WHAT ARE THE KEY AREAS OF ML-OPS / DL-OPS IN BUSINESS PROBLEMS FOR COMPANY GROWTH USING CLOUD ENVIRONMENT?

by

Gokul Chandrakant Talele,

M.Sc. Biostatistics, M.Sc. Data Science

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

MARCH 2024

WHAT ARE THE KEY AREAS OF ML-OPS / DL-OPS IN BUSINESS PROBLEMS FOR COMPANY GROWTH USING CLOUD ENVIRONMENT?

by

Gokul Chandrakant Talele,

APPROVED BY

dr. Jaka Vadnjal
<Chair’s Name, Degree>, Chair

<Member’s Name, Degree>, Committee Member

<Member’s Name, Degree>, Committee Member

RECEIVED/APPROVED BY:

Admissions Director

Dedication

I wholeheartedly dedicate this thesis to my beloved wife, Manjusha, and our precious children, Pauravi and Reyaansh. Your constant love, encouragement, and support have been my beacon throughout this academic voyage. Your faith in me, coupled with your countless sacrifices, has been the cornerstone of my perseverance, helping me navigate through the myriad of challenges. This accomplishment is as much a testament to your contributions as it is to my efforts. I am profoundly grateful for being my endless source of inspiration and motivation.

Gokul Chandrakant Talele

Acknowledgments

I wish to express my deepest gratitude to my thesis advisor, Dr. Mario Silic., PhD. His invaluable guidance, steadfast support, and encouragement were pivotal throughout the journey of my thesis. His profound knowledge, insightful feedback, and dedication played a crucial role in shaping the direction and caliber of my work.

I am also thankful to the members of my thesis committee. Their constructive feedback, insightful suggestions, and generous sharing of expertise significantly enriched and solidified my thesis.

My heartfelt appreciation goes to the Swiss School of Business and Management Geneva. They provided the essential resources, facilities, and opportunities fundamental to my research and academic growth.

I owe a profound debt of gratitude to my family for their constant love, encouragement, and understanding. Their unwavering support has been my cornerstone, inspiring me to overcome challenges and pursue excellence.

I also want to express my sincere thanks to my friends and colleagues. Their encouragement, companionship, and intellectual exchanges have greatly enhanced my academic journey and made it more fulfilling.

Lastly, I acknowledge the vast number of researchers, scholars, and practitioners whose works have inspired and informed my research.

This thesis stands as a testament to the collective support, wisdom, and encouragement from all the individuals mentioned. I am deeply thankful to each one for their invaluable contribution to this significant academic endeavor.

Gokul Chandrakant Talele

ABSTRACT

WHAT ARE THE KEY AREAS OF ML-OPS / DL-OPS IN BUSINESS PROBLEMS FOR COMPANY GROWTH USING CLOUD ENVIRONMENT?

Gokul Chandrakant Talele
March 2024

Dissertation Chair: Dr. Mario Silic., PhD

This research delves into the exploration and comprehension of Machine Learning (ML) operations within industrial environments, highlighting the growing necessity for data-driven organizations to adopt AI and ML. Operating and maintaining ML models in industrial production settings pose significant challenges. The integration of DevOps principles has revolutionized how software engineers release products, fostering efficiency and creativity. A parallel trend is observed in the machine learning sphere, where data science teams are beginning to integrate these principles into ML operations, termed MLOps. This literature review aims to shed light on the current obstacles encountered in the productionization of machine learning, drawing upon academic sources to scrutinize the prevailing challenges in MLOps. The focus is twofold: firstly, on the critical role of MLOps principles in industrial contexts, and secondly, on the application of these DevOps principles to enhance the operationalization of machine learning projects.

However, in recent years, the use of Machine Learning (ML) has witnessed a significant increase, but many organizations still face challenges when operationalizing ML. This thesis investigates the current best practices, challenges, and potential solutions associated with developing an MLOps process in the cloud from a RE perspective by exploring the intersection between machine learning operations (MLOps) and Requirements engineering (RE). This thesis aimed to create an artifact that would guide MLOps implementation in the cloud from an RE perspective, thus offering a more systematic approach to managing ML models in production by establishing goals and attitudes toward their development in the future. Three research questions were investigated using the Design Science Research methodology during design artifact creation. The study examined existing barriers to the MLOps process's establishment, found possible ways to overcome these difficulties, and assessed the efficiency. The study followed three cycles, each answering all the research questions but mainly concentrating on one question, allowing initial artifact creation and subsequent refining depending on data collected during each process. By establishing a better in-depth understanding of how these two spheres interact and providing some practical guidance for implementing MLOps processes from the RE perspective, this study advances both MLOps and RE fields. In terms of theoretical evaluations, quality feedback was collected about the artifact. In this regard, one major limitation is that an assessment of the artifact's efficiency in real-life situations needs to be made. As a result, future research should evaluate this artifact's effectiveness by conducting case studies in real-world settings and enhancing its limitations.

Keywords

Machine learning operations, Machine learning, Requirements engineering, ML, RE, MLOps, Design science research, cloud

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I:

INTRODUCTION

1.1 ML-Ops and DL-Ops: Vital for Cloud Company Growth

Numerous organizations, including tech giants like Google and Amazon, are leveraging machine learning (ML) and deep learning (DL) to enhance their services, like creating personalized recommendations by analyzing customer data (Weber,2019). However, building and maintaining these ML/DL models is complex, requiring significant data scientist effort due to ongoing development and updates (Capizzi,2020). MLOps, which integrates ML development and operations, is essential for managing this lifecycle efficiently, drawing on DevOps successes to streamline model training, deployment, monitoring, and management.

MLOps is becoming crucial in data science, aiding organizations to derive long-term value and manage risks in ML/DL/AI projects (Sculley,2015). Many firms struggle with developing and deploying multiple models efficiently. As automated decision-making grows in importance, effectively managing model risks becomes essential. Major companies like Amazon and Google offer MLOps solutions that centralize data storage and processing, enabling businesses to streamline their data operations and reduce dependency on local data centers (Karlaš et al.,2020).

To achieve rapid and substantial company growth, leveraging data-driven insights, as well as machine learning (ML) and deep learning (DL) technologies, is crucial. This necessity arises from the ongoing advancements in technology. Industry giants like Google, Amazon, Facebook, and Twitter have underscored the importance of ML and DL, showcasing their effectiveness in analyzing customer purchasing patterns and offering personalized product recommendations. Empirical evidence suggests that such approaches can significantly improve overall company performance. However, machine learning scientists face various challenges, such as model monitoring and version control, when implementing ML or DL models in real-world scenarios. They must meticulously manage previous versions and stay abreast of ongoing developments and modifications in model iterations. Maintaining a single model demands considerable effort and attention to detail. Therefore, it is imperative to adopt MLOps throughout the entire model lifecycle, with a specific emphasis on addressing these challenges.

The research emphasizes the importance of a specific approach in managing machine learning (ML) models, as outlined by Weber (2019). This thesis explores how ML-Ops and DL-Ops can address business challenges, driving growth in a cloud-based environment. It integrates insights from relevant studies and industry expertise, aiming to optimize strategies for adopting MLOps. ML and DL are potent tools for innovation and competitiveness, necessitating a comprehensive process model to manage their full lifecycle effectively. (Karlaš et al.,2020) stress the integration of continuous integration and deployment processes into ML-Ops and DL-Ops workflows, promoting stability and efficiency. Cloud-based MLOps solutions, such as Google's TFX platform, offer scalability and flexibility, mitigating challenges like data security and processing power limitations. Collaboration among data scientists, developers, and operations teams, supported by automation and integration of various technologies, enhances productivity and decision-making. Bergstra et al. (2015) introduce Hyperopt, facilitating hyperparameter optimization crucial for model performance. As ML and DL adoption grows, managing technical debt and integrating MLOps with established software engineering principles become imperative. The thesis employs a Design Science Research approach to address gaps in understanding and implementing requirements engineering for MLOps. By leveraging RE principles, organizations can realize similar benefits observed in traditional software development. The thesis structure includes problem statement elucidation, literature review, methodological insights, and findings analysis, offering avenues for further research and conclusive judgments.

1.2 Problem Statement

Integrating machine learning (ML) in the industry can yield a multitude of advantages (Capizzi,2020; Dotscience Blog,2022; Sculley,2015). Nevertheless, several firms may face challenges when implementing machine learning in their production processes (Knauss,2021). The challenges include meeting the criteria for effective and economically beneficial models, monitoring model performance and accuracy, and requiring data validation and preprocessing (Guru99,2019). To surmount these challenges and guarantee the triumphant execution of machine learning (ML), a methodical and effective strategy, such as MLOps, is necessary to oversee the progression, deployment, and upkeep of ML models. According to the literature study conducted in this thesis, no research has been published yet on the methodology of implementing requirement grounded MLOps. Prior studies have investigated the use of RE (needs engineering) for ML (machine learning) (Miao,2017; Villamizar,2021). However, due to the distinct needs of individual ML models compared to ML pipelines and the broader MLOps process, further investigation is necessary in this domain. There may be differences in prerequisites between ML models and MLOps procedures, such as: When training ML models, it is crucial to carefully select the data to be used. When considering MLOps operations, it is critical to determine the frequency of uploading the training data to the ML or data pipeline.

Considered one of the technologies that have contributed to the rising value of businesses in recent years, machine learning (ML) views data as a vital resource for making more flexible, logical, and occasionally even automated judgments. The advent of big data technologies, which allow for the large-scale development of new algorithms, and the rise in data generation over the past ten years have been the main drivers of this (Adadi, 2021). According to estimates by Jones et al. (2022), there will be seven billion linked devices in 2022 and 22 billion by 2025.

Furthermore, the increasing ubiquity and declining cost of particular hardware technologies, such as graphics processing units (GPUs), will create a favorable atmosphere for encouraging businesses to explore artificial intelligence (AI) technologies.

However, it is important to go beyond the model-building stage to reap the benefits of machine learning. The machine learning life cycle still needs to address a number of issues. For example, data scientists use a variety of strategies, techniques, and algorithms, such as regression, classification, optimization, or clustering, to build mathematical or machine learning models that increase computers' intelligence. They are good at making these kinds of models.

As Romero (2020) emphasizes, software engineers typically need to acquire software engineering skills to deploy their work in real-world production settings. This is a result of their usual strong foundation in mathematics. However, the skills needed are different, requiring, among other things, a thorough understanding of various environments, file formats, protocols, and networks, creating a distributed, scalable, and fault-tolerant application, or applying particular strategies to improve and expedite the software development life cycle.

The machine learning operations (MLOps) paradigm has arisen to support this idea. As Alla and Adari (2021) noted in (Alla, 2021), MLOps can be considered the meeting point between DevOps and machine learning methodologies. According to Leite (2021), DevOps is a collection of techniques designed to enhance the software life cycle and ensure continuous delivery. Similarly, MLOps focuses on leveraging the DevOps viewpoint to manage high-performance machine learning models, allowing for continuous delivery to improve their life cycle and, consequently, alleviating the responsibilities of data engineers and analysts.

One of the primary challenges for MLOps is the quick deployment of machine learning models in cloud systems in production contexts. To complete this MLOps process, they suggest addressing the following issues in the book Cloudera (2022):

A. Model packaging must be considered to automate the machine learning life cycle, which offers a vast tool ecosystem.

B. Model deployment involves providing the model built for production environments and making it accessible to software clients and applications.

C. Model monitoring must be conducted to detect model degradation and performance issues automatically. When anomalous behaviors are identified, models can be retrained.

D. Model governance makes it possible to track models. The primary strategy is to provide a model catalog with which relevant meta information can be connected. As a result, finding and identifying models is simple. A sound model library also makes setting up authorization and authentication rules for the models and related auditing systems easier.

On the other hand, cloud computing, on-premises servers, and the newly emerging cloud computing paradigms provide certain advantages. For example, to reduce latency, some time-critical steps could be executed near the edge (where data are generated) rather of running the entire pipeline in the cloud. Furthermore, as the cloud usually gets its data from edge devices in the same region, it can handle several aggregations. But this strategy also suggests further difficulties. One of them entails allocating computing and storage resources—edge and cloud—adequately and in a way that maximizes analytic pipeline performance and throughput while concurrently minimizing latency and costs. Furthermore, the need for more sophisticated and varied skills makes managing a heterogeneous infrastructure even more difficult.

One additional difficulty is that in certain cases the necessary infrastructure may not yet be in place. It takes an expert software engineer to complete this creation procedure because it is not simple. In addition, automating this process is a great way to improve environment upkeep, conveniently duplicate environments, and prevent human error. Furthermore, the proper execution of the analytic pipelines requires the independent execution—whether or not it is available—of certain applications and configurations, preferably also in an automated manner. As such, the work necessary to accomplish this goal is more substantial.

The proficiency of data engineers is critically important for deploying and summarizing machine learning models or analytic pipelines into testing and production environments. It is also necessary to establish and thoroughly configure target machines prior to initiating the deployment stage. Therefore, it would be quite beneficial to use software clients that are easy to use and automated procedures to make these activities easier for non-expert users. Regretfully, this is typically not the case.

