: Queries per second

TFX: TensorFlow Extended

List of figures

List of tables

1. INTRODUCTION

1.1. Motivation

Machine Learning has emerged as a critical technique for unlocking the power of data and enabling businesses to become more innovative, efficient, and sustainable. Machine learning has become essential for companies seeking to maximize data value and improve their operations (Chen et al., 2014). Companies now have access to vast amounts of data that can provide valuable insights thanks to the advent of big data. However, the sheer volume of data makes it difficult for companies to extract meaningful insights manually. This is where machine learning comes in. Companies can quickly and accurately analyze large volumes of data, identify patterns, and make data-driven decisions by leveraging machine learning algorithms and techniques. Machine learning algorithms, for example, can be used to detect fraudulent financial transactions (Kshetri, 2018), forecast customer churn, optimize supply chain management (Li et al., 2018), and automate manufacturing processes (Zeng & Shi, 2019) by applying MLOps principles.

MLOps is a collection of practices aimed at streamlining and automating the process of deploying and managing machine learning models in production. With the increasing adoption of machine learning in businesses, MLOps has become an essential discipline for ensuring machine learning systems' reliability, scalability, and efficiency (Géron, 2019). MLOps practices can assist businesses in managing the entire lifecycle of machine learning models, from training to deployment and monitoring, and ensuring that models remain accurate and up to date over time. Businesses can reduce the time and effort required to deploy machine learning models in production and reduce the risk of model failures and downtime by implementing MLOps practices (Wu et al., 2020). Furthermore, MLOps can assist businesses in increasing their agility and innovation by allowing them to rapidly iterate on and improve their machine-learning models in response to changing business needs and market conditions (Sculley et al., 2015).

MLOps is thus an essential component of a successful machine learning strategy that can assist businesses in staying ahead of the competition in today's data-driven economy (Bonomi et al., 2020). MLOps is a practice in which three disciplines collaborate: machine learning, software engineering, and data engineering. MLOps aims to bridge the gap between development and operations by utilizing machine learning systems in production (Kreuzberger et al., 2022).

According to Kreuzberger et al. (2022), MLOps aims to make machine learning product development easier by leveraging the following principles: CI/CD automation, Workflow orchestration, Reproducibility, Versioning of data, model, and code, Collaboration, Continuous ML training and evaluation, ML metadata tracking and logging, Continuous monitoring, and Feedback loops. MLOps is an approach that includes best practices, concepts, and development culture for machine learning products from start to finish conceptualization, execution, monitoring, deployment, and flexibility (Kreuzberger et al., 2022).

1.2. Problem statement

There is still a gap in the literature about MLOps (Testi et al., 2022). This academic gap leads to many challenges at the industry level when they want to implement MLOps. This research intends to bridge the literature gap and address the open challenges to adopting MLOps.

1.2.1. At the industry

To take advantage of ML's benefits, businesses are embracing MLOps. The adoption of MLOps is a challenging and demanding process. The researcher met with a team of engineers from a retail company to discuss the topic while he was conducting interviews. The group of software developers, ML engineers, AI architects, and MLOps engineers. On February 15, 2023, a group of ten professionals was working on the MLOps project for the second year at the time of the meeting. The project was not finished yet. The MLOps project's implementation calls for a multidisciplinary team and managerial support. The group described the difficulties they were having on the journey. In the findings and discussion section, Tables 1, Table 2, and Table 3 list these challenges along with the corresponding solutions.

1.3. Research objectives

- This study aims to explore MLOps in real teams to determine whether it is or is not a game changer in the Machine Learning industry.
- According to Kreuzberger et al. (2022), companies face three categories of open challenges when implementing MLOps; Organisational, ML system, and Operational challenges. This study will propose solutions to these challenges.
- The study will propose an MLOps pipeline to implement product recommendations on e-commerce platforms.

1.4. Research question

The study will answer the research question below at the end of the research.

RQ. Is the MLOps paradigm - a game changer in Machine Learning Engineering?

2. BACKGROUND

2.1. Machine Learning

Machine learning has significantly progressed in recent years, particularly in production applications. Deep learning models have achieved cutting-edge performance in tasks such as natural language processing, image and speech recognition, and self-driving vehicles, to name a few. Another development is the application of reinforcement learning to optimize industrial processes and robotics. Transfer learning has also been used to improve the efficiency of training models on small datasets, and explainable AI techniques have been developed to improve ML model interpretability. ML has been used in manufacturing for various applications, including predictive maintenance, quality control, and supply chain optimization. Advances in automated machine learning (AutoML) and distributed computing have also aided ML in production. AutoML tools are designed to simplify the ML process and reduce the need for manual intervention, whereas distributed computing enables ML algorithms to scale and handle large amounts of data.

Several academic references back up these developments. A Google paper, for example, demonstrates how TensorFlow and Kubernetes can be used to build scalable and efficient ML pipelines in production (Abadi et al., 2016). Another IBM study emphasizes the advantages of AutoML in reducing the time and cost of ML development (Palmes et al., 2021). OpenAI discusses recent advances in deep learning and their impact on various fields, including production applications, in a GPT-4 technical report (OpenAI, 2023). These examples demonstrate the significant impact that machine learning has had on production applications and the potential for future growth.

2.2. Machine learning Operations (MLOps)

Today's literature demonstrates collaboration between academia and industry to improve the productivity of ML. According to Andrew (2021), model development has received much attention over the last decade. According to Kreuzberger et al. (2022), the academic community has focused on machine learning model development and benchmarking rather than operating complex machine learning systems in real-world scenarios. The MLOps is the solution to the existing gap between development and production. The ML engineering team aims to deploy and run the model component in the production environment (Andrew, 2021). Once the entire ML project or system is operational, it is time to consider monitoring and maintaining the system without interfering with production. Sculley and his colleagues use the technical debt metaphor to explain how difficult and expensive it is to maintain ML systems (Sculley et al. 2015). Ward Cunningham introduced this metaphor in 1992 to help explain the long-term costs of rapid development in software engineering. Thinking through the ML project lifecycle is an efficient way to plan out all the steps to be completed throughout the process. The MLOps discipline has a set of tools and principles to help with the progression of ML projects through the lifecycle (Andrew, 2021).

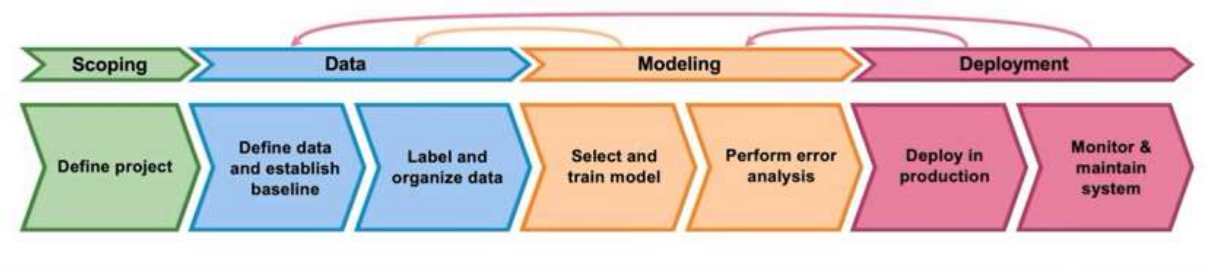

Figure 1: MLOps life cycle, *Credit: Andrew, 2021.*

In its life cycle, an ML model is developed and improved in an analytics pipeline or workflow that touches stakeholders around the organization (Sweenor et al., 2020). Realizing the value of data science and machine learning is a matter of reducing friction throughout those pipelines and workflows. ML models need constant refinement. They include data transformations and embody relationships with continually changing data, and data drift affect their predictions' accuracy. As such, operationalizing ML pipelines has management implications that differ from traditional software application engineering. In particular, long-term accuracy depends on periodic tuning, retraining, and even complete remodeling (Sweenor et al., 2020). Innovative organizations are currently adopting MLOps. They structure their data science and machine learning pipelines in three pipelines: data pipeline, model pipeline, and deployment pipeline (Sweenor et al., 2020). According to Testi et al., (2022) the current literature about MLOps is still mostly disconnected and sporadic. Even if the literature has different steps and pipelines, the main idea and logic behind the steps are the same.

One of the exciting and challenging moments of any ML project is when people get to deploy the model. Two main issues make ML model deployment hard: machine learning issues and software engineering issues. When people deploy an ML system, they must ensure they have a mechanism to manage changes such as concept drift and data drift on a continuously operating production system. Conversely, software engineering has many issues when deploying an ML model: real-time or bath, cloud vs. edge/browser, computer resources (CPU/GPU/memory), latency, throughout (QPS), logging, security, and privacy. In addition, the production system must run continuously at the lowest possible cost while producing maximum output (Andrew, 2021).

MLOps helps improve production models' quality while incorporating business and regulatory requirements and model governance. MLOps is trying to solve problems like:

- Inefficient workflows: MLOps provides a framework for managing the machine learning lifecycle effectively and efficiently (Hewage et al., 2022; Garg et al., 2022). MLOps creates a more structured, iterative workflow by matching business expertise with technical prowess.
- Bottlenecks: With complicated, non-intuitive algorithms, bottlenecks can often happen. MLOps facilitates collaboration between operations and data teams, helping to reduce the frequency and severity of these types of issues (Hewage et al., 2022). The MLOps encouraged collaboration leverages the expertise of previously siloed teams, allowing for more efficient development, testing, monitoring, and deployment of machine learning models.

However, solving these problems takes time and effort. Most MLOps practitioners point out the mindset and culture in ML governance, where a significant number of skilled personnel is vital to the performance of MLOps. Concepts and data drift are also a big challenge in automating operations, ML infrastructure, and software (Garg et al., 2022; Kreuzberger et al., 2022) all these challenges are discussed in the discussion section.