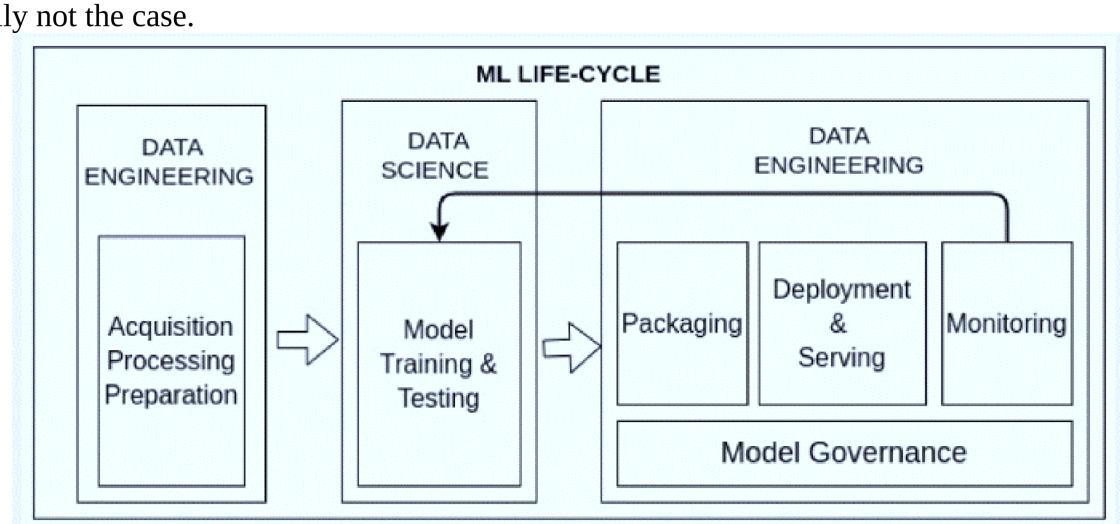


Figure 1.1 The main phases of the ML life cycle are aimed at making ML models work in production environments.

In addition to putting the model into production, additional testing is required to guarantee integration with other systems once models are built frequently. Moreover, the strategy can involve not only the output of a model but also the execution of several chained steps involving acquisition, processing, or data preparation, as shown in **Figure 1**—that is to say, an analytic pipeline. For this purpose, it can be beneficial to test such steps isolated in different hosts to represent the target environment better and to be able to focus on each step of the process separately in the evaluation phase. For these activities, it can be useful to evaluate the entire pipeline to better understand how the processes are acting in an isolated setting by building a development environment and imitating the production one.

The present thesis utilizes a Design Science Research (DSR) methodology to address this knowledge deficit and examine research inquiries concerning RE for MLOps, as illustrated in Section 1.2. In addition, an ongoing parallel endeavor aims to create an artifact that would assist organizations in managing and administrating their MLOps requirements. RE is a critical element in traditional software development as it guarantees that the application sufficiently satisfies the needs and desires of stakeholders and consumers (Hevner, 2007). The process comprises methodically identifying, documenting, and monitoring the features of a product or system (Vogelsang, 2019). Chapter 4 offers an in-depth description of the artifact, with precise and explicit questions on the requirements. It also lists the assigned responsibilities for posing these queries and provides samples of typical responses. Chapter 4 delves more deeply into the specifics of this information. When organizations employ requirements engineering (RE) in MLOps, they can achieve similar advantages as observed in software development.

The structure of this thesis is as follows: The first segment of the dissertation provides a concise exposition of the problem statement, research inquiries, and overarching objective. Subsequently, we will present a comprehensive analysis of existing research and give relevant contextual information. In addition, this study outlines the specific methods used, including design science research (DSR), literature review, interviews, and data analysis procedures. Moreover, the outcomes encompass both the produced product and the discoveries derived from the research inquiries. In conclusion, this thesis includes a comprehensive review of the results, suggests potential avenues for further investigation, and establishes conclusive judgments.

1.3 Research Questions

The study's primary goal is to answer three research questions (RQs), each with a problem statement, an approach to the problem, and an evaluation of the resolution and its usefulness:

Q1: What are the current challenges in designing an MLOps process in cloud environments, and how do they relate to requirements knowledge?

Q2: What potential solutions exist to mitigate the challenges of developing an MLOps process in cloud environments based on requirements engineering?

Q3: How well does the potential solution mitigate the requirements-related problems with developing an MLOps process in cloud environments?

1.4 The Purpose of the Study

This project aims to improve our current knowledge of the intersection of MLOps and RE, with the goal of creating an artifact to guide MLOps implementation using cloud from a RE perspective. To reach this goal, the research will look into current approaches for building MLOps procedures. The study will also look into the current best practices and problems that come with creating an MLOps in cloud environment. Additionally, it will explore the relationship between these problems, understanding the requirements, the available solutions, and the effectiveness of these solutions in mitigating the identified issues. This artifact aims to establish a method for embedding requirements engineering into MLOps procedures, resulting in a more systematic and reliable approach to managing ML models in production.

CHAPTER II:

REVIEW OF LITERATURE

This chapter will thoroughly cover selected information relevant to this thesis by addressing the following topics: MLOps, ML, DevOps, and RE for MLOps. The collected information consists primarily of academic literature. However, more peer-reviewed literature is needed since MLOps is a relatively new topic area. Therefore, we sourced some information from specialty courses by reputable experts and leading industry blogs rather than relying solely on peer-reviewed papers. Strict deliberation was devoted to the sources' selection process because of their provenance; we were utilized solely in situations where peer-reviewed literature was unavailable to provide an equivalent. In conclusion, this thesis offers an account of the sources it employs.

2.1 Introduction to ML/DL

ML is the discipline that uses the data set for learning and develops the best algorithms for prediction. This aspect is known as AI. In classical programming (Fig. 2.1 B), we will have to write an algorithm using programming languages to generate the desired outcome if someone provides the data to a computer. Whereas in the ML framework, we are providing data and products so that it will create an algorithm that might have used a novel or different combination of weights and features (Fig. 2.1A). (James G, 2013; Hastie T, 2009). These are the four most common ways to create learning algorithms for solving different problems: supervised, semi-supervised, unsupervised, and reinforcement learning. (James G, 2013; Hastie T, 2009).

A computer is given a dataset and associated outputs as part of machine learning. The AI algorithm learns about the relationship between the dataset and output. Deductions regarding forthcoming datasets are possible by employing this algorithm. Classical computing (A) gives computers facts and rules to follow. The computer uses the algorithm to determine how to process the dataset to produce results. The world is moving towards a data-centric method. ML/DL is used to gain knowledge in various application areas, such as customer shopping patterns (e.g., recommendation engines), maintenance problems, and medical diagnosis.

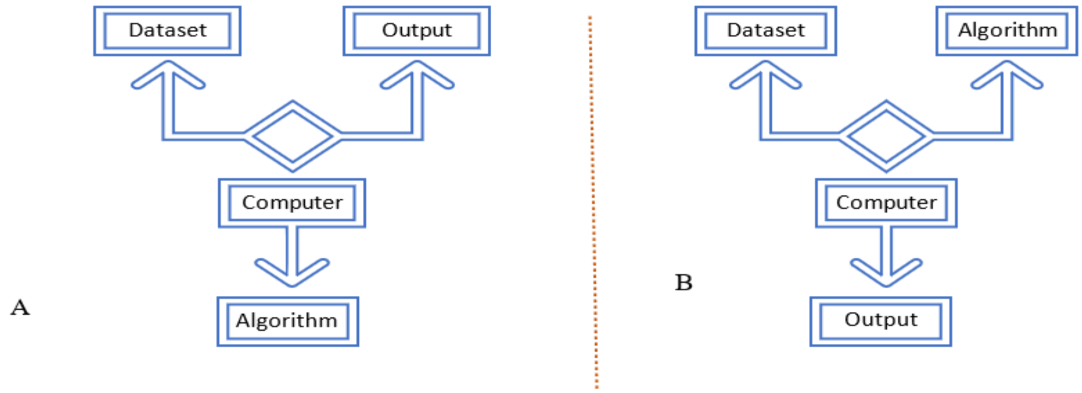


Figure 2.1: Classical programming versus machine learning paradigm. (A)

Each of these use cases has one common feature that is common to all of them: the development of ML/DL models (James G,2013; Hastie T,2009). The ML/DL Model Development life cycle poses a new problem different from DevOps and traditional software development. DevOps or traditional software development considers a defined set of product features. However, data scientists are always trying out new things with a dataset, new tools, hyperparameters, and other items to get the business's best results.

In this literature review on "Machine Learning in Production." It may prompt inquiries like: How exactly does machine learning function in manufacturing environments? How exactly does machine learning change things when applied in both the academic and industrial sectors? There is a clear divide between ML in research and industry. According to recent blogs by "Christopher Tao" ("Christopher's blog") and "George Seif" ("George's blog"), ML in academics and exhibitions focuses on different aspects. An ML researcher's research is more about science and theory; they work to improve the model's accuracy, concentrate on a specific part of the model, or develop models to solve custom scientific problems. The nature of industrial-based ML projects is more about engineering than science. Scalable use of ML systems requires integration, SaaS delivery, and the application of DevOps techniques to provide continuous delivery in production. Structured and clean datasets drive many competitions or challenges in academic research.

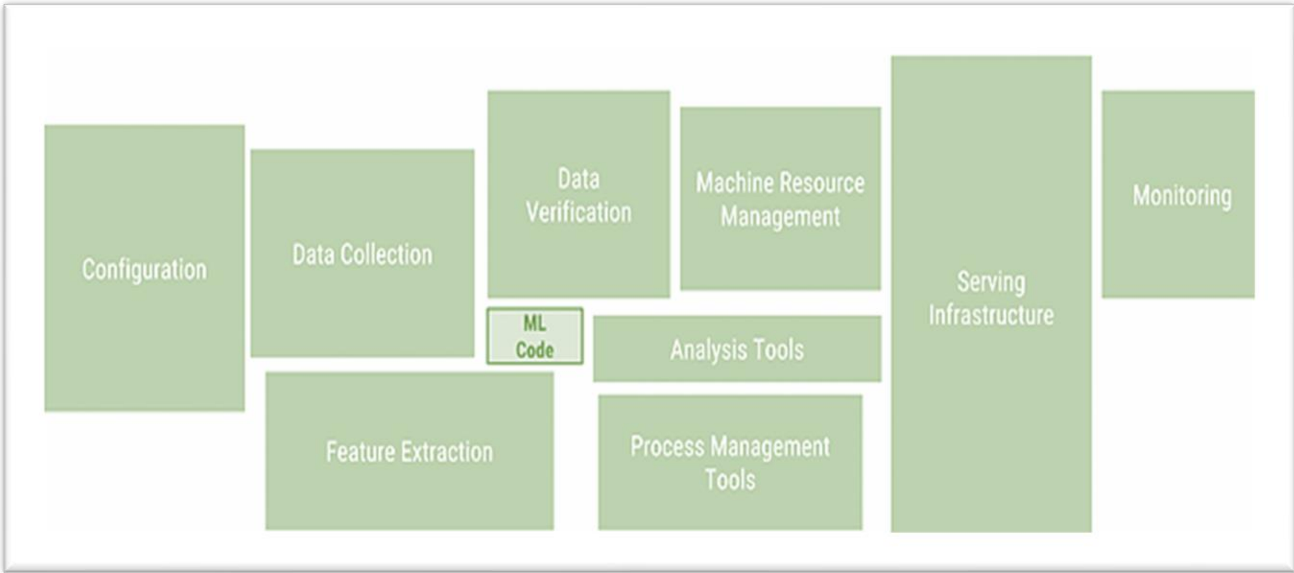


Figure 2.2: There is very little machine learning code in real-world systems. (Scrulleys,2015)

While their primary concern is not computing costs, the accuracy of ML models is of greater importance. Almost in an industrial setting, data is always more complex and unstructured so that the focus may be on something other than algorithm development but instead on data cleansing, formatting, and delivering the output as fast as possible. As shown in Figure 2.2, Scrulleys (Scrulleys,2015) has made a surprising statement that the ML code is a tiny part that gets utilized for production. In contrast to what ML researchers believe, ML systems often include various components beyond machine learning code.

2.2 Machine Learning Projects Face Numerous Challenges

A machine learning system differs from traditional software systems, which require several more components. Management and maintenance of machine learning systems are challenging due to these factors. A machine learning system incurs these long-drawn-out costs, as outlined in Sculley (2015):

- **Management Debt**

Management debt is a management decision with a short-term, practical outcome and a long-term, expensive consequence. Software management methods that deal with code levels rather than systems levels cannot address additional issues specific to machine learning. Within the production environment, second-level machine learning issues can occur. With ML systems, hundreds or even thousands of ML models can run simultaneously, but maintaining and running them in an ML system takes a lot of work. Debt management may become more complex due to multiple methods for assigning and modifying the system configuration to monitor the delays above, simplify experimental outcomes, and ensure comprehensive monitoring of every production line.

- **The Costs of Data Dependency Are Higher Than the Cost of Code Dependency.**

Machine learning development includes both coding and dataset, although traditional software development takes up most of your time while coding. There is a theory that data utilized by ML systems could be unstable. As ML systems learn, the data could become uncertain over time, resulting in a model change. According to this theory, different data sources can result in other behaviors in ML systems, causing us to lose control over them. In an ML system, versioning data is often unnoticed due to its potential value in handling unstable data. If something goes wrong, we can trace any unexpected behavior to the previous version. For example, the maintenance of multiple versions of the system comes at a cost. Furthermore, it is imperative to perform data testing to verify its functionality in light of its volatile characteristics.