2.2.1. Definition of MLOps

Previous researchers, Symeonidis et al. 2022; Garg et al. 2022; Kreuzberger et al. 2022; Hewage et al. 2022, have defined MLOps in many ways. This study combines these definitions into the following MLOps definition; MLOps is a practice combining DevOps principles with machine learning to automate and manage the entire machine learning lifecycle, from development to deployment and monitoring. It aims to streamline the process of building, testing, deploying, and monitoring machine learning models in production by incorporating best practices from software engineering and data science. MLOps also focuses on collaboration between data scientists, developers, and DevOps teams to ensure seamless integration and smooth operation of machine learning models in production. MLOps aims to increase the speed and reliability of machine learning model deployment while ensuring the ethical and responsible use of models.

2.2.2. Involvement in MLOps process

The Machine Learning Operations implementation process generally includes the following steps:

- Data Collection and Preparation: Gathering and preparing data for model training and evaluation.
- Model Development: Developing and testing machine learning models, including feature engineering, model selection, and hyperparameter tuning.
- Model Deployment: Deploying the model in a production environment, such as a cloud-based platform or on-premise infrastructure.
- Monitoring and Maintenance: Monitoring model performance, updating the model as needed, and maintaining the infrastructure that supports the model.
- Continuous Integration and Delivery (CI/CD): Automating and integrating the model deployment process with the overall software development pipeline.
- Model Management: Managing multiple models and versioning them, ensuring that the correct model is used in production, and maintaining a history of models over time.
- Governance and Compliance: Ensuring that models are aligned with organizational policies and regulations and that the data used for training and evaluation is appropriately secured and managed.

2.2.3. Challenges in implementing MLOps

Researchers have highlighted different open challenges companies face implementing MLOps (Garg et al., 2022; Kreuzberger et al., 2022). These open challenges are grouped into three main categories:

- Organizational challenges: In organizational settings, the mindset and culture of data science practice are common challenges (Kreuzberger et al., 2022). To successfully develop and deploy ML products, a change in culture from model-driven machine learning to a product-oriented discipline is required. Because developers are not vocal about automatically delivering each change, they are still more inclined to manually test and deploy AI models in production (Garg et al., 2022). According to Garg et al., (2022) the lack of communication and collaboration among development and production teams emphasize the manual working routine.
- ML system challenges**:** Today, companies have vast amounts of data from different sources. Data changes every time, and concepts also may change. Developing a robust ML system to adapt to this dynamic environment is challenging.
- Operational challenges**:** Monitoring the effectiveness of AI models is one issue that needs to be addressed early on. Even if the collected data is error-free, changes in data are more frequent, interrupting the entire model (Garg et al., 2022). Manually operating ML in productive settings is difficult due to the various stacks of software and hardware components and their complex interactions. Therefore, High-level automation is required, and the constant inflow of new data necessitates retraining capabilities (Kreuzberger et al., 2022). Monitoring and governance are critical aspects of managing machine learning models. Monitoring assists in identifying and addressing performance issues and data drift, whereas governance ensures regulatory and ethical compliance. Robust monitoring and governance processes necessitate a mix of automated and manual checks, such as regular audits, explainability analyses, and ongoing assessment of model fairness and bias. Proper monitoring and governance are required to ensure that machine learning models are dependable, accurate, and accountable throughout their lifecycle.

3. THEORETICAL FRAMEWORK

According to Andrew (2021), Garg et al. (2022), and Kreuzberger et al. (2022) MLOps is a practice that focuses on the operational aspects of deploying and managing machine learning models in production. It involves collaboration between data scientists, developers, and DevOps teams to automate the entire ML lifecycle, from development and testing to deployment and monitoring (Kreuzberger et al., 2022). The shared vital concepts include;

Automation: Automating the ML pipeline, including model training, testing, deployment, and monitoring.

Collaboration: Encouraging collaboration between data scientists, developers, and DevOps teams to ensure seamless integration of machine learning models into production.

Model Management: Managing and tracking the different versions of models and their respective configurations.

Model Monitoring: Monitoring the performance of models in production and detecting drift, which refers to the difference between the expected output and the actual output.

Model Governance: Ensuring the ethical and responsible use of machine learning models, such as preventing discrimination and ensuring data privacy.

MLOps aims to bring the principles of DevOps to the world of machine learning, making the deployment of ML models more reliable, scalable, and efficient. In the following part, these fundamental concepts are highlighted to show how MLOps is a need in the industry today.

3.1. Industry needs MLOps

The industry needs MLOps because as machine learning models become more widespread, it becomes increasingly important to ensure that these models are continuously monitored and maintained to ensure they continue to deliver accurate and reliable results. MLOps helps to automate and streamline the process of deploying, monitoring, and maintaining machine learning models in a production environment, enabling organizations to quickly identify and resolve issues, improve the performance and accuracy of their models, and ultimately drive better business outcomes (Makinen et al., 2021). Additionally, MLOps helps improve collaboration between data scientists and IT operations teams and helps ensure that models are correctly governed and meet compliance requirements (Kreuzberger et al., 2022).

Since introducing MLOps and its principles, the industry has shown they already lack the advantages MLOps can provide in ML products. The advantages of MLOps in revolutionizing ML industry can be summarized in the following key concepts:

- Collaboration: In the previous environment, all teams involved in ML projects, like ML engineers, Data scientists, software developers, and IT operations engineers, worked in an isolated manner (Kreuzberger et al., 2022). This old fashion delayed the project and made it hard to manage separated teams. MLOps allows the entire team to work cooperatively.
- Automation: Achieving desired business outcomes necessitates the benefits of automating software and developing machine learning. By automating the lifespan of ML-powered software, diverse teams can focus on more critical business issues, resulting in faster and more reliable business solutions (Battina, 2019).
- Effectiveness: MLOps boosts the productivity of all production teams and the entire machine learning project development workflow, from conception to deployment (Battina, 2019).
- Workflow: Every machine learning project entails teams of data scientists and machine learning engineers working together to create cutting-edge models by hand or computer. Data scientists consider how the model has evolved and the model's complexities when selecting an ML model for training. Development and test environments are distinct from staging and production environments. ML manual process models frequently fail to adapt to changes in production environment dynamics or changes in data representing production environments (Battina, 2019). MLOps comes in as a solution to facilitate workflow automation. Collaboration of involved teams to look for one solution leads to a fully automated workflow.

3.2. Relevant theories

The MLOps framework developed by John et al. (2021) is the foundation of this study. The framework explains the pipelines and steps involved in MLOps adoption, including the data pipeline, model pipeline, and release pipeline. Every pipeline has several steps to execute and a responsible professional(s).

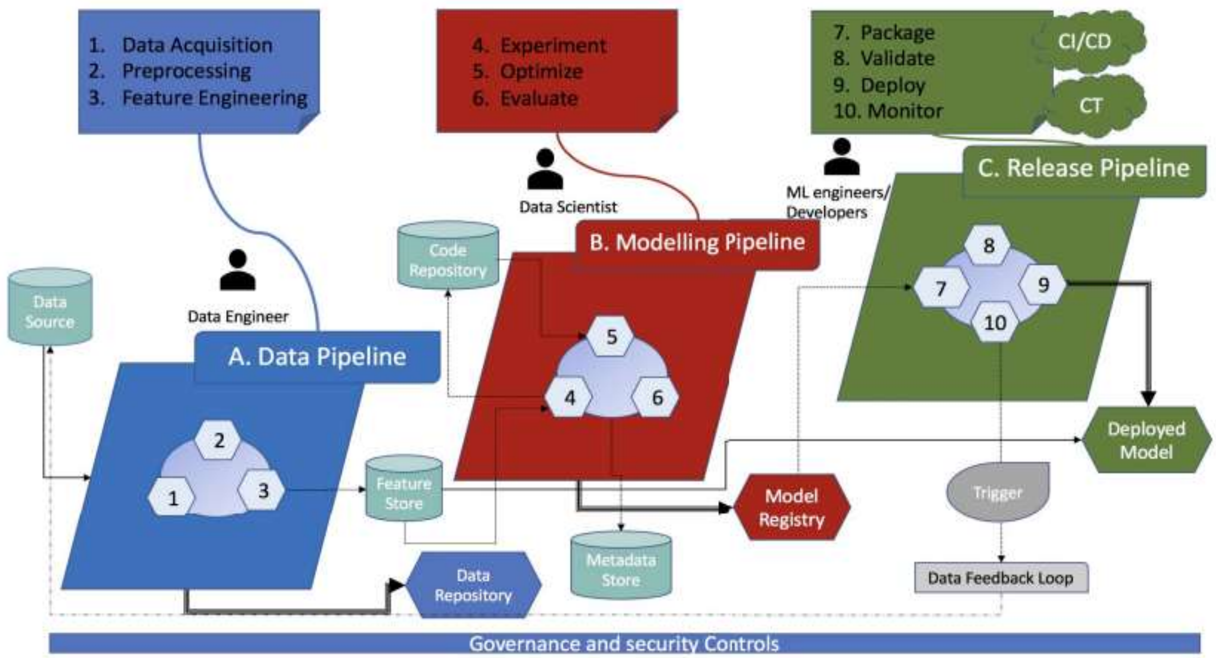

Figure 2: MLOps framework, *Credit: John et al., 2021.*

3.3. Contribution of this study

According to John et al. (2021), MLOps is still in its infancy in practice, with only a few standard guidelines for successfully incorporating it into existing software development methodologies. The contribution of this study on the growing MLOps concepts is to provide a compilation of the previous research on the topic and critically addresses the challenges companies currently face when implementing MLOps. The study also provides a step-by-step MLOps pipeline to implement product recommendations on e-commerce platforms.

4. METHODOLOGY

In this study, using both quantitative and qualitative methods provides a robust and comprehensive approach to investigating an MLOps paradigm and answering the research question. For questionnaire data, the quantitative method is used to allow statistical analysis. Quantitative methods can provide precise and objective data that can be analyzed statistically to identify patterns, trends, and relationships between variables (Creswell, J. W. 2009). This enables researchers to draw quantitative conclusions and generalize about a larger population. The qualitative method is used to help researchers comprehend the complexities, context, and subjective experiences of interviewers. Qualitative methods enable in-depth investigation of human behaviour, attitudes, beliefs, and motivations (Alshenqeeti, 2014). According to Creswell, J. W. (2009), qualitative data can provide rich and detailed insights that quantitative methods cannot alone. This combination helps the researcher to have grounded and reliable findings.

4.1. Research process

The figure 3 shows the process guided by the researcher in this study. Even if the study aims to determine if the MLOps paradigm is the game changer in the machine learning industry, there are two sub-tasks the researcher needs to work on; MLOps implementation challenges and solutions and MLOps pipeline for product recommendations on e-commerce platforms. To address these sub-tasks, two categories of data are in place; primary and secondary. The researcher used questionnaires and conducted interviews to get primary data. Secondary data are from previous similar work. The data go through analysis to get the research findings.

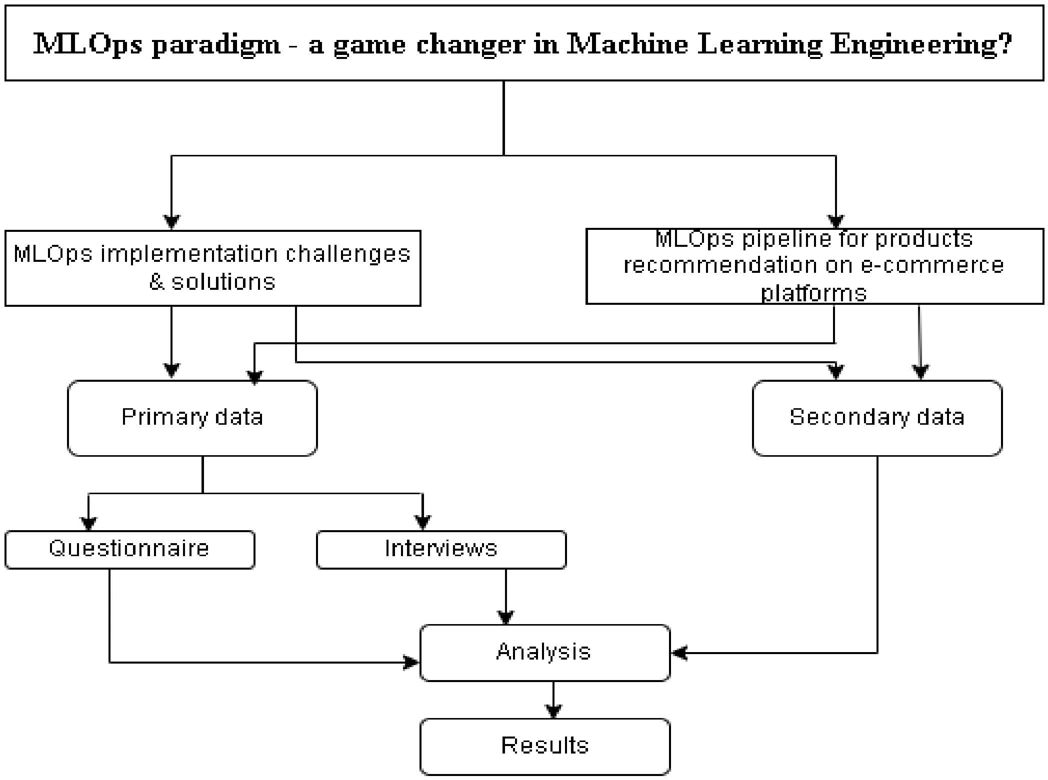

Figure 3: Research process

4.2. Data collection methodologies

The researcher has conducted mixed-method research, including a literature review, survey questionnaire, and interviews with people who work with ML, data engineering, and software engineering from various organizations. The targeted professionals are MLOps engineers, ML engineers, Data Scientists, Data engineers, DevOps engineers, Software engineers, Backend developers, and AI Architects. The study discovered how these interviewees and respondents use, perceive, and comprehend MLOps. To find more people to do interviews and questionnaire respondents, the researcher used professional networks, social media, and online platforms like LinkedIn to target people in the specified field of research interest, which helps to get reliable feedback.

4.2.1. Secondary data

The Google Scholar search engine is the primary source of the majority of literature in this study to access credible scholarly databases. Search queries like MLOps survey, MLOps machine learning and MLOps DevOps led a researcher to many valuable articles, books, and blogs all talking about MLOps. Machine Learning Operations is an interdisciplinary field combining software engineering practices, data engineering, and machine learning to improve machine learning development and deployment speed and efficiency. There has been a growing body of literature on MLOps, with research and development efforts to improve the end-to-end machine learning development process. These efforts aim to increase the speed and efficiency of model deployment and improve machine learning systems' reliability, security, and transparency.

To facilitate the development of MLOps, different agencies have developed tools and frameworks. MLOps tools and frameworks like TensorFlow Extended (TFX), Kubeflow, Apache Airflow, AWS SageMaker, and Google Cloud AI Platform are essential for the MLOps life cycle. These tools provide a unified platform for developing, deploying, and managing machine learning models and offer a range of capabilities such as automated model training, model serving, and monitoring.

4.2.2. Primary data

4.2.2.1. Questionnaire

During the study, a survey questionnaire was used to reach more people. Most of the respondents were reached through online platforms.

Why an online survey?

For several reasons, online surveys are a popular method for collecting data in research studies: Convenient and accessible: Online surveys are simple to administer and can reach many people quickly and efficiently. The survey can be completed from any location with internet access. Cost-effective: Online surveys are more cost-effective than paper surveys or in-person interviews. They minimize the need for printing and mailing costs, and there are numerous free or low-cost options. Increased response rates: Because online surveys are more convenient for participants, they often receive higher response rates than traditional methods. Respondents are more likely to participate if the survey can be completed quickly and easily. Real-time data collection: Online surveys allow for real-time data collection, making tracking and analyzing responses as they come in easier. Improved data quality: Online surveys frequently produce more accurate and reliable data because participants can review and change their responses before submitting them.

The researcher developed an online survey form targeting people in the field. The targeted respondents were the MLOps engineer, ML engineer, Data Scientist, Data engineer, DevOps engineer, Software engineer, Backend developer, and AI Architect. After many alliterations and discussions between the researcher and the supervisor, they came up with the final version of the online survey questionnaire on 28th January 2023. Monday, 30th, 2023, the form was published online.

The following channels were used to reach out to many participants; LinkedIn: According to Wikipedia, LinkedIn is a business-oriented social media platform with websites and mobile apps. The platform is primarily used for professional networking and career development, with job seekers and employers able to post curriculum vitae and job openings. This platform served as a great channel in the data collection journey. The researcher has a significant connection in the target field on the platform allowing him to collect more responses from LinkedIn. Kaggle: The researcher posted the survey link in the discussion forum on the Kaggle platform. Kaggle is a data scientist and machine learning practitioners online community. This is an excellent community of data science and machine learning professionals with whom to collaborate in order to obtain trustworthy input. Direct messages via emails, Facebook, and WhatsApp are other essential channels to reach the actors in the ML and development domains.

4.2.2.2. Interviews

The interview method is a common data collection technique because it allows the researcher to gather in-depth information and understand the studied subject (Alshenqeeti, 2014). It also allows the researcher to ask follow-up questions and probe for additional details. Interviews can be conducted in person or over the phone and structured (using a pre-defined questionnaire) or unstructured (allowing for a more open-ended conversation). The flexible method can be adjusted to suit specific research needs and goals. Additionally, interviews can be used to gather data from a small sample of participants, making it suitable for exploratory or qualitative studies. A purposive or strategic sampling was used to target professionals in the ML field. In qualitative research, purposeful sampling is commonly used to ensure that the participants chosen are the best fit for the research question and can provide the most relevant information (Lemp J. et al., 2012). The researcher may select participants based on their experience, expertise, knowledge, or perspective on the topic under study. Purposive sampling, as opposed to random sampling, involves selecting participants based on specific characteristics or criteria relevant to the research question or hypothesis.

Three semi-structured interviews were used in this research. The first interview was conducted on 15 February 2023, where the researcher met a group of professionals in a retail company. Eight people working on the MLOps project in the company contributed a lot to this study. It was a mixed team of MLOps engineers, ML engineers, AI architects, and Software developers. The second interview was on 23 February 2023 with an AI research scientist. The third interview was with a senior data consultant on 27 February 2023. A semi-structured interview is a research method that combines the unstructured interview's flexibility with the structure of a structured interview. The researcher has a list of predetermined questions in a semi-structured interview. However, the interview also allows follow-up questions and probes to delve deeper into participants' responses. In qualitative research, semi-structured interviews are commonly used to collect information about participants' experiences, perspectives, and attitudes. They allow for a more in-depth exploration of the participant's perspectives than a structured interview while maintaining some structure and consistency across interviews.

To better understand the courant situation in MLOps implementation, the researcher met with the engineering teams of different companies to discuss MLOps, its advantages and challenges. The following questions were discussed during the interview meetings;

Q1. Can you explain in a few words what MLOps is and why it is becoming so important in the ML field?
Q2. How does MLOps differ from traditional DevOps?
Q3. What are some of the key challenges in implementing MLOps? And how to overcome these challenges?
Q4. What are the biggest advantages of incorporating MLOps into a company’s machine learning workflow?

Q5. How does MLOps help to improve collaboration and communication among ML engineers, Data scientists, and operations teams?
Q6. How do you ensure the robustness and reliability of models once they are deployed in production?
Q7. How do you see the future of MLOps?
Q8. What technologies will shape the future of MLOps?

These open questions grounded in open discussion contributed a lot to find the current and future role of MLOps in the ML engineering industry. The researcher met people in the targeted discipline to discuss the above questions. In the analysis block, the researcher highlights the outcomes from the interview meetings.