- **Complex Modular Boundaries**

Traditionally, software engineers have used modular design to make software easier to maintain. Modular design permits modifications to one element to impact the remaining elements.ML systems cannot be subjected to this approach because one attribute of the machine learning model depends on another. Any alteration in one aspect, such as the significance of distribution, will affect another part.

- **Configuration Debt**

In the production environment, many components come with preinstallation, which can minimize the configuration debt. During the configuration phase, it is imperative to take into account a multitude of factors. These include the utilization of features, preparation of data, details regarding pre- and post-processing, and building feedback loops. It is often costly to make mistakes in the configuration because they cost labor, time, and computing resources. Data management challenges are similar to management challenges, according to Schelter (Schelter,2018). The machine learning lifecycle needs to be accelerated and automated; it is necessary to keep track of the training dataset, validation dataset, and model hyperparameters like accuracy or any other. Tracking experiment results and storing ML objects in a centralized repository are essential to model management, as is managing multi-language codebases. There are usually several different types of objects in the codebase of ML systems, as we discussed in Sculley's paper (Scrulleys,2015). Several programming languages can be selected to develop the program's various components. Many popular libraries for machine learning are written in Python, such as Spark, to deal with the massive scale of data. They may require different configurations because of this characteristic. Data exchange between components must be reliable and efficient to ensure data continuity. In order to leverage ML, data integrity and variability are fundamental requirements, as the algorithm can only acquire knowledge from instruction. Data challenges are similar to those encountered in production when using ML, as stated in the article by Ansreas (A. Mayr,2019). In a machine learning project, data preparation is the most time-consuming process. Some steps are required to achieve industry standards, such as data preparation, cleaning, and validating the model.

2.3 Importance and Difference of MLOps in Production

Machine learning Ops can handle production software systems using DevOps practices, but software development might get delayed in the absence of DevOps, which might take months to release software products. Is it crucial to apply the MLOps concept in the ML pipelines of productions the same way we apply CI, CD, and DevOps to software engineering? The purpose of iterative ML lifecycles is to help us fix issues as they arise throughout the process. According to the article (Dr. Nick Ball,2019), this is one of the benefits of ML in the real world. One compelling rationale for the iterative nature of the machine learning lifecycle is the ability to identify and resolve issues at various stages, including data preparation, model development, production deployment, and more. DevOps does not require a fixed end-to-end workflow. However, there is the requirement of the process, which can help us track versions of different components such as experiments, data, code, results, and models to provide complete control over the system.

The benefits of DevOps include tracking changes and replicating each step of the ML project lifecycle. Then, machine learning will be just like software engineering was in the past before DevOps existed. It is challenging to implement MLOps in the ML lifecycle. Therefore, we need to apply the DevOps methodology to the ML lifecycle; as a result, it is crucial to use MLOps principles as a part of the ML lifecycle. In his paper, Yuri (2020) acknowledged that continuous development, continuous integration, and continuous delivery are the factors that give rise to enterprises adopting the DevOps methodologies. Machine learning systems' development and deployment processes are comparable in complexity to conventional software. However, the ML infrastructure differs from software engineering as it involves different processes, e.g., data preprocessing and cleaning. It is important to note that machine learning workflows consist of many aspects, like models, datasets, matrices, and hyperparameters, which can differ from the traditional DevOps methodology. Tracking datasets, metrics, models, and other resources is also essential. We must track the version of the code. Yuri (2020) mentioned in his paper that there are some differences when we discuss MLOps and DevOps; generally, DevOps is helpful for data engineers while maintaining, building, and improving the ML lifecycle of services. MLOps's data science team mainly focuses on building machine learning models that the production team uses. ML models continue to undergo advancements, experimentation, and continuous improvement even after delivering the merchandise to production. Using MLOps, data scientists can work with the scalability and reliability of data.

2.4 MLOps Methodology Based on DevOps Methodologies

A DevOps system comprises a collection of practices and tools that enable organizations to deliver services and applications more rapidly than with conventional software development. Alternatively, the term DevOps could refer to a set of software engineering practices employed by software engineers. As an alternative, the Agile methodology distinguishes itself through its iterative approach to managing software. Agile and DevOps frameworks enable teams to develop and deliver applications and services perpetually. MLOps demonstrates the applicability of numerous DevOps and Agile principles in an additional domain; thus, Yuri's (2020) article describes this strategy for integrating DevOps into data science initiatives. The research process and its associated methodologies share numerous similarities. The generation of hypotheses, the planning and execution of experiments, the collection and analysis of data, the development of models, the performance of evaluations, and the implementation of enhancements are a few examples. Documenting, retaining, linking, and ensuring the replicability of all stages are often imperative for achieving reproducibility. Continuous refinement is another characteristic that frequently defines data scientist workflows. Karamitsos (2020) emphasized in his paper that implementing the DevOps principle for machine learning systems requires continuous integration and delivery. CI facilitates the validation and testing of machine learning system data and models, in addition to code. It is imperative to incorporate automation and monitoring mechanisms into every stage of the machine-learning procedure. Integrating, releasing, and deploying are all components of production. This article emphasizes the CI and CD phases of the machine learning lifecycle, which they also comprise. Continuous integration (CI) encompasses data preparation, analysis, and cleaning, after which models undergo testing and validation to ascertain their capability to execute the newly added duties. The products also employ models for testing and validation. CD aims to ensure that production environments can access ML models by deploying machine learning objects to staging and production environments. Deployment of machine learning models to staging/production environments occurs after their generation through the CI steps. CD denotes the delivery of machine learning models generated during the CI phase. Deploying the outcomes of this stage will also occur in production environments and the staging environment. Karla (2020) posits that machine learning initiatives like CI and CD can benefit from applying the DevOps mindset. Engineers can automate application construction, testing, and deployment as part of the software development cycle with a continuous integration service. Developers advance their work by iterating on machine learning applications until they are content with the model's quality. They proposed a framework for continuous integration that includes build, test, and release regarding machine learning applications. By executing a code segment, the model will independently initiate its training process utilizing the preprocessed data, parameters, and objects produced during the development phase. Following the construction phase, models are evaluated as part of the model development process and returned to the developers for additional refinement by executing test procedures. To initiate the discharge procedure in CI for machine learning, developers must be content with the utmost possible model results.

In contrast to conventional software testing, testing outcomes are not static. Recent research utilizes the MLOps/DLOps/model life cycle to monitor model development changes in response to dataset modifications. Two scholars, Weber (2019) and Ashmore (2019) provided descriptions of a model life cycle. In contrast, Miao (2017) expounded upon the deep learning methodologies. The following section provides an account of the life cycles of models.

Model Planning: During this preliminary stage, we will establish the aims and objectives of the machine learning project requirements. The engagement of all pertinent stakeholders in the discussion about the projects is analogous to the operation of conventional software development or DevOps.

Model Building and Validating: This phase encompasses various tasks, including data exploration and preprocessing, model training, hyperparameter optimization, and model evaluation. Feature extraction occurs during data preprocessing, which involves data cleansing and transformation. The algorithm's input determines the execution of those stages.

Model Deployment: After setting up the model's features and running tests, the next step is to put the model to use. MLflow is frequently employed with ML models, whereas Apache Spark helps with analytics, bulk processing, and data streaming.

Model Monitoring and Use: - Upon deployment into designated environments, a machine learning model becomes available for utilization by the project stakeholders. When using the model, we can increase the model output performance by training the model online or offline. While doing this, the data scientists can work on the model monitoring step. In addition to monitoring its useability (for example, requests received per second), it is possible to detect drift in basic ideas by examining the input dataset.

Outdated Model: As part of the MLOps process, several factors can make updating the machine learning model necessary. Examples include spontaneous shifts observed in the market sell patterns, poor data quality issues, infeasibility, and economic problems such as the COVID-19 crisis. With the widespread adoption of artificial intelligence (AI) and machine learning (ML), which have expanded exponentially in recent years, the new methods of deep learning and machine learning research are increasing the value of these technologies for an ever-expanding array of organizations and applications. Significant development has occurred in machine learning. Many industries rely on machine learning to support their goals and critical business processes using academic, sample, and competition datasets generated in sandbox environments. As soon as machine learning and artificial intelligence systems reach the assembly servers, they will likely accumulate technical debt, a challenge associated with rapidly implementing innovative technology in the industry production space. Ward Cunningham (Ward Cunningham, 2016) introduced the concept of technical debt. Technical debt is essential for the quick development of new systems, according to Ward Cunningham (Ward Cunningham, 2016). It may be unintended, but unmanageable debt can seriously harm business, individual, and economic health. In his explanation, Ward Cunningham (Ward Cunningham, 2016) used the example of debt as a metaphor to illustrate how 'fast and nasty' methods can create multiple problems inside the code. Consequently, this can lead to more work in the final steps, which can lead to reduced confidence among teammates in the end and an increase in the cost and concern related to debt. Machine learning and AI research communities have introduced technical debt (for example, A person holding a credit card with a higher interest rate) to mitigate the unintentional effects of machine learning in production. In both tutorial and practitioner contexts, extensive research is ongoing. Ward Cunningham (2016) posits that technical debt may arise from many factors. However, in this thesis, we intend to investigate a few of the crucial purposes, how it can be recognized, and what mutual approaches can help to reduce the problem.

A. Why Technical Debt Arise in ML Projects?

When targets and speediness are a priority, teams frequently choose a less-than-optimal solution and implement it quickly with the tools they can access; technical debt results from this in software development. The reason for this may be misdocumented or unstreamlined system requirements and complexity, in addition to concerns regarding code quality, legibility, and testing. In addition to these situations, machine learning and artificial intelligence pipelines are also more susceptible to technical debt due to:

- Models built with machine learning are implemented as black boxes. The lack of explainability and inherent bias in these models is an ongoing research question. As time passes, unintended consequences frequently manifest in production channels predictions or classifications. Due to the absence of comprehension regarding the rationale behind
- Challenges that come with the measurements and scalability problems (the science and art)
- There is a phenomenon known as changing anything, change everything (CACE)
- The environment will likely undergo steady unobserved variations over time due to overfitting, feedback loops, and feedback.
- The reformation of the primary model obscures the status norm of downstream consumers, leading to unpredictable performance.
- Instability in data and features due to high data dependency.
- The capabilities of machine learning are to delve into data from various applications to discover findings from it.
- Due to multiple platforms, languages, and versions, there are jungles of glue code and pipelines. There is an abundance of platforms, technologies, modeling approaches, and programming languages when it comes to AI and ML. Models are frequently used in computer languages other than those for which they originate, putting systems in danger.

There is a problem with misaligning stakeholder groups. Organizational research laboratories frequently launch data science projects after considering the impact on all stakeholders and their business areas.

B. Identifying Potential ML Debt in Ongoing Projects

Consider the technological debt that current implementations have imposed when planning a replacement project.

- Examine these data infrastructures closely for the issues mentioned above.
- Verifying and reviewing the model's scope limits and standards is essential for ensuring their identification. Does the model contain a lot of abstractions?
- How are data collected, analyzed, and used at each level of the system integration in each upstreaming and down-streaming process?
- Maintaining, monitoring, and improving the model daily.

C. Strategies for Mitigating Technical Debt

Some software engineering practices apply to existing and new projects to mitigate technical debt. It is essential to refactor and improve the readability of source code, testing (like regression testing, unit, and integration), and evaluate the configuration, procedures, and tools. It is even possible to mitigate technical debt with ML systems in addition to plain techniques through the following methods:

• Assumptions made in the past can be reassessed and reevaluated to remove the risk of assuming assumptions to be inaccurate or obsolete. Identifying data, concepts, and feature drift can be accomplished through continuous testing. Other methods include testing the equivalence of training models and synchronizing them. In cases of significant disparity, the model's underlying assumptions may have undergone a modification, leading to overfitting.

• Establishing a well-thought-out, standardized, documented process is another mitigation technique. While some processes like CRISP-DM exist, there are also some guidelines from Google, Facebook, etc. Many organizations also use their processes. Quite the methodology, it is vital to possess a relevant and documented process to be employed by data scientists and engineers.

• By utilizing versioning and pipeline management, it is possible to simplify redundant computations and concurrently execute multiple versions of dataset derivations. It may be possible to decrease the time spent in the investigation on resolving the data issues and continuously monitor the results across many pipelines since pipeline stages may have to be updated. When consumers change, the channel must be adapted accordingly.

• By preserving the data used to build the models, the financial industry could also inspire ongoing model risk management. Data-keeping policies and controls help verify the integrity of the data used to construct the model, thereby ensuring the audibility of machine learning models.

• Including other departments in data science training can be very valuable. Data scientists will be able to communicate and understand the models used in production and software quality assurance so they can issue an alert if models are no longer valid. Even in the assembly environment, human involvement could be a good idea. After inputting the data into the subsequent application, the systems that are uncertain about the predictions will reassess them.

D. The Path Forward

A balance between technical debt and unpaid debt is required. ML and AI systems are frequently becoming more precise as human intervention decreases daily. ML model construction involves applying a limited number of data science techniques. Maintaining their relevance and functionality in actual production environments is an additional significant obstacle. To achieve the transformational potential of AI and machine learning solutions, it is crucial to managing the process of data science, machine learning models, and their deployment and continuously evaluate results. Technical debt plays a vital role in such projects where organizations must understand the requirements of modeling, hardware, and software and learn about managing and optimizing their technical debt.

2.5 MLOps Philosophy:

1. **"The Seven Key Principles of MLOps" (Laura Norén):** Laura Norén has talked about seven key principles fundamental to any MLOps framework, with a significant focus on collaboration, iteration, and reproducibility. The article introduces the reader to the core value systems that guide the implementation of MLOps within organizations functioning in their respective industries.

    Collaboration and Communication: It stresses practical cooperation and communication between data scientists, engineers, and other stakeholders across different stages of the ML lifecycle.

    Iterative Development: It epitomizes how model development is supposed to be iterative by encouraging continuous improvement based on feedback received as well as changing requirements as they may occur.

    Reproducibility and Audibility: This emphasizes the significance of reproducible machine learning procedures in enabling the ongoing auditing of experiments and models.

    Automation for Efficiency: To enhance efficiency and reduce manual errors in the machine learning process, it recommends automating repetitive tasks associated with it.

    Explainability and Transparency: The emphasis here is on the need for machine learning models to be understandable, open, and transparent so that stakeholders can follow the decision-making process underlying predictions.

Feedback Integration: This approach argues that feedback loops should be incorporated into MLOps so that iterations can occur based on real-world performance.

Cultural Transformation: It pinpoints the need for organizational cultural change where MLOps becomes a part of the company's overall culture; it breaks silos and promotes cooperation.

Agile Development: This entails advocating for agile development as an approach within MLOps. It supports teams' ability to respond to changes and requirements in an iterative way quickly.

Documentation Standards: In this context, it stresses the importance of having comprehensive documentation standards in MLOps to ensure clarity, repeatability, and knowledge transferability.

Continuous Improvement: Here, we get advice on creating a continuous improvement culture where teams consistently assess and improve their activities for evolving conditions.

Inclusive Collaboration: The emphasis here is building inclusive collaboration, which calls for diverse views and skills to help contribute to MLOps processes.

Experiential Learning: This strongly advocates experiential learning in MLOps teams where people learn through first-hand experiences and share insights.

2. **"MLOps: A Complete Guide" (Piyush Kumar):** Piyush Kumar, the author of this guide on Analytics Vidhya, has provided an extensive overview of MLOps, including its philosophy, challenges, and best practices. By incorporating real-life illustrations and case studies, this ebook enhances comprehension of MLOps principles and provides further clarification.

    Industry Case Studies: Contains industry case studies to demonstrate how MLOps principles can address problems.

Challenges and Solutions: Challenges commonly faced with implementing MLOps are dealt with pragmatically, offering practical ways to overcome them.

Continuous Learning: Emphasizes the significance of continuously adopting new information and adapting it to suit current trends to stand out among other players within the sector. Practitioners must always be up-to-date with emerging technologies and methodologies.

Adoption Strategies: Gives insight into how organizations can adopt MLOps practices, including dealing with cultural and organizational hurdles.

Human-Centric Design: Draws attention to the need for designing human workflows, ensuring technology supports them rather than hinders aligning technology flows between humans, thereby concentrating on user experience when developing MLOps processes.

Cross-functional Collaboration: Encourages multi-stakeholder collaboration, which is not limited to just data science and operations teams but extends to business units that help achieve technical-business objective alignment.

Model Explainability: It deals with explanations of models, emphasizing that models must provide understandable insights, particularly when determining the fate of an individual or a society at large.

User-Centric Design: The MLOps lifecycle should be user-centric by considering the end-users and end-users and user’s needs and standpoints during the design process.

Scalability for Models and Processes: Considering expansion and heightened intricacy, the paper addresses model scalability and escalating MLOps process scalability.

Maturity Model Frameworks: Describes maturity model frameworks as a tool that will gradually raise their MLOps maturity through self-evaluation and improvement of organizations.

Scalability for Models and Processes: Explores scalability in terms of models and the scalability of MLOps processes, providing for growth and increased complication.

Maturity Model Frameworks: It provides maturity model frameworks that enable organizations to assess their MLOps maturity levels to improve them progressively.

2.6 MLOps Practices:

1. **"Implementing MLOps on Azure with Azure DevOps and Azure Machine Learning" (Nagesh Pabbisetty):** Nagesh Pabbisetty created a Microsoft Docs tutorial that outlines practical steps for implementing MLOps practices using Azure DevOps and Azure Machine Learning. It shows how MLOps fits into the larger Azure ecosystem.

Azure-Specific Implementation: Walks through implementing MLOps practices on Azure DevOps and Azure Machine Learning and smoothly integrates all relevant Azure services.

Pipeline Automation: Illustrates automatic setup of machine learning pipelines to generate a replicable and efficient workflow.

Scalable Deployment: The article discusses methods used in deploying machine learning models at scale, considering such factors as model versioning, testing, and monitoring in a production environment.

Integrated Monitoring: Demonstrates how to incorporate monitoring solutions into the MLOps pipe for real-time model performance tracking.

Containerization for Portability: Explains how containerization such as Docker can enhance reproducibility, ensuring that machine learning models are portable across different environments.

Model Monitoring Best Practices: Discusses critical practices when conducting model monitoring; this includes defining metrics that matter, setting alerting mechanisms, and determining thresholds for model performance.

Data Ethics: Involves talking about ethical issues related to data usage and model deployment, emphasizing the importance of incorporating ethical considerations into MLOps workflows.

Model Lifecycle Awareness: Advises communicators on knowing the entire life cycle of a model from data acquisition until it is ready for use, therefore ensuring well-rounded and responsible decision-making.

Continuous Integration Best Practices: This section highlights some of the best practices regarding continuous integration in MLOps, such as automated testing and version control integration, for a perfect collaboration.

Pipeline Orchestration: This section elaborates on the orchestration of machine learning pipelines to manage the various duties involved in the deployment and development of a model.

Cross-Environment Portability: This paper emphasizes methods for achieving cross-environment portability, enabling the seamless deployment of models across development, testing, and production environments.

Resource Scaling Strategies: Explores methods of efficiently scaling computational resources based on the workload, ensuring that MLOps environments have optimum resource placement.

Versioned Configuration Management: Best practices are discussed for versioning configuration settings to maintain consistent environments and enable reproducibility throughout different stages of the MLOps pipeline.

Model Drift Monitoring: The discourse revolves around implementing model drift monitoring, which aids teams in discerning fluctuations in model performance throughout a specified duration and executing essential remedial measures.

2. **"Practical MLOps for Enterprise" (S. Stanley Young, et al.):** The Harvard Data Science Review featured this article, which examines the practicalities of implementing MLOps in enterprise contexts, including scalability, security, and collaboration concerns.

    Governance Framework: Discusses why it is essential for organizations to implement a governance framework when dealing with machine learning models at scale.

    Scalability Considerations: This one explores scalability problems in MLOps and methods of handling increased requirements better in a corporate environment.

    Model Lifecycle Management: It expounds on machine learning models' lifecycle management, from their development and training to their deployment and eventual retirement.

    Security Protocols: It addresses security considerations in MLOps, providing insights into implementing robust security protocols for protecting sensitive data and models.

    Dynamic Scaling: The discussion is about the dynamic nature of machine learning workloads, the need for scalable infrastructure, and strategies used in resource scaling based on demand.

    Continuous Compliance: This paper investigates methods for maintaining regulatory compliance throughout the MLOps regime to ensure precise adherence to data protection laws and privacy regulations.

Failure Recovery Strategies: It examines strategies that could handle failures in MLOps, such as good error handling, logging, and rollback mechanisms, to ensure the stability of a system.

Knowledge Transfer: It emphasizes the significance of knowledge transfer within MLOps teams, where all team members should understand models and pipelines.

Regulatory Compliance Automation describes regulatory compliance automation checks that ensure all machine learning workflows conform to industry regulations.

Multi-Cloud Deployments: This study investigates the various options available for multi-cloud deployment, which involves deploying models across multiple cloud providers. It is of the utmost importance to safeguard against service interruptions in which no losses will occur.

Model Retirement Strategies: This expounds on how to gracefully retire models by considering options for decommissioning models that are not effective or relevant anymore.

Cross-Functional Incident Response: It supports the idea of cross-functional incident response groups that facilitate quick and collective troubleshooting in MLOps workflows.

2.7 MLOps Tools:

1. **"MLOps Tools Landscape" (S. Böhm, et al.):** The MLOps.community’s resource provides an exhaustive landscape of MLOps tools, categorized and described according to data versioning tools, model training, deployment, and others. The document serves as a good reference point for studying the ecosystem of MLOps tooling.

    Tool Evaluation Criteria: An analysis of the criteria for classifying an instrument as an MLOps tool and the factors to consider when deciding on a purchase.

Emerging Tools: Recognizing that the world of MLOps tooling is changing fast, it concentrates on emerging ones with their potential impact.

Community Engagement: The importance of community engagement is stressed while choosing MLOps maps tools by considering active community support, documentation, and user feedback.

Interoperability: Interoperability focuses on the need for tools that can blend seamlessly into existing ecosystems without creating data silos.

Explainable AI Tools: Highlights are given to new arrivals with tools focused on explainable AI, helping organizations stay tuned to the demand for interpretable and understandable machine learning models.

AutoML Integration: Automation in certain aspects of the machine learning process can be facilitated by exploring the integration of autoML (Automated Machine Learning) within this context.

Model Versioning: These delve into tools specifically designed to deal with modeling versions; for this reason, the team can easily trace and manage different machine learning versions.

Collaboration Platforms: Talks about the growing number of collaborative platforms specific to MLOps that promote teamwork and communication among data scientists, software engineers, and other stakeholders.

Model Serving Platforms: This section explains specific tools geared towards serving models in production, making it easier to deploy and handle machine learning models in a production environment.

Visualization Tools: Examines visualization tools in MLOps that aid in understanding how a model performs, its impact on comprehensive datasets, and other essential measures.

Integration with CI/CD Pipelines: Presents a series of tools that fit into the Continuous Integration/Continuous Deployment (CI/CD) pipeline to ensure a smooth transition from model development to deployment.

Experiment Tracking for Collaboration: Explores advanced experiment tracking capabilities within these tools, enabling collaboration by enabling colleagues to understand each other's experiments.

2. **"DVC: Open-source Version Control System for Machine Learning Projects" (Dmitry Petrov):** Dmitry Petrov has written a paper on DVC, or Data Version Control, which is helpful in machine learning projects. The guide helps users understand how to make the most out of DVC as an essential part of MLOps toward reproducibility and collaboration.

Decentralized Versioning: This section focuses on decentralized versioning inherent in DVC, allowing team members not to depend on any central server.

Integration with Git: It talks about how DVC can quickly go with Git and serve as a sound version control system for machine learning projects in particular.

Data Versioning: A look at how DVC differs from other tools by providing a way to track changes made to the code and the datasets throughout the process of building models.

Collaborative Workflows: How does DVC support collaborative workflows where different people are working on other parts of a machine learning project at the same time?

Data Versioning for Reproducibility: For all stages of the model development lifecycle, changes occurring in datasets should be traceable and reproduced; hence, data versioning is significant.

Experiment Tracking: This section talks about DVC's capability to track experiments. For example, it allows teams to record and compare experiments and model iterations.

Integrations with ML Frameworks: DVC can be used with several machine learning frameworks, making it adaptable to several technologies and environments.

Support for Large Datasets: This tool can manage vast amounts of data in experiments by creating and maintaining copies for each one, which aids in organizing data in extensive machine-learning projects.

Integration with MLflow: DVC integrates seamlessly with MLflow, allowing for comprehensive administration of all components in a machine learning workflow.

Collaboration Features: This part will cover some collaboration features in DVC, such as sharing, version control, etc. These are useful when many people come together for teamwork purposes in machine learning.

Support for Large Data Files: It can efficiently handle vast datasets by processing large files without requiring significant memory space, thus avoiding storage duplication.

Compatibility with Jupyter Notebooks: At the same time, they have an integration called DVC, which incorporates Jupyter Notebook, thereby offering a solution for storing code along with interactive notes on the go alongside it.

2.8 General MLOps Resources:

1. **"MLOps: From Model Development to Deployment" (Tirthajyoti Sarkar, et al.):** This Towards Data Science article covers a wide range of MLOps topics, encompassing the principles, practices, and tools involved. It is a one-stop shop that provides a panoramic view of MLOps.

    Cross-disciplinary Collaboration: Emphasizes the need for cross-disciplinary collaboration by breaking down data science silos between operational and business teams.

Model Interpretability: Covers the importance of model interpretability in MLOps, enabling stakeholders to understand and have faith in machine learning models.

Ethical Considerations: Discusses ethical considerations in MLOps, touching on responsible AI and the implications of machine learning on diverse communities.

Model Governance: Considers model governance concepts, including establishing policies and controls for managing end-to-end life cycle management of machine learning models.

Bias Mitigation: Proposes techniques to handle and reduce biases in machine learning systems to realize unbiased outcomes.

Hyperparameter Tuning Best Practices: This chapter explores best practices for hyperparameter tuning in MLOps, which mainly involves optimizing model performance by fine-tuning the parameters during the training phase.

Cross-functional Training: According to this section, cross-functional training within teams is a catalyst that helps individuals learn skills outside their primary areas of expertise, thereby promoting multidisciplinary approaches towards MLOps.

Model Robustness Strategies: Conversely, another chapter deals with strategies to improve models' robustness. These involve addressing issues about adversarial attacks, noisy data, and changes in distribution.

Knowledge Transfer Platforms: The authors suggest knowledge transfer platforms as they would help refine ML Ops among professionals by sharing best practices, experiences, and insights.