4.3. Ethical considerations

As with any technology, MLOps raises several ethical considerations that must be addressed to ensure that the models are developed, deployed, and operated responsibly. Some of the key ethical considerations in MLOps are discussed below.
Bias and fairness: One of the most serious ethical concerns in MLOps is the possibility of bias and unfairness in machine learning models (Niemelä et al., 2022). Models can be biased if they are trained on unrepresentative data or if they encode pre-existing biases in the data. Bias can lead to incorrect or unfair decisions, especially regarding lending, hiring, and criminal justice applications. MLOps teams must collaborate to ensure that models are trained on diverse and representative data and tested for bias before being deployed in production. Privacy and security: In MLOps, privacy and security are critical considerations. Machine learning models frequently make predictions using sensitive data, such as financial information, medical records, and personal information. MLOps teams must ensure the security of the data used to train and deploy models and the models themselves from unauthorized access or manipulation. Explainability and transparency: Machine learning models can be complex and challenging to comprehend, making it difficult to determine why a specific decision was made. This lack of transparency can breed doubts and distrust in machine learning (John et al., 2021). MLOps teams must ensure that models are transparent and easy to understand so stakeholders can understand how decisions are made. Ethical use: MLOps teams must also consider the ethical use of machine learning models.

For example, models used in the criminal justice system must be fair and unbiased, while models used in the medical field must be used to benefit patients and not to exploit or harm them. MLOps teams must ensure that models are developed, deployed, operated ethically and used for the intended purpose (Niemelä et al., 2022).Continual monitoring and improvement: Machine learning models must be continuously monitored and improved by MLOps teams to ensure that they perform as intended and that any biases or other ethical considerations are addressed. This could include updating the data used to train models, improving the algorithms, or changing how models are deployed in production. MLOps teams must ensure that machine learning models are developed, deployed, and operated ethically and responsibly. MLOps teams can help ensure that machine learning models benefit society rather than harm it by addressing the considerations above.

This study was conducted following research ethics at every stage. The message on the questionnaire stated clearly that no personal questions would be asked and that questions would be completely anonymous. Participants were informed before the interview that the interview would be transcribed and that the researcher may quote them in the text if necessary. The researcher then read them their rights, including the right not to answer any uncomfortable questions and leave the interview at any time. They were also informed that their identities would be anonymized and that their data would be stored and processed in accordance with GDPR regulations.

5. RESULTS

In this chapter, the researcher analyses the primary data, both interviews and questionnaires. Interviews are analyzed using a thematic approach. The analysis of questionnaire data is done in Jamovi software to develop statistical analysis.

5.1. Interviews

5.1.1. Analyzing interviews

After collecting all the interview data, it was thematically analyzed. A thematic analysis was chosen. According to Braun & Clarke (2006), thematic analysis is a technique for systematically identifying, categorizing and providing insight into meaning patterns (themes) in a data set. It seeks to reveal patterns within the data collected, which was especially useful for this project because it aims to explore MLOps and address challenges companies face when implementing it. Furthermore, data were analyzed using Nvivo, a qualitative data analysis software program. The first step in a thematic analysis was to become familiar with the data and better understand how to sort data into different themes. Five themes were developed based on the interviewees' responses: importance, challenge, teamwork, future trend, and policy. Figure 4 of the coding chart illustrates the six themes and their percentage weights.

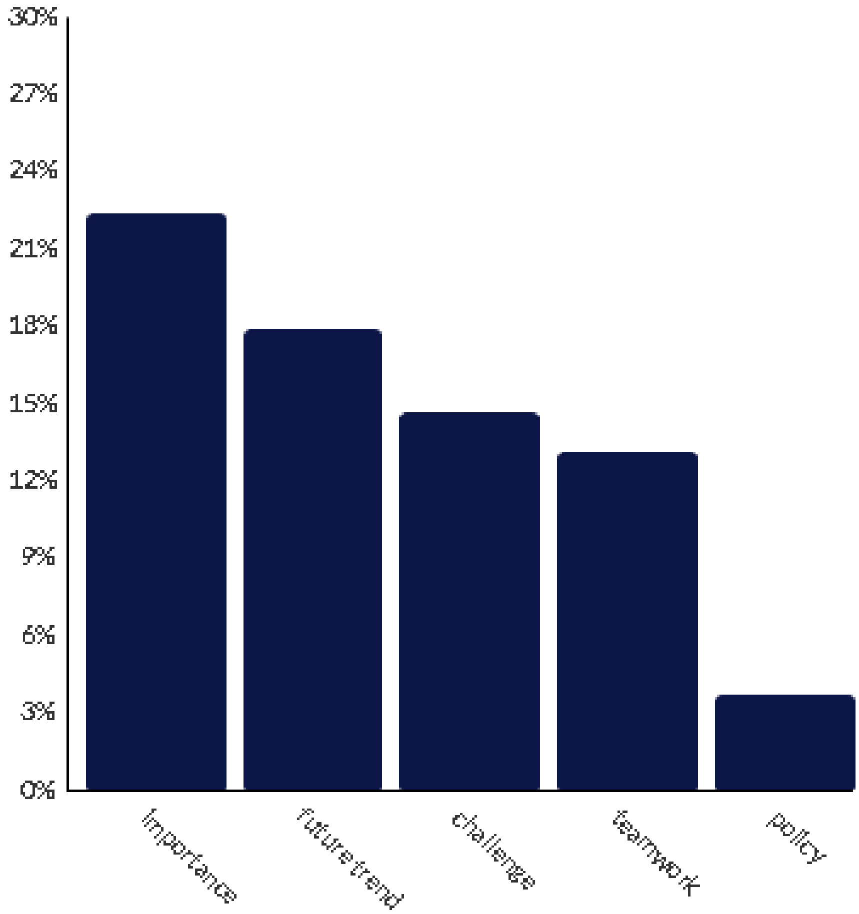

Figure 4: Coding chart

Theme 1: Importance of MLOps in the ML industry

MLOps is critical for building and deploying ML models at scale. It is becoming essential to the ML industry as businesses seek to use ML to push business performance**.**

This theme aims to understand MLOps, the difference between MLOps and DevOps, and its importance in the ML industry. One interview stated that the main difference between MLOps and DevOps is “*Data, model, and concept versioning over code versioning.*” *“MLOps is essentially a specific implementation of DevOps for machine learning projects and pipelines.” “MLOps entails automating the entire machine learning lifecycle, beginning with data acquisition and ending with model deployment and monitoring.”*

Theme 2: Challenges of implementing MLOps

MLOps is incorporating machine learning into the software development and deployment process to ensure that ML models are deployed in a reliable, repeatable, and scalable manner. Implementing MLOps necessitates a mix of technical expertise, effective coordination, and a comprehensive understanding of the business and regulatory landscape, making it a difficult but necessary practice for companies to embrace to maximize the value of their ML investments.

5.1.2. Organizational challenges

Business domain (understanding): MLOps implementation in an organization requires the approval of top managers. Sometimes senior managers need help understanding why a company can invest in this technology. MLOps includes many domains of expertise, which may result in a misunderstanding at the horizontal level. ***Mindset***: There is an old informal principle of developers saying, "… *if it works, do not touch it.*" This mindset still exists in the development industry, where people argue that companies can not invest in new tech if the existing technology does the same. They need to consider added values like accuracy, speed, and robustness. ***Microeconomic and MLOps implementation cost***: A complete MLOps life cycle casts a lot of resources, including money and personnel. One of the companies the researcher interviewed said ten full-time professionals, Software engineers, ML engineers, Data scientists, and MLOps engineers, are working on the MLOps project. During the time of the interview, it was the second year the team had been working on the MLOps project, and still, the project still needed to be completed. Every one can imagine how much money the company is paying the team, and the cost is not limited to personnel. ***Skilled professionals***: MLOps is quite a new concept. Many people in ML, data science, and software development are learning the concept. Skilled professionals are there, but with specific skills putting all together to come up with the entire MLOps life cycle is still a challenge. ***Collaboration***: For the MLOps project to succeed, collaboration among the staff is the key instrument. It is not easy to bring together all these experts working on the same project with different responsibilities but having the same goal.

5.1.3. ML systems challenges

Data protection laws: In some cases, the data protection laws become a barrier to automating the MLOps process. Personal data must be processed lawfully, fairly, and transparently under the GDPR. Individuals must also be informed about the processing of their data and given certain rights, such as the right to access, correct, or delete their data and object to its processing. Furthermore, the GDPR imposes specific requirements for AI processing of personal data, such as ensuring data accuracy, minimizing the amount of data collected, and providing transparency regarding AI algorithms. Automating the lifecycle: An MLOps engineer said, "…*automating the entire lifecycle of machine learning products is the key mandate of MLOps*" However, it is a challenge to make it fully automated. "…*we always face problems in the automation process*".

Data drift and concept drift, deployment complexity, and model explainability and interpretability are other ML systems challenges.

5.1.4. Operational challenges

Monitoring [model layer and platform layer]: Monitoring is an essential aspect of MLOps, and a solid monitoring strategy that addresses both the model layer and the platform layer is required. MLOps teams can quickly identify and address issues by monitoring performance metrics, resource utilization, errors and failures, and security threats, ensuring their machine learning models perform optimally.

Theme 3: Collaboration between teams involved in the MLOps life cycle

MLOps necessitates collaborative effort among data scientists, software engineers, and operations professionals. It is not easy to maintain collaboration among the teams involved in the MLOps life cycle, and the success of the ML projects results in that collaboration. Teamwork between teams involved in the MLOps life cycle is essential to practical machine learning model development, deployment, and maintenance. One interviewee said, "… *the sprint meetings help maintain good team collaboration.*" Another interviewee said, "...*we have to be professional if you understand your task, you work on it, and respect the schedule*." The MLOps cycle includes diverse workgroups with roles and responsibilities, such as data scientists, data engineers, DevOps engineers, and business stakeholders. Collaborating among teams involved in the MLOps life cycle is critical for model development and deployment success. Teammates can build robust and scalable machine-learning solutions by establishing shared goals, communication channels, and workflows.

Theme 4: The future trend of MLOps

MLOps is becoming increasingly important as more businesses adopt machine learning and artificial intelligence, ensuring efficient and effective deployment of these models. Machine learning professionals argue that MLOps is here to stay. They base their arguments on the importance of MLOps in transforming ML engineering.