Interdisciplinary Training Programs: Similarly, it supports multidisciplinary training programs that enable one to be skilled in data science, engineering, and operations.

AI Ethics Education: This section examines how AI ethics education can be integrated into MLOps training, promoting responsible AI practices within the MLOps community.

2. **"MLOps, or DevOps for Machine Learning" (V. Feinberg, et al.):** The book by O'Reilly explains MLOps as an expansion of DevOps ideas to machine learning. It advises on implementing MLOps workflows in practice and is a good resource for professionals.

Cultural Transformation: This calls for a cultural shift in organizations to facilitate collaboration between data science and IT operations groups.

Continuous Monitoring: Concentrating on constant monitoring in MLOps to detect and address gradual model performance degradation with time.

Feedback Loops: This argues for having feedback loops between data scientists and operations teams that would help continuously improve model performance and deployment.

Resource Optimization: Discuss computational resource optimization strategies to ensure cost-effectiveness in MLOps environments.

Model Deployment Patterns: Examining varying deployment patterns within MLOps, such as canary releases and A/B testing, enables organizations to deploy models safely and incrementally.

End-to-end Automation: Calls for end-to-end automation throughout the life-cycle of MLOps, from preparing data to deploying models, thereby reducing manual intervention and speeding up the deployment cycle.

Cost Management Strategies: These are the strategies that can be employed for cost management in MLOps to optimize RUs and make machine learning workflows efficient from a cost perspective.

Continuous Learning Platforms: An Explanation of the Integration of Constant Learning Platforms in MLOps and Their Enhancement of Practitioners' Expertise and Understanding.

Community-Driven Practices: Describes the community-driven practices in MLOps whereby collaborative efforts and shared resources enhance field progression.

Model Retraining Strategies: Continuous model retraining techniques guarantee that ML models remain pertinent and accurate in changing real-life scenarios.

Knowledge-sharing Platforms: Knowledge-sharing platforms play a crucial role in disseminating insights, code snippets, and best practices among members of MLOps.

Distributed Team Collaboration: How is it possible to establish effective collaboration within distributed teams involved in MLOps work while dealing with geographical problems?

These summaries provide a comprehensive range of perspectives on MLOps, exploring in detail the philosophical underpinnings, operational practices, and nuanced toolset behind this specialized field. These summaries, therefore, go even more profound than surface-level discussions to underline the centrality of inclusivity in MLOps methodologies, with mentions of stakeholders and diversity throughout the ML life cycle. Discussions about scalability transcend basic questions because they involve adaptiveness to dynamic machine learning workflow trajectories, making MLOps resilient and robust. The maturity models discussed within these summaries offer organizations a guide on navigating the complexities of MLOps adoption. Insights into incident response strategies emphasize proactive measures that will be important for maintaining operational integrity in case of contingencies.

Moreover, delving into specific functionalities offered by individual tools about MLOps enables users to weigh options consistently based on their peculiar projects' needs. These summaries focus on the human aspect and, in so doing, underscore the value of good documentation and user-centered design, which facilitate teamwork and knowledge exchange among diverse groups. A study into cultural and ethical dimensions in MLOps reveals that responsible practices in ML have a great deal to do with integrating ethical considerations into decisions made during each step of the machine learning process. The range of inquiry also includes critical subjects such as model explainability, cooperative dynamics, and compliance, thereby providing an all-around view of their interrelated functions within the broader MLOps landscape. Going through these elaborate summaries immerses scholars in MLOps in such a profound manner that they can tackle future challenges and opportunities associated with ever, changing ever-changing machine learning operations, making them useful for research that demands comprehensive insight into the topic.

2.9 Summary: -

While practicing and researching machine learning operations, I have already seen many challenges, as stated in our literature survey. Operationalizing machine learning in day-to-day life and processes and the AI team requires a sound understanding of best practices around MLOps. A few constraints apply to this literature review: it is limited to scholarly articles on machine learning operations, which cover various topics associated with DevOps and MLOps. This thesis provides a comprehensive understanding of architecture and emphasizes the significance of using machine learning in industrial environments, among other subjects. These disciplines have undergone a substantial increase in popularity in recent years. The importance of operationalization, architecture, and machine learning in industrial contexts, among other issues, is gaining traction. The issue needs to be more widely researched and described in academic papers, even though many technical blogs, such as Azure MLOps, have introduced MLOps and their approaches. There is still relevant literature on this subject, even though additional research is required to investigate and develop the viability of implementing machine learning in production. As we have seen, DevOps has taken some time to get acceptance in software development. In the starting phase of DevOps, we have seen many challenges while implementing it. The same things are happening in data engineering and machine learning, known as MLOps, as we have seen in the literature survey. I have divided them into seven different stages.

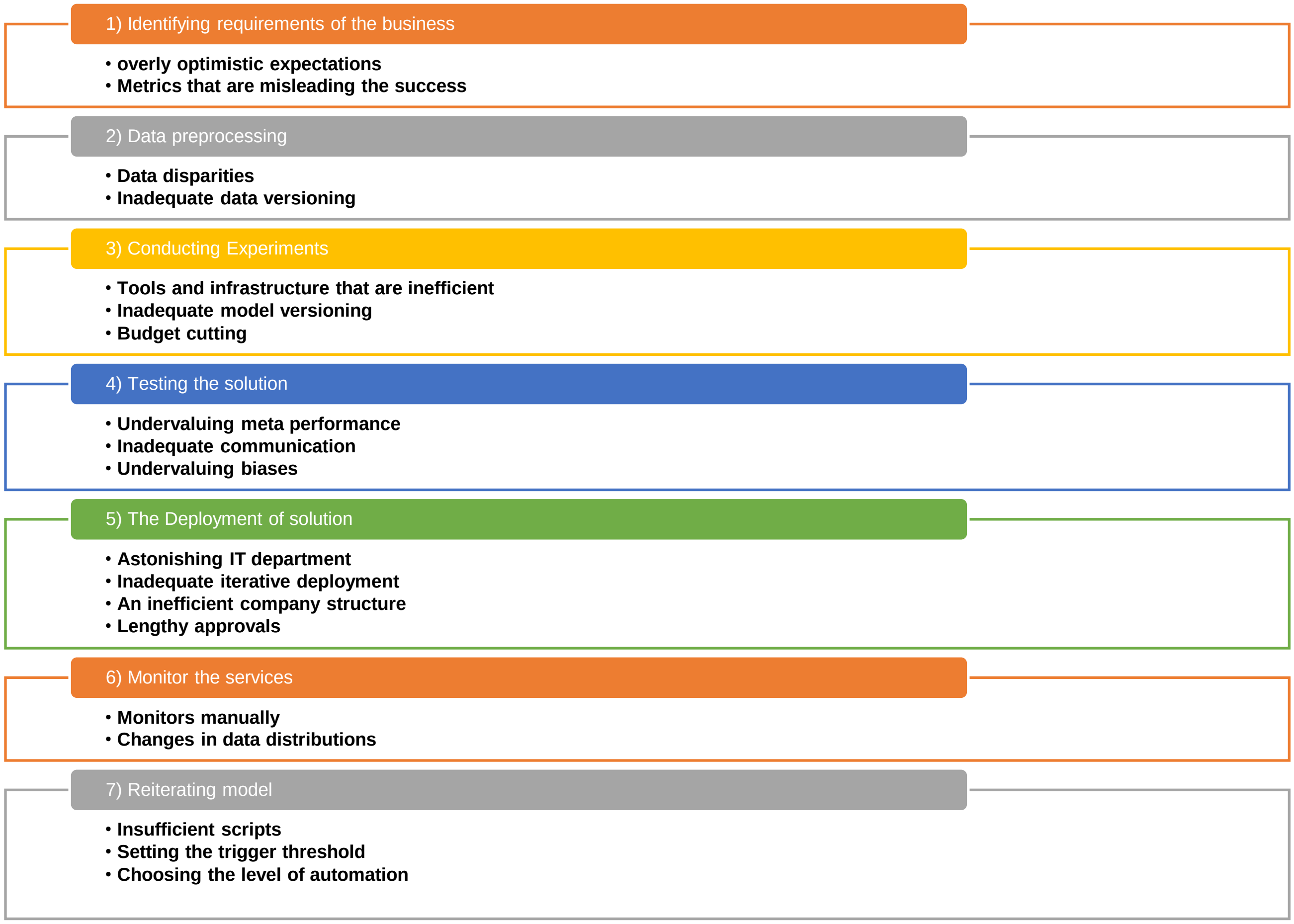


Figure 2.3 Seven different stages of MLOps life cycle (https://www.bitstrapped.com/blog/mlops-lifecycle-explained-by-stages)

CHAPTER III:

METHODOLOGY

3.1 Research Methodology

This chapter provides a comprehensive overview of the methodological framework adopted in this thesis. It encompasses a rigorous examination of the research approach, the nature and extent of collaboration with the partnering company, contextual factors shaping the investigation, the intricacies of employed data-gathering methods, and the subsequent analytical approach applied to the collected data.

The research methodology for this thesis can be as follows:

3.2 Literature Review

Before beginning the research process, it is critical to do a thorough literature review of ML-Ops and DL-Ops. To do this, individuals must actively absorb literary materials such as books, papers, and research articles written by renowned academics and specialists in the relevant field. Individuals who participate in this activity may get a solid foundation of learning and competence in the current state of ML/DL-Ops. It was necessary to fully understand the current information and study on study Question 1 (RQ1) before starting the investigation. Following the steps shown in Figure 3.1, the study tried to find gaps in knowledge, combine previous research, create a project framework, and assess the first answer suggestion based on current findings and assumptions.

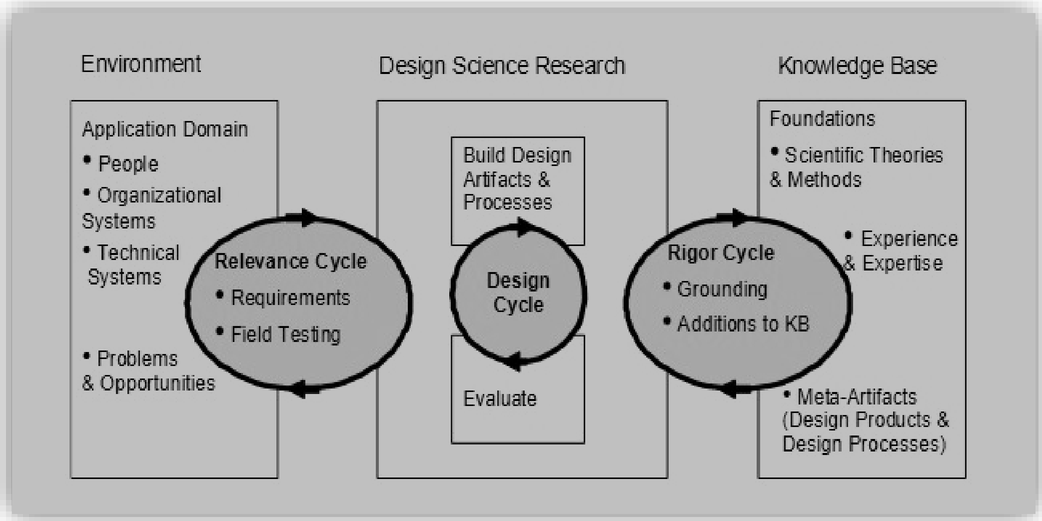


Figure 3.1:The illustration below has been adapted from Hevner (2007) to present the iterative workflow used in design science research.

For instance, Table A.1 in Appendix A shows how outcomes of literature analysis played a significant role in shaping the trajectory of development for the artifact. The critical examination of current problems and best practices from the literature led to an exhaustive list of Requirement Questions. Eventually, it served as a reference for all Design Cycles that followed this set. Nevertheless, there was no way to avoid repetitive cycles since very few people within MLOps were knowledgeable about it. This methodological pattern affected the artifact's initial structuring and subsequent iterative cycles. Thus, by embedding insights gained from the literature into this study, we put proactive strategies in place to address subtle complications characteristic of MLOps, ensuring a sturdy foundation with enough knowledge for subsequent steps of our project. It is worth noting that thorough reviews involve search engines such as Google Scholar and Web of Science, known worldwide for their extensive literature collections.

These platforms systematically combed through established and pertinent journals for this study, including IEEE and ScienceDirect. The initiation of the search involved the utilization of key terms such as "RE for MLOps," "RE for ML," "MLOps challenges," "DevOps for ML," and "MLOps best practices." We used the "snowball" method to find even more related studies after seeing a few useful ones. It was clever to add "gray literature," which only came in handy when peer-reviewed materials were considered and comprised a small percentage of the sources we discussed. This methodological approach ensured a rigorous and comprehensive exploration of the literature landscape, aligning with the standards of scholarly research.

3.3 Interviews with Experts

This section will articulate the various aspects of the research, including the guiding principles and the nature and philosophy of the research.

There are two main research designs used in academia, quantitative research design and qualitative research design. The quantitative design is primarily about examining the relationship between variables. It involves generating data from samples and analyzing them using statistical techniques and works well with the deductive approach. Qualitative research design, on the other hand, is used with the inductive and abductive approaches. This research design often involves interviewing people, asking probing questions and deriving insights. While quantitative research design involves examining the relationship between variables, qualitative research design involves examining the relationship between entities. The research being presented is a descriptive study and looks to properly explain the various phenomena around the success or failure of businesses. Therefore, qualitative research design will be employed.

The authors define semi-structured interviews as a widely used qualitative research method. Specific questions can allow the researcher to seek clarification or further exploration of any issues that may occur during the study (Doody and Noonan, 2013). As a result, interview methods systematically assisted similar data gathering from participants, providing a sense of orderliness during the investigation. We made sure the questions were open-ended and adaptable so that you could adjust the order of the questions, change the language, or add more questions if needed. We have done this on purpose so that when conducting these interviews, we could alter our initial plan if circumstances required it.

The data collection process is summarized below: -

1. Create a google survey for potential interviewees.
2. Create a set of probing questions to ask each interviewee. Each interview should last an estimated 45 – 60 minutes.
3. Fix a period for the interviews to be held.
4. Reach out to potential interviewees through various channels such as LinkedIn, email, messaging services etc.
5. Ask each responding individual to complete the google survey.
6. Potential interviewees who meet the criteria of the survey will be shortlisted and an email will be sent to each of them introducing the research.
7. Each interviewee will be sent the interview consent form and details of the study, along with an explanation of their rights.
8. Each interviewee will be asked to sign and return the consent forms.
9. A date and time will be fixed with each interviewee for the interview.
10. The interview will be conducted over a video conferencing tool such as Microsoft teams and will be recorded. Interviews will be conducted in English.
11. A copy of the recording will be provided to the interviewee for fact checking and confirmation. The recording will be transcribed soon after the conclusion of the interview.
12. Each interviewee will be asked a series of probing, open ended questions to best capture their life experience.

13. The researcher will use observation techniques to ask deeper, more pointed questions based on the interviewee's answers.
14. The core question and sub-questions of the research will be answered through an abductive approach.

The interview is constructed as a set of semi-structured questions presented to the interviewee in a set order. The researcher may choose to ask more pointed questions to obtain more details or gain more insights.

The participants for this research will be chosen carefully across freshers to senior roles, including and up to co-founders and CTO. The potential candidates will be chosen from business of a variety of sizes, revenue, and age. Candidates chosen this way will provide a rounded and more accurate depiction of the various problems that their businesses had faced.

The interview questions are prepared well in advance, and the interviewee is made aware of the time and date of the interview beforehand so that they can best prepare for the process. Since this research involves discussion of potential MLOps secrets, internal knowledge, and business health, it is important that the researcher gain the utmost confidence of the interviewee. This is done by explaining to the interviewee their rights and by explaining the interview consent form in detail. Furthermore, the researcher will remind the interviewee about their rights at the beginning and the end of the interview.

In addition, we have also conducted the pilot interview at each iteration stage to test and refine the interview methodology before proceeding with additional interviews. These interviews were video-conferenced and mechanically transcribed via the Microsoft Teams service. To minimize the possibility of any errors in the transcriptions that came through automation, both researcher carefully analyzed these recordings and corrected any mistakes they found. We have a meticulous approach to ensure that data collected during semi-structured interviews is reliable and valid.

Then, you may need to interview specialists dealing with ML-Ops / DL-Ops. These might include data scientists, engineers, and business leaders who have handled machine learning and deep learning models' deployment and management in the cloud. Such interviews will highlight the most significant challenges and best practices in ML-Ops / DL-Ops.

During the interview, the researcher will not use detailed written notes, and rather record the entire exchange. The interview will be done over a videoconferencing app like Microsoft teams, with video turned on for both the researcher and the interviewee. This is done so that the researcher can observe the interviewee's reactions to questions, and their body language while answering. The researcher may take short written notes to capture important pieces of information and to phrase proper follow up questions.

After the data collection process is completed, the researcher will codify all the important sections of each interview and use an abductive approach to identify one or many key metrics and business practices that occur commonly across these businesses that contributed to their success. As a result, the themes that emerge will be used to create a framework for aspiring businesses to adopt and increase their chances of success.

3.4 Data Analysis:

This research made use of qualitative data. Hence, coding and theme identification, according to Saldaña were, are the means of analysis (Saldaña,2021). Coding is a process where qualitative data is analyzed to develop patterns and themes that corresponding codes can depict. Appendix D.1, Codebook, contains the compilation of the final codes. While going through interview transcripts, for example, open or initial coding was done to enable the researchers to allocate codes for each data item related to these questions. Afterward, axial coding enabled us to examine interrelationships among codes, leading to higher-order themes. By practicing this methodology, we identified trends and patterns in our data, which allowed us to develop a deep understanding of our subject matter. In addition to initial coding and axial coding, "Themeing the Data," as presented by Saldaña (2021), was also employed here to bring out overarching themes from the corpus of collected data (Saldaña,2021).

So, common patterns appeared in many different sources of information, and papers ultimately fell into groups based on those patterns. The purpose here was, therefore, aimed at how different sources would fit together while also identifying any mismatches or inconsistencies leading towards an integrated storyline. Several methods serve to ensure that the data analysis is reliable and trustworthy.

For a start, the author carried out their separate data analyses. Subsequently, these studies' results were prominent in guaranteeing a consistent interpretation of the data. Getting as much information as possible from various sources, such as books, interviews, or others, was also possible. This method avoided having outcomes influenced by only one data source.

3.5 Research Design Limitations

This research does have some limitations that may reduce the generalization of the findings. Since this was a study conducted using interviews, where the sample size was limited, and the interviewees were selected using random sampling and a google survey, it may be the case that the experiences of these interviewees do not fully capture the experience of all data driven businesses in India. That said, it is imperative to state that the size of the sample is less critical than the quality of the data being generated and analyzed through these interviews.

Furthermore, while the interviewees were given ample time to prepare for the interview, it is possible that they did not recall incidents as they actually happened and may have missed details that would affect the outcome of the research. Additionally, a fundamental assumption of this study is that the interviewees had the relevant experience and were considered experts in their domain at the time of the interview.

Lastly, while the interviewees were informed and assured that their answers would be kept confidential, there is the possibility that their answers did not accurately depict their lived experience.

3.6 Conclusion

The researcher has explored the qualitative methods of research design. Qualitative research design is applicable when the phenomenon in question is related to the lived experiences of the people involved in the research. The research instruments used for this research were a google survey to screen potential interviewees, and a semi-structured interview comprising of open-ended, probing questions.

The answers provided by the interviewees were the main source of data for this study, and the responses to the google survey brought in context to some of the answers. The various procedures for data collection, coding and analysis used in this study were presented and discussed in the above sections. The coding techniques by Saldaña (2021) were explained and used to transcribe the interviews, create relations with the data extracted and the findings of the research. Finally, the limitations of the research were stated and discussed.

CHAPTER IV:

ARTIFACTS

This chapter examines the artifact constructed through three Design Cycles to address the challenges identified as the results of Research Question 1 (RQ1). What are the current challenges in designing an MLOps process, and how do they relate to requirements knowledge? The artifact offers a resolution to Research Question 2: What potential solutions exist to mitigate the challenges of developing an MLOps process based on requirements engineering? This chapter thoroughly examines the artifact's concept and design, precisely the Requirements Form for machine learning operations. Additionally, it includes valuable guidance on the optimal utilization of the form. Chapter 5 elucidates the rationale behind the artifact design and presents the discoveries.

The MLOps Requirements Form is a tool designed to aid individuals, teams, or both in collecting MLOps requirements throughout the implementation of MLOps. The artifact aims to align with the sequential stages of an MLOps process, namely scoping, data, modeling, and deployment. Depicting the graphical representation of the information transmission process from one phase to the following, Figure 4.1 offers a comprehensive analysis of the phases.

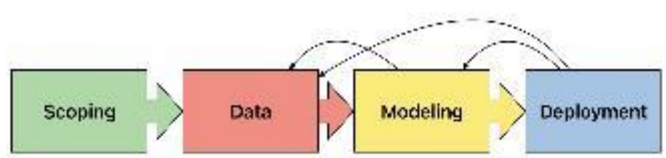


Figure 4.1 An examination of the steps of MLOps and its iterative characteristic.

The MLOps Requirements Form comprises two components: The initial section is an introductory tutorial for users, providing instructions on utilizing the form successfully. For further reference, please see Table A.1 in

A. The second component is the actual form, derived from the research of the 4 Artifact problem in cycle three and the assessment conducted in cycle two. Please refer to Table 4.1, 4.2, 4.3, and 4.4 for visual representation.

4.1 Design Science Research (DSR): -

This study adheres to a Design Science Research (DSR) methodology. *Design science* is a research approach that focuses on creating and assessing practical solutions, such as models, theories, and prototypes, to solve real-world issues and enhance scientific knowledge (Mayr,2020). Design science frequently addresses complex real-world problems that conventional research methods cannot fully comprehend or address.

According to Knauss (2021), DSR generally follows a methodical and repetitive procedure that includes recognizing a problem, creating a solution (the design artifact), and assessing the efficacy of that solution. Figure 4.1 provides a comprehensive summary of the strategies employed in each iteration. In this chapter, we will elaborate on each approach in detail. Designers design artifacts to resolve recognized problems in practical situations, tailoring them to specific audiences or user groups.

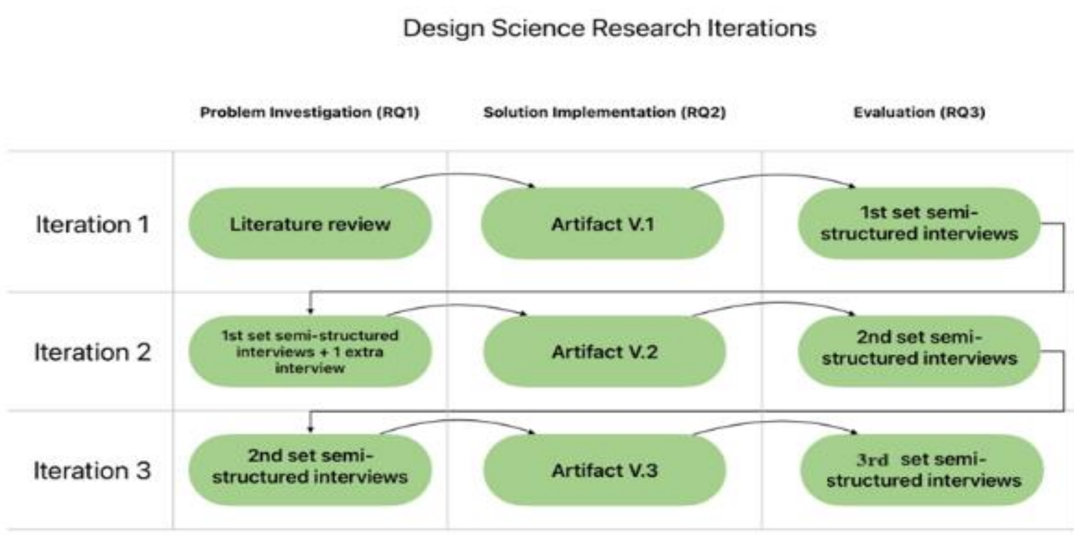


Figure 4. 2 An overview of the methodology utilized throughout the three DSR iterations in this study.

Furthermore, the figure illustrates the flow of data and knowledge from one DSR stage to the next. Additionally, the columns show how each event relates to the DSR stages and the RQs.The focus is on explaining the artifact designed to address the challenges identified as results of the RQ1: What current challenges are involved in developing an MLOps process, and how do these challenges correspond to requirements knowledge? The artifact acts as the solution and results of RQ2: What potential solutions are available to address the challenges of developing an MLOps process rooted in requirements engineering? This chapter presents the design and the idea behind the artifact, which is an MLOps Requirements Form, and gives suggestions on ways of using it. An explanation of the artifact's structure is provided, along with the results.

Part of the ML Lifecycle	**Roles to ask**	**Requirement Question**	**Requirement Question Answer**	**Examples**
Scoping:				
	Business stakeholder	What specific challenges is the business facing?		Battery optimization, Fraud detection, Demand forecasting
	Data Scientist	Is machine learning a viable solution for addressing these business problems, and if so, how?		Has it been done before, research proves it possible, still unclear
	Product owner	What metrics will be used to measure the		ROI, customer wishes

		success of the solution?		
	Product owner	What resources are required to implement the proposed solution?		Data, time, people
	Product owner, Business stakeholder, Data scientist	What is the budgetary constraint for the computation required to train the model?		If on-premise: 100h allowed, 50h, Unlimited If on cloud: Budget is $1,000, $5,000, $500
	Business stakeholder	Who constitutes the end user in this context?		Demographical information, Internal company users, Customers
	Business stakeholder	How will users interact with the model, and what interface is necessary for their interaction?		App, Voice-activated feature, Web page, API

	Business stakeholder, Product Owner, Data scientist, Data engineer	Who serves as the domain expert, and is there accessibility to them for consultation?		Doctors, Lawyers, Domain-specific researcher

Table 4. 1: Part one of the final artifact, includes requirement questions regarding the scoping stage of an ML system.