One interviewee summarised how he sees the future of MLOps and the technologies that will shape the future of MLOps. *"I believe the future of MLOps will be heavily influenced by the integration of AI and automation in the deployment, monitoring and management of ML models. The trend towards cloud-based MLOps solutions will continue to grow, along with the adoption of containerization and orchestration tools. The future of MLOps is bright, and I am excited to see how it will revolutionize how we approach machine learning operations."* Increased automation, integration with established DevOps practices, and a focus on transparency, security, and compliance will likely characterize the future of MLOps. MLOps will continue to play a critical role in ensuring these models' efficient and effective deployment in various applications as machine learning becomes more widely adopted.

One interviewee said the MLOps scope covers three essential phases: "*The complete MLOps process includes three broad phases; designing the ML-powered applications, ML experimentation, development and deployment and ML operations and observability*." MLOps entails automating the different stages of the machine learning funnel, such as data acquisition, data preprocessing, model training, testing, deployment, and monitoring. Automating the MLOps lifecycle can reduce manual intervention while also improving the overall efficiency of the machine learning pipeline. Companies may boost machine learning models' speed, reliability, and quality while lowering the risk of inconsistencies and increasing productivity by automating the MLOps lifecycle.

Theme 5: Policy

Existing policies and data protection policies can complicate MLOps implementation as they were created with traditional software development workflows in mind and may not account for the unique characteristics of machine learning models. Because machine learning models frequently rely on sensitive and personally identifiable information, data protection, privacy, and security policies may be more stringent (Char DS et el., 2018). Model transparency and explainability policies can also be challenging to implement because they may necessitate additional testing and documentation.

5.2. Questionnaire

5.2.1. Exploring the questionnaire data

In this study, 84 respondents answered the questionnaire. Among the respondents, 54 (64%) work in companies where MLOps is implemented or under implementation, while 30 (36%) know about MLOps, but their companies still need to implement MLOps. According to the survey results, 42 (50%) of respondents have experience of less than five years, 36 (43%) have five years to ten years of experience and 6 (7%) have more than ten years of experience. This makes sense because MLOps is a new concept, and a large number of respondents are using MLOps in their companies.

After one month of collecting feedback from MLOps' professionals, the researcher and his supervisor decided to close the survey. This number is sufficient to get a clear picture of the MLOps field because all 84 participants are in the targeted group who know more about the topic in the investigation. The targeted professionals are MLOps engineers, ML engineers, Data scientists, Software engineers, Data engineers, DevOps engineers, Backend developers, and AI architects. The Figure 5 of question one shows the distribution of the respondents. It is good that people responded the survey are in the positions and have background linked with machine learning engineering.

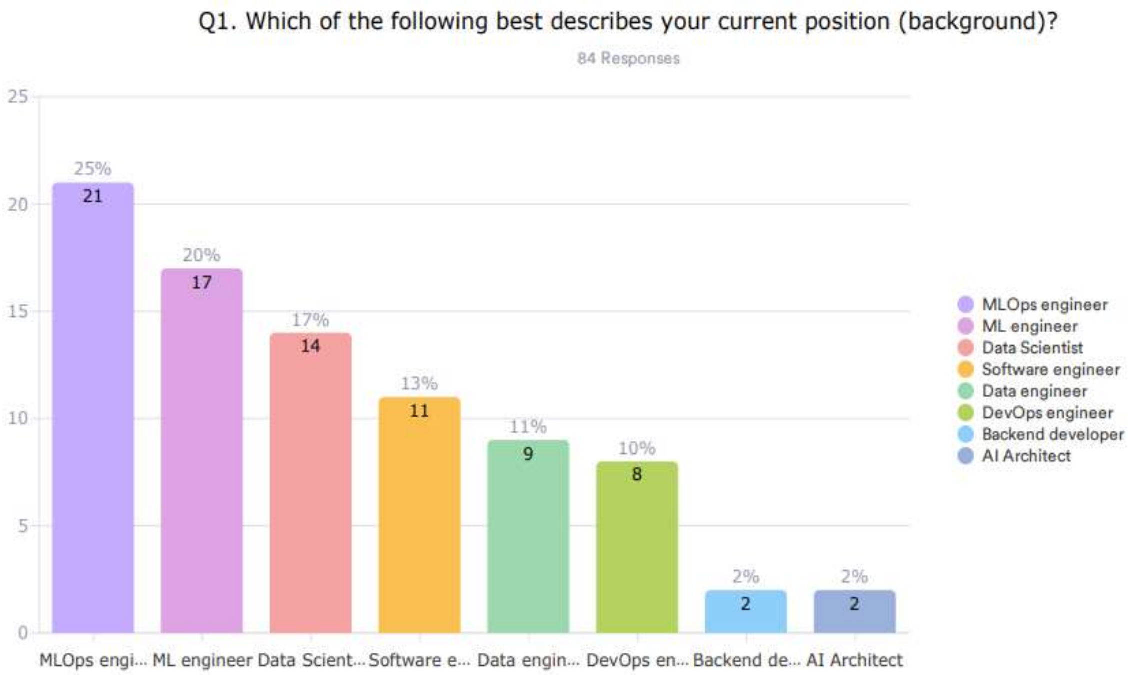

Figure 5: Distribution of respondents

Figure 6 of the third question shows the size of our respondents' company. The figure shows that 30% of respondents come from more prominent companies with more than ten thousand employees. This is understandable because the bigger the company, the more adoption is easier. Big companies have the resources to implement MLOps.

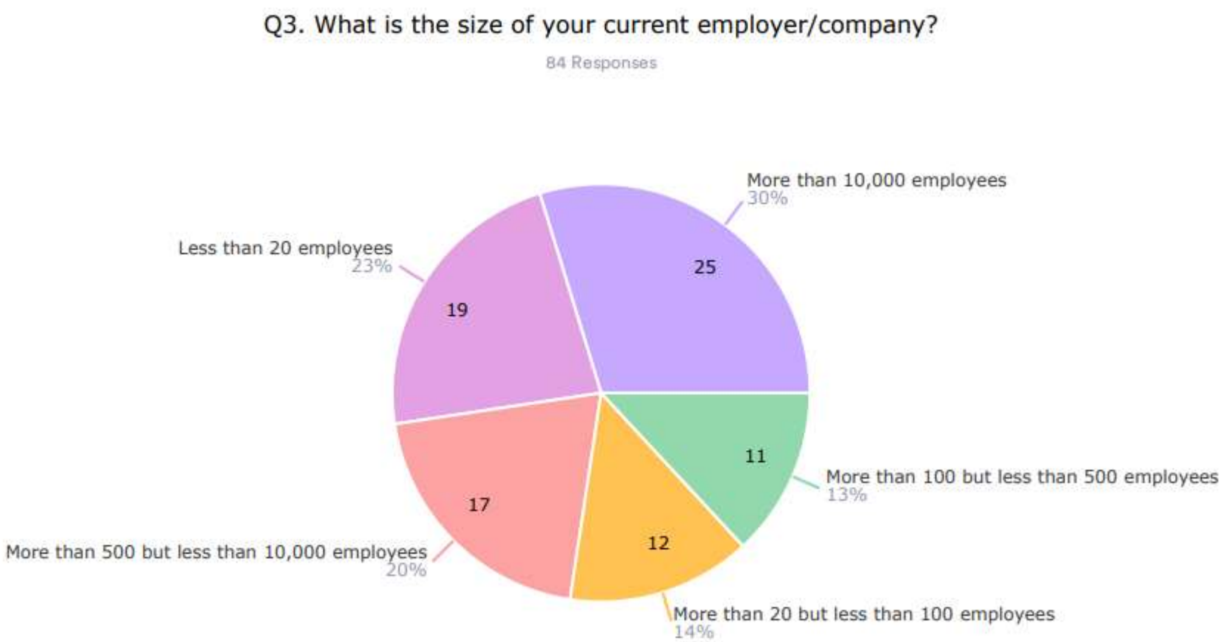

Figure 6: Size of companies

Figure 7 of fifth question illustrates the challenges companies face when implementing MLOps. The lack of skilled staff in the MLOps life cycle top up a long list of challenges. In the analysis section, every challenge is discussed with appropriate solutions. Challenges are grouped into three categories: organizational challenges, ML systems challenges, and Operational challenges.

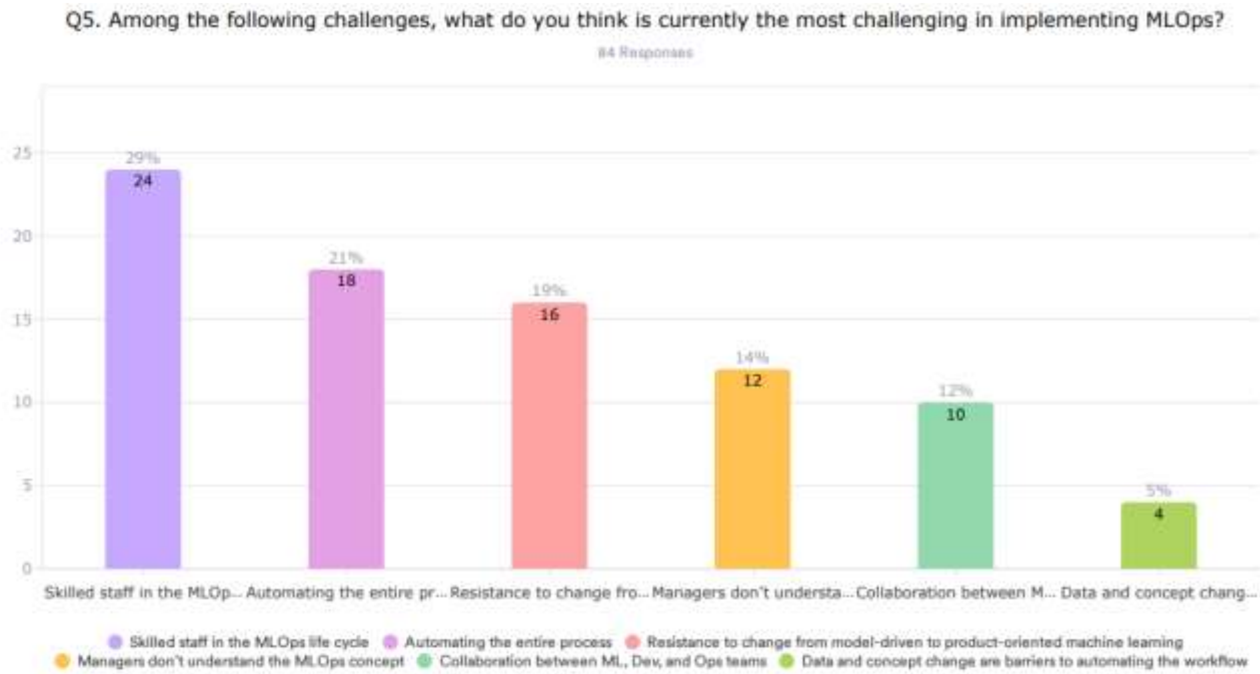

Figure 7: Challenges companies face when implementing MLOps

The feedbacks show the value respondents give to the MLOps in transforming and revolutionizing the machine learning industry. Again in the next analysis section, the researcher extracts insights from our respondents' feedback.