Part of the ML Lifecycle	**Roles to ask**	**Requirement Question**	**Requirement Question Answer**	**Examples**
Data:				
	Business stakeholder, Product Owner, Data scientist, Data engineer	Where is the data sourced from?		Owned data, crowdsourced, purchase data, purchase labels
	Data scientist, Data engineer	What format is designated for the data?		Structured, unstructured
	Data scientist, Data engineer	What preprocessing		Remove data, remove duplicates

		steps are required for the data?		
	Data scientist, Data engineer, Domain expert	What guidelines exist for labeling the data?		On images: Label each scratch independently on the screen, label each animal separately in the field
	Product owner, Business stakeholder	By whom will the data be labeled?		In-house resources, Crowdsourced, Outsourced, Mixture of resources
	Data scientist, Data engineer, Product owner	What meta-data needs to be gathered alongside the data?		Time, system model, factory, device type
	Data Engineer, Legal team, Business stakeholder, Product owner	Are there privacy considerations related to the data?		Names, Emails, Addresses, Phone numbers, general GDPR concerns
	Data Engineer, Legal team, Business	Are there specific ownership		Data is owned by us, it's open source, and

	stakeholder, Product owner	considerations for the data?		another party owns all data
	Product owner, Data scientist, Data engineer	What is the anticipated volume of stored data?		~10TB
	Product owner, Data scientist, Data engineer, Domain expert	When does the data reach a point of irrelevance?		Never, new product versions are released, annually
	Data engineer, Domain expert	Are there recurring patterns or cycles in the data?		Seasonal sales cycle, full-day cycle
	Data scientist	What is the minimum data quantity required for model training?		10k images, 100 GB worth of 1080p mp3 video recordings
	Data scientist	What is the minimal data point frequency		Every 5ms, Every 1s, Every data point

		required for streaming data to align with the business goals?		
	Product owner, Data scientist, Data Engineer	How is the data acquisition process structured?		Automated tool, manually collected, purchased

Table 4. 2: Part two of the final artifact, includes requirement questions regarding the data stage of an ML system.

Part of the ML Lifecycle	**Roles to ask**	**Requirement Question**	**Requirement Question Answer**	**Examples**
Modeling:				
	Product owner, Data scientist	What constitutes the baseline for the model?		Human-level performance, A previous system's performance, Dummy model
	Product owner, Legal	Is model auditing required, and if so, who is responsible for conducting it, and		Yes/No. Business stakeholder, Third party, Data scientists. Transparency, Equality, Fairness, and Accountability...

		what is the focal point of the audit?		
	Data scientist, Data engineer	What potential biases should be acknowledged and addressed in the model?		Gender bias, Brand bias, Ethnicity bias
	Product owner, Data scientist	How is the input data presented to the model?		Batch data, Real-time data
	Data scientist, IT Architect	Where is the appropriate storage location for the outcomes of experimental data?		Database, Excel document, JSON-file
	Product owner	What key business metrics should the ML model prioritize?		Business required classifications performance, different from general ML model performance
	Data scientist	Which experimental data points should be monitored?		Dataset used, Hyperparameters, Results,

				Results with metric summary/analysis, Training resources, Training time),
	Data scientist, Software engineer, DevOps engineer, MLOps engineer	What constraints are in place for model deployment?		None, Edge device's hardware capabilities

Table 4.3: Part three of the final artifact, includes requirement questions regarding the modeling stage of an ML system.

Part of the ML Lifecycle	**Roles to ask**	**Requirement Question**	**Requirement Question Answer**	**Examples**
Deployment:				
	Product owner, MLOps engineer, DevOps engineer	What is the recommended approach for managing the deployment process?		Canary releases, A/B releases, Shadow releases
	Product owner,	Where is the designated		Cloud or edge device

	MLOps engineer, DevOps engineer	location for the prediction device?		
	DevOps engineer, MLOps engineer, Software engineer	Which software metrics should be closely monitored?		Memory, computing power, latency, throughput, server load
	Data scientist, Data engineer, MLOps engineer	What input metrics are considered crucial for monitoring?		feature types (INT or String), feature range, Data schema validation
	Data scientist, Software engineer, MLOps engineer	What output metrics are essential for ongoing monitoring?		# times users redo the search, avg. prediction accuracy
	Product owner, MLOps engineer, Data scientist	At what frequency should the model undergo retraining using the data collected during deployment?		Every Monday, once a month, based on deployed input/output metric triggers

	Product owner, DevOps engineer, Data scientist	Are there any explicit performance requirements that need to be met?		Latency requirements, Query per seconds requirements

Table 4.4: Part four of the final artifact, includes requirement questions regarding the deployment stage of an ML system.

The MLOps Requirements Form systematically arranges requirement questions based on their relevance to different stages. Furthermore, this prioritization ensures a logical progression of questions by aligning them with specific steps and incorporating thoughtful sequencing within each stage. The meticulous structuring replicates the natural flow of inquiries within the context of an operationalized machine learning project. However, these queries have naturally emerged during the various phases of operationalizing an ML project. The strategic organization improves the form's effectiveness for providing all-inclusive information and facilitating smooth integration into agile workflows for machine learning projects. Mainly designed to elicit feedback from pertinent team roles, specific requirement inquiries within this form may transcend the scope of implementation groups. Each question has a dedicated space for detailed response writing.

Similarly, the requirement responses provide practical examples to illustrate the interpretative aspect of the requirement question. These are, therefore, recorded responses to the prerequisite inquiry. These recorded responses capture key insights and perspectives from different MLOps team members on MLOps, resulting in informal requirements. Users can utilize these resulting requirements as is or as templates for more formalized requirements development in the context of MLOps without any modifications, or they can serve as the initial basis for such action.

The MLOps Requirements Form is designed explicitly with universality as a deliberate emphasis, making it universally applicable across industries and projects of diverse natures. A thorough literature review encompassing well-established best practices and common challenges encountered while implementing MLOps processes has ultimately influenced this set of inquiries. An iterative refinement process followed through interviews with experienced practitioners with practical knowledge of MLOps or related areas. The questions within the form have been improved over time using insights gained from these engagements.

For each question on the form, the design is such that it can change to risks and other best practices learned from analyzing literature reviews and putting together interviews with practitioners. It is essential to make this form work for different types of businesses by ensuring it follows best practices and asking people who have been in that line for a long time for their advice. Methodically, this approach shows the form's adaptive capabilities and situates it within a complex response regime, which addresses varying challenges and best practices associated with MLOps implementations.

The MLOps Requirements Form draws on existing literature and interviews with industry experts in MLOps. Through synthesizing this information, practitioners get an all-inclusive yet flexible means for gathering and documenting requirements during an MLOps process.". This approach empowers those using the product to think systematically about many important factors and conditions. This expansive perspective intends to transcend the limitations of specific projects or industries, reducing exposure to common issues that might hinder the efficiency of implementation in MLOps.

The active incorporation of knowledge from academic sources and experienced MLOps practitioners enhances its robustness, making it a sophisticated tool for addressing multi-faced considerations. Such strategic alignment with different types of expertise makes the form more valuable and suitable as a resource to help professionals navigate through complex MLOps requirements. Ultimately, the main objective is taking proactive measures toward reducing common risks, thus fostering resilience and success in implementing MLOps.

4.2 Practical Application: -

The use of the artifact depends on several factors, such as the individual's professional background, the size of the project, and company organization. For example, during a project scoping process, a practical approach might be to have an all-inclusive meeting that involves relevant stakeholders from various roles and initiate collective discussions to respond to all Requirement Questions and queries. On the other hand, a more personalized method may be an individual assigning one person to interact with each requirement question.

By considering these variations within the project scope, this adaptive model acknowledges that diversity in preferences and availability is vital. An inclusive group meeting promotes collaborative decision-making, thereby ensuring a collective understanding of requirements. In contrast, a one-on-one setting allows for a more personalized probing deep into each query's minute details, which may unearth some hidden angles related to people's thinking patterns. Adapting the artifact to the specifics of the setting, the project, and the organization's work makes it more useful as a flexible tool for gathering and documenting requirements under the MLOps regime.

The artifact will function only by beginning a collaborative engagement with the stakeholders through the MLOps Questions Requirements. It is possible through an organized gathering of project stakeholders during the scoping phase, where experts from different fields work together to answer each requirement question. Another option is a one-on-one method, in which a specific person leads discussions between various stakeholders.

It is necessary to document these engagements' responses formally or informally. These answers should be shared widely with relevant parties, including implementing teams and stakeholders, promoting transparency and alignment among team members. Such an interpretative nature leads to flexibility in their adoption. Depending on the details of a project, they can either act as formal needs or function as immediate requirements. This strategic flexibility helps the artifact fit into various MLOps scopes (project size, organizational structure, and personal preferences), making it applicable across diverse contexts.

CHAPTER V:

RESULTS

This chapter thoroughly evaluates and analyzes the results obtained from the thesis project. Each of the three parts represents a cycle or iteration. Problem Investigation (RQ1) has sub-sections in each process that explore the question of the study in greater detail. After that, the Solution Candidate (RQ2) gives a possible way to deal with this problem. The Solution Candidate (RQ3) Evaluations then critique this suggested solution. This classification strategy makes it possible to present research results systematically within every iterative phase.

5.1 Findings from the First Design Cycle

A synthesis of existing literature was necessary for identifying initial challenges, best practices, and prerequisites for MLOps. Subsection 4.1.1 builds on this idea by introducing primary challenges and best practices, among other things necessary for understanding their foundations; Table A.8 Annex provides eight requirement questions that emerged from the insights gathered during this stage, among others in Subsection A.9 Annex 'Requirement Questions.' It is a well-structured presentation that thoroughly examines the first stage of the Design Cycle's achievements.

5.1.1 Problem Investigation

This subdivision outlines the difficulties and patterns discovered via an extensive literature review chosen for their universality and usability in various MLOps scenarios. Consequently, we formulated specific questions (refer to item 4.1.2) to address these challenges and implement more effective methods. The feedback to these queries could act as the initial requirements for the MLOps process, which can be refined or improved over time. In such an analytical procedure, a comprehensive grasp of foundational aspects is drawn from literature analysis to create a solid framework for MLOps processes.

Challenges Found in the Literature: -

P1 - Data Drift: Here, the input data's distribution changes from that used for training the model, while the desired prediction output remains unchanged. This kind of data change can reduce their predictability for models trained on different datasets. A data drift describes this. Drift can move at a fast or slow pace. In tackling this problem, monitoring systems must be put in place to supervise changes happening throughout distributions and output and input modifications (Paleyes,2022). Retraining the model and incorporating pertinent updates will proactively resolve drift and maintain the system's precision and reliability (Paleeyes, 2022).

P2 - Concept Drift: A change in the definition of output is necessary whenever there is a change in the inputs to a model since this is known as idea drift. Consequently, patterns acquired before by such models become obsolete, deteriorating their prediction accuracy. Addressing this issue necessitates strategies like those used for P1, including frequent retuning sessions integrated with continuous monitoring of input/output distributions (Paleyes,2022).

P3 - Inter-team Communication: Within MLOps, communication problems can start surfacing due to various roles and different expertise levels among practitioners, as Kreuzberger et al. (2022) noted. According to Ng (2023), there are times when an ML model might be performing well on test sets but needs to meet business objectives, leading to discord between ML and business teams. Evaluating models based on average error rates alone may miss critical cases, thus resulting in failed deployments. Furthermore, more is needed to rely on individual responsibilities (Kreuzberger et al., 2022). However, the successful completion of MLOps initiatives depends on promoting effective communication amongst interdisciplinary teams.

P4 - Performance During Serving: Two mechanisms related to post-deployment performance typically account for these difficulties. First, traffic management concerns involve network latency, ML system throughput, and access points. Secondly, issues within this class include whether accurate labels are available for data passed through models used in predictions. The current status of a deployed model is challenging to monitor due to the sporadic availability of accurate labels (R. Ashmore, 2019). Serving is a common term for this stage of MLOps, and the literature review on MLOps (Paleeyes, 2022) and Ng's expert course (2023) highlight this step's unique difficulties.

P5 - Disorganized Data: To train the model, we use data from a variety of sources. Consequently, using this raw data as input for the model is usually tricky, given that such formatted data may take time to handle the model (Letouzey,2022).

P6 - Sustainable MLOps:

Tamburri et al. (2022) extensively examine the development and implementation of sustainable MLOps, elucidating three key components: explainability, fairness, and accountability. They emphasize the importance of explainability, which involves explaining why automated decision-making occurs as it does. It talks about fairness, which brings up the fact that ML systems need to ensure everyone has an equal chance to make decisions while also trying to avoid bias or discrimination. Furthermore, they discuss accountability, focusing on fixing misaligned attributes and identifying those responsible through blame assignment. Tamburri et al. (2022) emphasize the need to operationalize these concepts for sustainability, considering their interconnectedness and the ethical use of ML in general. Explainability enables the observability and self-improvement of MLOps; fairness is necessary to sustain social contracts, and accountability reflects the legal systems. However, Villamizar et al. (2021) state that the biggest problem with ML is the need for more awareness of specific non-functional requirements (NFRs) like explainability, fairness, and accountability. They have noted that there needs to be more awareness about NFRs, and practitioners face challenges in defining and refining NFRs within an ML context.

Best Practices Found in the Literature: -

BP1 - Versioning: The authors emphasize implementing versioning on data, models, experimentation logs and code to increase system reproducibility and traceability (Kolltveit,2022). Ashmore (2021) says that each system part needs its storage area. For example, there should be a model registry, a feature store for features, a pipeline store for data and machine learning pipelines, a regular source code repository for machine learning, and Infrastructure as Code (IaC) scripts. It is in addition to storing metadata that includes hyper-parameters and model metrics! By using this approach, firms can keep a complete record of any changes they have made to their systems, making it easier to discover mistakes or bugs introduced during development. Revisioning also allows data scientists to confidently reproduce results, validate findings, and build upon prior research more effectively.