5.2.2. Analyzing the questionnaire data

Figure 8 shows that many respondents agree that MLOps is more concerned with combining the work of ML engineers and developers to automate ML products. This shows how automating ML works is important in revolutionizing the ML industry. As the concept is still new for many professionals, respondents do not want to agree with the statement strongly.

Question 6 was “To what extent do you currently agree with the following statements? MLOps is more concerned with combining the work of ML engineers and developers to automate ML products.”

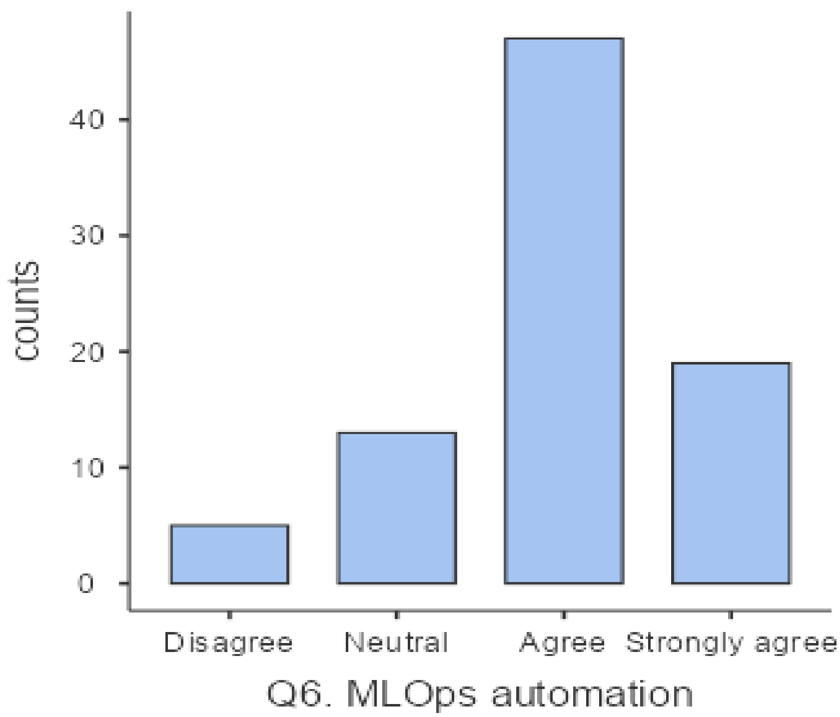

Figure 8: A plot of MLOps automation

The results show that out of 84 respondents, 47 (56%) agree with the statement, 19 (23%) strongly agree, 13(15%) are neutral, and 5 (6%) disagree with the statement.

In the interview section, we have seen that companies face different challenges in the MLOps adoption journey. Even though development is still underway in many companies and the challenges they face, professionals feel very positive about the contribution MLOps is bringing to the industry.

In figure 9, the question was MLOps adoption and development are still underway in many companies. What is your overall feeling about MLOps?

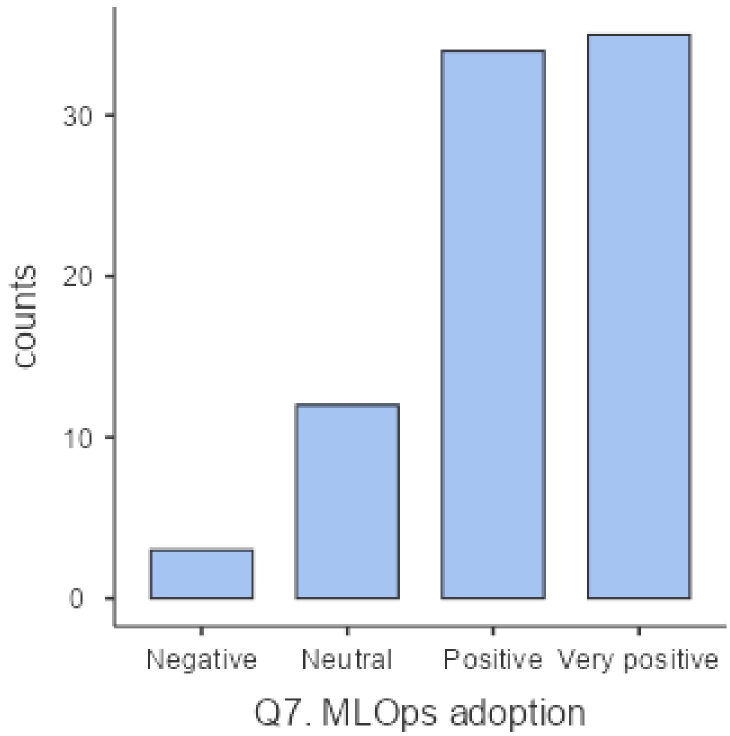

Figure 9: A plot of MLOps adoption

According to the survey results, 35 (42%) of the respondents are very positive about MLOps, 34 (40%) are positive, 12 (14%) are neutral, and 4 (4%) are negative about MLOps.

From the beginning of the study, the researcher is concerned to see if MLOps is revolutionizing the ML engineering industry. Figure 10 shows that respondents strongly agree that MLOps is revolutionizing the ML engineering industry.

Question 8 was To what extent do you agree with the following statement? MLOps is revolutionizing Machine Learning Engineering.

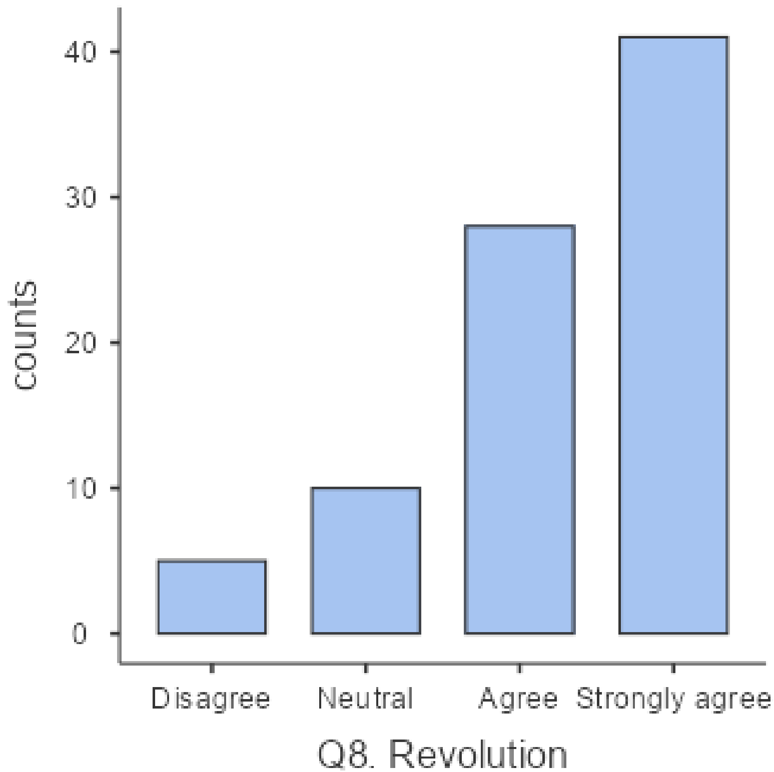

Figure 10: A plot shows how MLOps is revolutionizing ML field

According to survey results, 41 (49%) of respondents believe (strongly agree) that MLOps is revolutionizing machine learning engineering while 28 (33%) of respondents agree with the statement.

Figure 11 of the density plot below shows that professionals from different companies have almost the same understanding of the use of MLOps and its contribution to revolutionizing the machine learning engineering field. They support the idea that MLOps is revolutionizing the ML engineering sector.

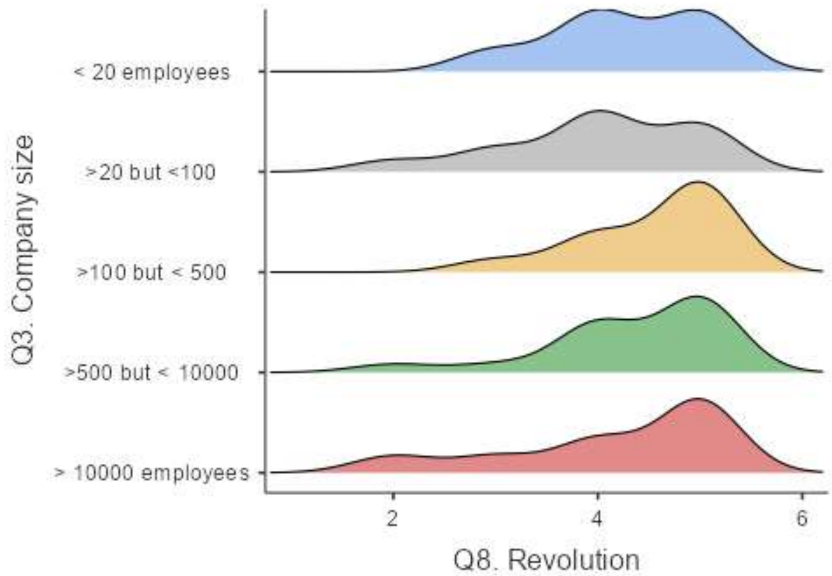

Figure 11: A density plot shows how MLOps is revolutionizing ML field

Analysis shows that big companies adopt and implement MLOps more than small and medium companies. As we discussed above, the implementation and adoption of MLOps require companies to have enough resources in terms of skilled staff and necessary budgets. It is easier for established companies to afford the cost of MLOps implementation than startups or growing companies.

6. FINDINGS AND DISCUSSION

6.1. MLOps implementation challenges and solutions

According to conversations with professionals during interviews, companies may find it challenging to implement MLOps, but building and maintaining reliable machine learning models in production environments is critical. MLOps is a set of best practices and tools that allow businesses to develop, deploy, evaluate, and sustain large-scale machine learning models. However, there are several obstacles that businesses may encounter when implementing MLOps. Lack of collaboration among data scientists, machine learning engineers, software developers, and operations teams is one of the most significant challenges in MLOps implementation. This can result in siloed teams, lack of communication, and model deployment delays.

Businesses can form cross-functional teams with defined roles and responsibilities to address this issue and establish a communication and collaboration process, such as regular meetings, joint planning, and work documentation. Another issue is reproducibility, which is difficult because models can be sensitive to small changes in data or environment. Companies can use version control to manage code, data, and model artifacts to ensure reproducibility. They can define a clear workflow for developing, training, and deploying models and use containers to ensure consistency across environments. Another challenge in MLOps implementation is monitoring. Companies must monitor model performance metrics, data drift, and other production issues. To address this, they can set up monitoring and alerting systems and automated testing and validation to detect issues as they arise. Deployment can also be complex due to integrating with existing systems, managing dependencies, and ensuring scalability. Companies can address this by containerizing models and managing dependencies with tools like Kubernetes and Docker. They can use infrastructure as code to automate deployment and configuration and CI/CD to speed up the deployment process. Because machine learning models can potentially expose sensitive data or be vulnerable to attacks, security and privacy are significant concerns in MLOps implementation. Access control, encryption, and data anonymization are security and privacy measures businesses can use to address this challenge. They can use secure coding practices and secure deployment best practices.

The following tables show the synthesized challenges and corresponding solutions. There are grouped into three categories; organizational challenges, ML systems challenges, and operational challenges. Synthesized challenges and their corresponding solutions are results from interviews and questionnaire data.

6.1.1. Organizational challenges

Table 1: Organizational challenges and solutions

MLOps implementation challenges	**Solutions**
Inadequate cross-functional collaboration	To ensure MLOps is aligned with business goals, encourage collaboration among data scientists, IT, and business stakeholders by maintaining regular communication and holding joint training sessions. Create cross-functional collaboration with clear roles and responsibilities.
Change resistance	To overcome resistance to change, provide education and training on the benefits of MLOps, and implement change management strategies
A scarcity of resources and qualified professionals	Invest in data science, machine learning engineering talent, cloud platforms, and pre-built MLOps tools to reduce the strain on internal resources.
Inadequate process alignment	Ascertain that MLOps is integrated with existing IT processes and infrastructure and that all stakeholders are on board with the MLOps process.

6.1.2. ML systems challenges

Table 2: ML systems challenges and solutions

Challenge	Solution
Data management 1. Problems with data quality 2. Data silos and fragmentation 3. Fairness and bias in data 4. Data drift 5. Data governance and privacy issues	1. Implement data validation and cleaning procedures, and ensure that data is appropriately documented and versioned. 2. Implement a data integration strategy that combines all relevant data sources and tracks data lineage and provenance using a data catalog. 3. Use bias detection and debiasing techniques, and ensure diversity and representation in the data used for training. 4. Retraining, remodeling, removing drifting features, and sample reweighting are some methods for dealing with data drift. 5. Create data governance policies compliant with regulations such as the GDPR and the CCPA, and ensure that data privacy is considered throughout the MLOps lifecycle.
Model development 1. Ad-hoc model development 2. Slow model training and testing 3. Model deterioration and drift 4. Model versioning and management	1. Use version control for code and models and implement a model development framework to standardize the process and ensure reproducibility. 2. Implement automated testing and validation to detect errors early and use techniques like distributed training and model parallelism to accelerate training. 3. To maintain model performance over time, keep an eye on the currently used models and put strategies like retraining, model versioning, and A/B testing into practice. 4. Utilize version control to manage models, track changes, and ensure models are tested properly and documented.
Model deployment 1. Difficulty in deploying models to production 2. Deployment complexity 3. Model explainability and interpretability	1. To speed up deployment and lower error rates, use automated deployment tools like Kubernetes. 2. Utilize containerization and orchestration tools like Docker and Kubernetes to ensure scalability and simplify deployment. 3. Ensure models are understandable and interpretable using feature importance analysis and model explainability frameworks.

6.1.3. Operational challenges

Table 3: Operational challenges and solutions

Challenge	**Solution**
Model performance evaluation	Use monitoring tools to track model performance, detect anomalies, and log and alert to notify stakeholders of problems.
Inadequate standardization of frameworks and tools	Adopt standardized tools and frameworks integrated with existing IT processes and infrastructure.

MLOps implementation presents several challenges businesses must overcome to succeed. The challenges include a need for collaboration, skilled professionals, reproducibility, monitoring, deployment, security, and privacy. Companies can overcome these challenges by developing transparent processes, encouraging collaboration, implementing automation and monitoring tools, and ensuring security and privacy. They can successfully deploy and maintain machine learning models at scale this way.

6.2. Is the MLOps a game changer in Machine Learning Engineering?

Based on the results from interviews and questionnaire data, there is no doubt about the importance of MLOps in transforming machine learning engineering industry. By leveraging the best software engineering, ML engineering, and operations practices, MLOps has brought a new level of maturity and reliability to machine learning development, making building, testing, deploying, and maintaining models easier and faster. Using automated tools and processes, MLOps has reduced the manual work required in the MLOps life cycle, allowing teams to focus on higher-level tasks such as feature engineering and model optimization. MLOps is also revolutionizing machine learning engineering by improving model reproducibility and reliability. In the conclusion section, more arguments support the idea that MLOps is the game changer in ML engineering.

6.3. MLOps pipeline to implement product recommendations on e-commerce platforms

Based on the survey and interviews conducted, figure 12 illustrates step-by-step the MLOps pipeline to implement product recommendations on e-commerce platforms. The combination of the above solutions contributes to designing an adequate MLOps pipeline to implement product recommendations on e-commerce platforms.

6.3.1. Data Pipeline

Data collection: The first step is to gather information from various sources, including clickstream data, purchase history, product metadata, and customer behavior. The data can be collected in batch or real-time. *Data preprocessing*: The collected data needs to be preprocessed to eliminate any anomalies or inconsistencies and to convert the data into a format that can be used for machine learning. The data must be cleaned, filtered, and transformed in this step. *Feature engineering*: This step entails choosing and developing the features the machine learning models will use to produce recommendations. In feature engineering, new features are created based on selecting relevant products and customer attributes.

6.3.2. Model pipeline

Model development: Machine learning models are created in this step to produce product recommendations. There are various models, including hybrid, content-based, and collaborative filters. The models are trained on historical data using algorithms like matrix factorization, neural networks, and decision trees. *Model evaluation*: An evaluation must be conducted to determine the precision and efficiency of the trained models in producing recommendations. The performance of the models is measured using evaluation metrics like precision, recall, and F1 score.

6.3.3. Deployment pipeline

Model deployment: The models are evaluated before being deployed to a production environment where they can produce real-time user recommendations. The models can be deployed on cloud platforms like AWS or GCP, and the deployment can be carried out using containerization technology like Docker. *Model validation*: Compare the new model's performance to the current recommendation engine using A/B testing. To evaluate the new model's performance, use metrics like click-through rates, conversion rates, and revenue generated. *Model monitoring and maintenance*: The models must be monitored and maintained after deployment to guarantee they are operating at their best. Metrics like model accuracy, response time, and throughput are monitored in this process. It is necessary to retrain the model with new data if its performance deteriorates over time.

Figure 12 shows step-by-step the data flow during the product recommendation on e-commerce platforms. Specific tools and technologies are used to implement the process at every step. This study limits its scope to the steps and pipeline.

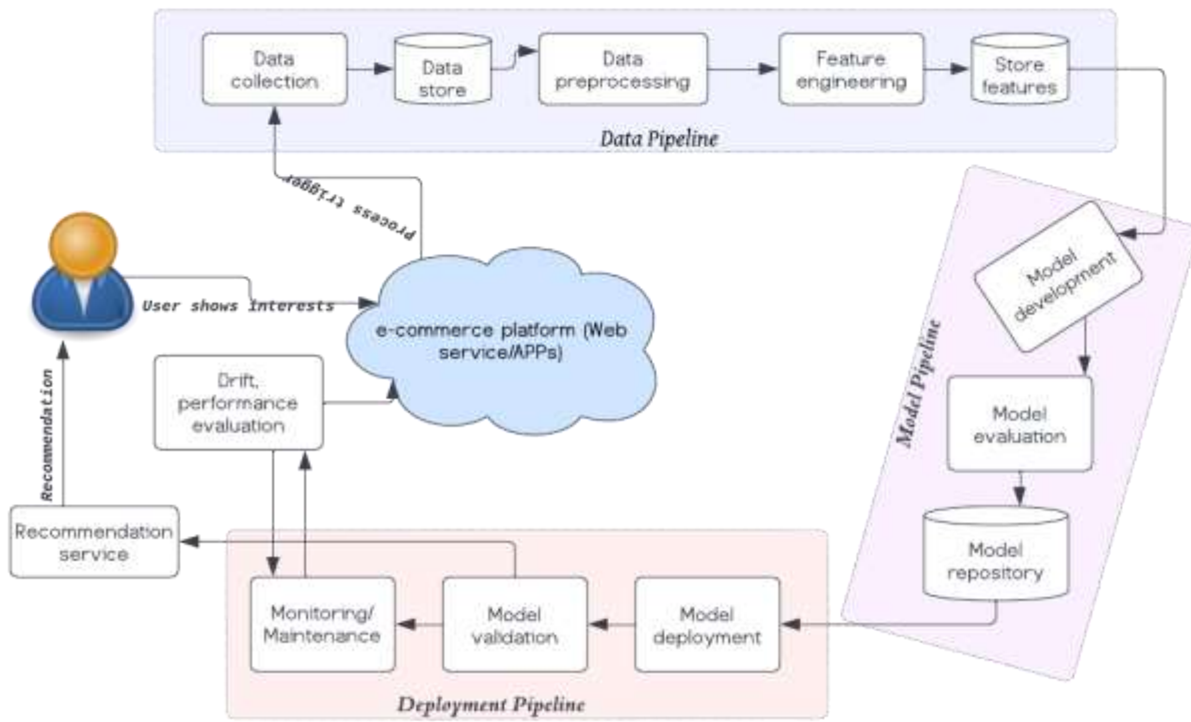

Figure 12: Data flow diagram for products recommendation on e-commerce platforms

7. CONCLUSION

MLOps has transformed machine learning engineering by enabling systematic and automated management of the complete machine learning process, from development to deployment and maintenance. MLOps has brought a new level of maturity and reliability to machine learning development by leveraging the best software engineering and operations practices, making building, testing, and deploying models easier and faster. MLOps is revolutionizing machine learning engineering by facilitating better collaboration among data scientists, machine learning engineers, software developers, and operations teams. These teams can use MLOps to collaborate seamlessly, with defined roles and responsibilities, to build and deploy machine learning models that meet the needs of the business.

MLOps also makes managing every aspect of the machine learning workflow easier, from data pre-processing to model training to release and monitoring. MLOps has reduced the manual work required using automated tools and processes, allowing teams to focus on higher-level tasks such as feature engineering and model optimization. MLOps is also revolutionizing machine learning engineering by improving model reproducibility and reliability. MLOps uses version control, automated testing, and continuous integration and deployment pipelines to ensure that models are built consistently and efficiently, lowering the risk of errors or bugs.

Furthermore, MLOps is assisting in addressing the issues of model deployment and maintenance. MLOps makes deploying models in various environments easier, from on-premises data centers to cloud-based platforms, by leveraging containerization technologies such as Docker and Kubernetes. MLOps also includes automated monitoring and alerting systems that allow teams to detect issues with production models and take corrective action quickly. MLOps has a bright future as more organizations recognize the benefits of using this approach to machine learning engineering. As machine learning becomes more predominant across industries and uses cases, MLOps will play a growing role in assisting organizations in delivering reliable and scalable models that drive business value. The emergence of more specialized tools and platforms designed specifically for machine learning engineering is one of the key trends in the future of MLOps. These tools will make it even easier to implement MLOps best practices and automate essential tasks like data preparation and model training. Another trend in machine learning models is the increasing importance of explain ability and interpretability. MLOps will play a critical role in ensuring that models are built and deployed in a way that meets these requirements as organizations strive to build models that are transparent and understandable to end users. Advances in AI and machine learning research will also shape the future of MLOps. As new algorithms and techniques are developed, MLOps must evolve to incorporate these innovations and allow organizations to capitalize on the most recent advancements in the ML field.

To conclude, MLOps is a game changer and is reshaping machine learning engineering by offering a systematic and automated approach to managing the entire machine learning lifecycle.

7.1. Future research

It is an opportunity for future researchers to investigate the tools used in each pipeline of the MLOps life cycle. MLOps practices can be carried out using a variety of tools. The scope of this study was limited to discussing the steps to take during MLOps implementation, identifying open challenges that companies face when implementing MLOps, and proposing solutions to these challenges. Another future project could stimulate the entire MLOps life cycle, from data collection, preprocessing, model development, model deployment, and maintenance. Moreover, apply data and model versioning principles in the automated production line.

List of references

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

Alshenqeeti, H. (2014). Interviewing as a data collection method: A critical review. English linguistics research, 3(1), 39-45.

Andrew, N. (2022, December 16). *Machine Learning Engineering for Production (MLOps) Specialization*. Deeplearning.ai. Retrieved January 3, 2023, from https://www.deeplearning.ai/courses/machine-learning-engineering-for-production-mlops/

Battina, D. S. (2019). An Intelligent Devops Platform Research And Design Based On Machine Learning. training, 6(3).

Braun Virginia & Clarke Victoria (2006) Using thematic analysis in psychology, Qualitative Research in Psychology, 3:2, 77-101, DOI: 10.1191/1478088706qp063oa

Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med. 2018 Mar 15;378(11):981-983. doi: 10.1056/NEJMp1714229. PMID: 29539284; PMCID: PMC5962261.

Creswell, J. W. (2009). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. (3rd Ed.). Thousand Okas, CA: Sage.

Garg, S., Pundir, P., Rathee, G., Gupta, P. K., Garg, S., & Ahlawat, S. (2022). On Continuous Integration / Continuous Delivery for Automated Deployment of Machine Learning Models using MLOps. arXiv.org. http://arxiv.org/abs/2202.03541

Hewage, N., & Meedeniya, D. (2022). Machine Learning Operations: A Survey on MLOps Tool Support. arXiv.org. https://doi.org/10.48550/arXiv.2202.10169

John M. M. , H. H. Olsson and J. Bosch, "Towards MLOps: A Framework and Maturity Model," 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Palermo, Italy, 2021, pp. 1-8, doi: 10.1109/SEAA53835.2021.00050.

Kreuzberger, D., Kühl, N., & Hirschl, S. (2022). Machine Learning Operations (MLOps): Overview, Definition, and Architecture. arXiv.org. https://arxiv.org/pdf/2205.02302

Lemp, J. D., & Kockelman, K. M. (2012). Strategic sampling for large choice sets in estimation and application. Transportation Research Part A: Policy and Practice, 46(3), 602-613.

Niemelä, P., Silverajan, B., Nurminen, M., Hukkanen, J., & Järvinen, H. M. (2022). LAOps: Learning Analytics with Privacy-aware MLOps. In CSEDU (2) (pp. 213-220).

OpenAI, 2023. GPT-4 Technical Report, https://cdn.openai.com/papers/gpt-4.pdf

Palmes, P. P., Kishimoto, A., Marinescu, R., Ram, P., & Daly, E. (2021). Designing Machine Learning Pipeline Toolkit for AutoML Surrogate Modeling Optimization. arXiv preprint arXiv:2107.01253.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., ... & Dennison, D. (2015). Hidden technical debt in machine learning systems. Advances in neural information processing systems, 28.

Sweenor, D., Hillion, S., Rope, D., Kannabiran, D., Hill, T., & O'Connell, M. (2020). ML Ops : operationalizing data science : four steps to realizing the value of data science through model operations (First edition.). O'Reilly Media. ML Ops: Operationalizing Data Science | ML Ops: Operationalizing Data Science (oreilly.com)

Symeonidis, G., Nerantzis, E., Kazakis, A., & Papakostas, G. A. (2022, January). MLOps-definitions, tools and challenges. In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0453-0460). IEEE.

Testi, Ballabio, M., Frontoni, E., Iannello, G., Moccia, S., Soda, P., & Vessio, G. (2022). MLOps: A Taxonomy and a Methodology. IEEE Access, 10, 63606–63618. https://doi.org/10.1109/ACCESS.2022.3181730

EU, GDPR https://eur-lex.europa.eu/eli/reg/2016/679/oj

RESEARCH APPENDIX

Questionnaire

Online survey questionnaire

Q1. Which of the following best describes your current position (background)?

1. MLOps engineer
2. ML engineer
3. Data Scientist
4. Data engineer
5. DevOps engineer
6. Software engineer
7. Backend developer
8. AI Architect
9. Other

Q2. How many years of experience do you have in the field?

1. Less than five
2. Between five and ten
3. More than ten

Q3. What is the size of your current employer/company?

1. Less than 20 employees
2. More than 20 but less than 100 employees
3. More than 100 but less than 500 employees
4. More than 500 but less than 10,000 employees
5. More than 10,000 employees

Q4. Do you use MLOps in your company? *Yes or No*

Q5. Among the following challenges, what do you think is currently the most challenging in implementing MLOps?

1. Skilled staff in the MLOps life cycle
2. Resistance to change from model-driven to product-oriented machine learning
3. Managers don't understand the MLOps concept
4. Data and concept change are barriers to automating the workflow
5. Collaboration between ML, Dev, and Ops teams
6. Automating the entire process
7. Cost to develop robust MLOps workflow

Q6. To what extent do you currently agree with the following statements?

MLOps is more concerned with combining the work of ML engineers and developers to automate ML products.

- *Strongly disagree – Disagree* - *Neutral* - *Agree* - *Strongly agree.*

Q7. MLOps adoption and development are still underway in many companies. What is your overall feeling about MLOps?

- *Very negative* - *Negative* - *Neutral – Positive* - *Very positive.*

Q8. To what extent do you agree with the following statement?
MLOps is revolutionizing Machine Learning Engineering.

- *Strongly disagree – Disagree* - *Neutral* - *Agree* - *Strongly agree.*

Q9. Any other thoughts about MLOps? Feel free to share.

Interview questions

1. Can you explain in a few words what it's MLOps, and why it is becoming so important in the field of ML?
2. How does MLOps differ from traditional DevOps?
3. What are some of the key challenges in implementing MLOps? And how to overcome these challenges?
4. What are the biggest advantages of incorporating MLOps into a company's machine learning workflow?
5. How does MLOps help to improve the collaboration and communication among the ML engineers, Data scientists, and operations teams?
6. How do you ensure the robustness and reliability of models once they are deployed in production?
7. How do you see the future of MLOps?
8. What technologies will shape the future of MLOps?