BP2 - Model Deployment and Serving: Kumara et al.'s (2021) "Model Deployment and Serving" is an essential concern in the MLOps spectrum. According to the authors, choosing the model prediction serving plan is one of the most critical steps. According to the authors, selecting the model prediction serving program is one of the most vital steps. There are three ways. The first is model-as-service, in which the model is made available as a web endpoint called precompute. Precompute, during which the model constructs estimates by awaiting the input of data groups to store the results for subsequent utilization., and Model-as-Dependency, in which the model starts up as the program runs. Furthermore, according to the authors, these serving methods may require specific architecture designs depending on whether online or offline serving is available and whether real-time or batch input data exists. These examples demonstrate key best practice considerations when establishing an MLOps architecture.

BP3 - Data Quality and Labeling: They talk repeatedly about the importance of maintaining proper data quality during the creation of ML systems. The writers assert that there are many times where when professionals use already available public datasets for training. Nevertheless, these datasets, though accessible, mostly have limited quality and unbiasedness. Vogelsand and Borg argue that shortcomings in publicly available datasets are due to their poor labeling. It stresses the need for adequately coordinated and transparently implemented labeling processes, especially when data quality is essential, as often happens when developing any ML system.

BP4 - Feasibility: According to Vogelsang and Borg (2013), data scientists tend to make technological choices in machine learning systems without considering stakeholders' business context and needs. Ng (2016) agrees with this idea in his expert course, where he emphasizes that finding projects suitable for machine learning is problematic in this field. According to Ng, instead of searching for machine learning problems, it is preferable to begin by analyzing business issues that machine learning can solve. It helps simplify things because once they understand their problem well, finding the solution becomes more straightforward. However, before embarking on an analysis of success metrics and resource budgeting, one should evaluate whether these solutions are feasible or beneficial.

5.1.2 Solution Candidates: -

After looking into the problems and best practices in the preliminary problem investigation mentioned in Section 5.6.1.1, we have developed a new version of the artifact. The artifact, as shown in Table A.1 in Appendix A, served as a foundation for subsequent iterations. Table 5.6.1 is a traceability matrix that relates the Requirement Questions to the artifacts' Best Practices and Challenges.

Additionally, this table is divided into rows grouping several MLOps stages as they are presented throughout chapter 4 of this thesis.

Table 5. 1: Traceability matrix showing how each Requirement Question originated from corresponding Best Practices or Problems identified within the related literature.

Problem (P):	**Best Practice (BP):**	**Source:**	**Requirement Question:**
Scoping:			
	BP4: Feasibility	(Miao,2017; Letouzey,2022)	What are the existing business challenges, and is AI a viable solution for addressing them?
			What metrics define success in this context?
			What resources are required to tackle these challenges effectively?
Data:			
P5: Disorganized data		(Kolltveit,2022 ;Letouzey,2022; Baier,2019)	How is data preprocessing to be conducted?
			What is the source of our data?
			Which data format is designated for use?
			What metadata should be gathered?

	BP3: Data quality and labeling	(Knauss,2021)	What standards guide the labeling of the data? Top of Form
Modeling:			
P3: Poor communication between ML and business teams		(Kolltveit,2022; Letouzey,2022)	What constitutes the model baselines? Which ML model should take into account crucial business goal metrics?
P6: Sustainable MLOps		(Ng2023; Villamizar,2021)	Is an audit of the selected model necessary, and who is responsible for conducting the audit? What potential biases should be acknowledged in the model?
	BP2: Model Deployment and Serving	(Ashmore,2021)	How is the model's input data presented? Are there specific constraints for model deployment?
	BP1: Versioning	(Kolltveit,2022; Ashmore,2021)	Where is the appropriate storage location for the outcomes of experimental data?

			Which experimental data points should be monitored?
Deployment:			
P1: Data drift, P2: Concept drift		(Paleyes,2022;Ashmore,2021; Ashmore,2019)	What software metrics should be closely monitored? Which metrics related to input are critical for monitoring? Which metrics related to output are crucial for monitoring? what frequency should the model undergo retraining with the deployed data?
P4: Performance during serving		(Knauss,2021 ; Letouzey,2022;)	Where is the optimal placement for the prediction device? Are there any specified performance criteria to be met? What is the recommended approach for managing the deployment process?

Making Requirement Questions was a carefully planned process designed to get essential information from different parts of a team to derive informal requirements that are important for an MLOps project. This way of doing things makes the requirement development process more efficient and helps it align with established best practices.

5.1.3 Evaluation: -

This subsection describes what happened during the evaluation phase in the first set of interviews, the goal of which was to gather feedback about this artifact's usability, usefulness, and content. The primary purpose behind this evaluation was to gain insights to improve the artifact based on the problems and trends identified in literature reviews. The following section, 5.6.2.2, highlights actions taken from the evaluation findings. From there, subsequent artifact versions identified and addressed areas that needed improvement.

Redundancy and Clearness

During the artifact evaluation, participants had to judge how clear and redundant the information was. But there was never any worry about unemployment. Nevertheless, one person inquired if this was the object's genuine purpose. Both writers learned from this question that many other people had been confused in the same way but differently. Because of this, it became clear that information about using it in business must be easy to share.

Artifact Appreciation

The purpose of asking whether the artifact would be valuable and why it might be helpful was to ascertain people’s opinions on its importance in MLOps implementation. Seven of the twenty participants believed it could be an excellent tool when implementing the MLOps process. Many respondents cited one reason – how well it can manage users' expectations. It is critical because it helps others understand what has to be done during the MLOps process and what they may anticipate from them.

"Yes, I think so, the project scope provided by the artifact helps identify discrepancies in people's expectations regarding different aspects of the product." – ID4

Another point that multiple interviewees keep on emphasizing is that this artifact functions as a necessary tool for understanding the prerequisites of an MLOps process before starting its implementation. It thus helps to identify and address some of the likely challenges in advance, avoiding expensive challenges emerging unnoticed.

Suggestions

Initially, two out of twenty participants provided insights on how to improve artifacts in the post-deployment phase. It included adding a requirement question that asks about the type of interface for end users, and another one mentioned end users' demographics as one of the factors that influence input data used for training models. Specifically, filtering the training data to represent the primary user group's features accurately would be necessary to tailor a voice recognition system for children.

"Including a question about end-user interface preferences is essential in an MLOps questionnaire. It guarantees conformity with user requirements and processes, resulting in enhanced acceptance and achievement. Assessing user preferences for graphical, command-line, or other interfaces provides valuable insights for making design choices and allocating resources. This strategy, which prioritizes the needs and preferences of the user, promotes cooperation and ensures that the MLOps pipeline successfully meets end-user expectations." ID 10

"The lack of end-user demographics in the questionnaire disregards a critical factor that affects MLOps training data. When customizing speech recognition systems for children, it is essential to selectively choose and improve training data, aiming at capturing specific attributes unique to them. Optimizing MLOps initiatives for specific target demographics requires understanding user demographics. For example, for the model to accurately adapt to children's speech patterns, it needs to know their age, gender, and location." ID 8.1

Four of the twenty interviewees recommended examining the allocation of roles within the artifact, particularly in scoping sections where they were the sole role assigned due to aspects, they deemed unsuitable for a business stakeholder. A single respondent disagreed, arguing that a data scientist or someone familiar with ML and the context would have been more suited to answer these questions than a business stakeholder acting alone. A pair of other interviewees echoed this sentiment, recommending either an ML expert or a data scientist for insights on the feasibility of using ML in problem-solving. According to another interviewee, the scoping and similar sections should involve all MLOps or DevOps engineers. The next version should incorporate thorough feedback and role changes according to the comments of all respondents, as they all agreed.

Four interviewees stated that the artifact's assigned responsibility requires further examination, primarily what they believe ought to be corrected for the benefit of a business stakeholder in the scoping sections, where we have exclusively given the responsibility to participants. Someone else who answered said that these questions might have been better replying by a data scientist or someone who knew ML and the situation better than by a business partner alone. Instead of relying merely on a business stakeholder, one respondent believed these questions would have benefited from the expertise of a data scientist or an individual familiar with ML and the context. A pair of other interviewees mirrored this sentiment, recommending either an ML expert or data scientist for insights regarding the feasibility of using ML in problem-solving. Also, another interviewee advised that the project manager should include either an MLOps or DevOps engineer in such sections as scoping. There was an agreement among all respondents that the next version should contain comprehensive feedback and changes of roles based on their comments.

One interviewee emphasized the crucial role of including subject-matter experts in the project. If there is no domain expert, it becomes necessary to identify an individual with domain knowledge and establish some communication links. Additionally, another interviewee supported this statement by saying that a domain expert is critical for supervised models using labeling. These people explained how domain experts work with data scientists to develop guidelines and standards for labeling tasks. In this scenario, the domain expert guides the correct labeling process while the data scientist handles the technology used for labeling. These responses demonstrated that it is necessary to include considerations about subject specialists in the next version of the artifact. It must consider the needs of subject specialists.

One interviewee underlined that including subject matter experts is crucial to the project. If there is no domain knowledge expert, it becomes essential to identify an individual with domain knowledge and establish some communication links. Another interviewee supported this statement by stating that a domain expert is necessary for supervised models using labeling. These people explained how domain experts work with data scientists to develop guidelines and standards for labeling tasks. Thus, in this scenario, the domain expert guides how to label correctly while the data scientist takes care of the technology used during the labeling process. These responses showed that it is necessary to include considerations about subject specialists in the next version of the artifact.

The interviewee proposed dissecting Requirement Question 1, which centers on discerning business issues and evaluating their appropriateness for AI solutions. While it is significant, it possesses an exceedingly wide range. After reaching a consensus, the participants suggested investigating the business challenges and exploring the possibility of using machine learning to devise business solutions. Both have enough mass to justify structural breakdowns to the required degree. Furthermore, it was somewhat incorrect to keep the term “AI,” and subsequent iterations of the object had it replaced with “ML.” By splitting the query into two sections, the artifact becomes more precise in identifying the responsibilities responsible for responding to each inquiry. Three of the twenty respondents criticized Table A.1 in the Appendix for not accurately displaying the data stage. They highlighted the significance of this phase and suggested including a broader array of inquiries. Two participants recommended including a question about the expected quantity of data, while another emphasized the importance of having a minimum number of data samples. One of the interviewees emphasized the importance of establishing a minimum level of data availability. He also suggested inquiring about the length of data retention before it becomes outdated. The significance of questions regarding the appropriate timing for data disposal and the point at which they become irrelevant is vast. Finally, an interviewee inquired about the criteria, specifically in terms of data privacy and ownership details.

An interviewee suggested breaking down Requirement Question 1 (What are the business problems, and can they be solved with AI?) because although relevant, it is highly broad-spectrum. “What are business problems?” was the consensus response to this suggestion. and “Is it possible to construct business solutions utilizing machine learning?” Both are sufficiently significant to warrant structural decomposition at the required level. Moreover, it was slightly inaccurate to maintain “AI,” and later versions of the artifact had it replaced with “ML.” Eventually, this question was segmented into two parts, making the artifact more precise in targeting the roles responsible for handling each inquiry. Three out of twenty interviewees criticized Table A.1 in the Appendix for insufficient representation of The Data stage. They stated that this phase is critical and recommended incorporating a wider variety of inquiries. Two respondents proposed a Requirement Question on anticipated data size, while another respondent noted that there should be a condition for a minimum number of data samples. In addition, an interviewee mentioned putting down a requirement for at least how much data should be present. Additionally, he suggested incorporating inquiries regarding the intended retention period of the data before its obsolescence. Crucial inquiries were those regarding the timing of data disposal and its obsolescence; ultimately, one interviewee requested Requirement Questions about the proprietorship and privacy of the data.

You may consider discussing privacy and data ownership during the scoping stage. It could be a manageable issue if it is all within one company, but when different entities are involved, it becomes a crucial problem. Handling data with care is critical, even within the same group. In addition, another interviewee emphasized that we should check for any cyclic behaviors in the data that might need monitoring. Furthermore, having a plan for detecting and handling data errors and faults would be beneficial. In subsequent sections of this thesis, we will delve deeper into these points. Finally, out of eight interviewees, two expressed their interest in having a list of requirement questions that would specifically address infrastructure requirements such as tools (e.g., specific tools for training), databases, or hardware required to support training. Although we considered this feedback, we decided it was unnecessary because we believed an IT architect could determine the needed infrastructure based on responses to other requirement questions.

5.2 Findings from the Second Design Cycle

This section focuses on the second iteration of the Design Science Research (DSR) approach. Like the previous section, this follows the sequence of problem investigation, solution candidates, and evaluation data integration. Unlike the version explained in Subsection 5.1.1, this iteration exclusively focused on knowledge acquired from semi-structured interviews, as outlined in Chapter 3. This chapter, however, provides a more detailed discussion about cycle two's connection with the challenges and best practices identified during the literature review in Chapter 2.

5.2.1 Problem Investigation

Below are the results of conducting a thematic analysis on the initial part of the first set of semi-structured interviews. Section 4.2, Subsection C, explained extensively how the survey aimed to identify common themes and patterns available with RQ1. This subsection presents these codes and themes, along with their definitions, as well as illustrative examples, and Figure 5.1. As discussed in this subsection, we use these discoveries to improve the artifact, as explained in Subsection 5.2.2.

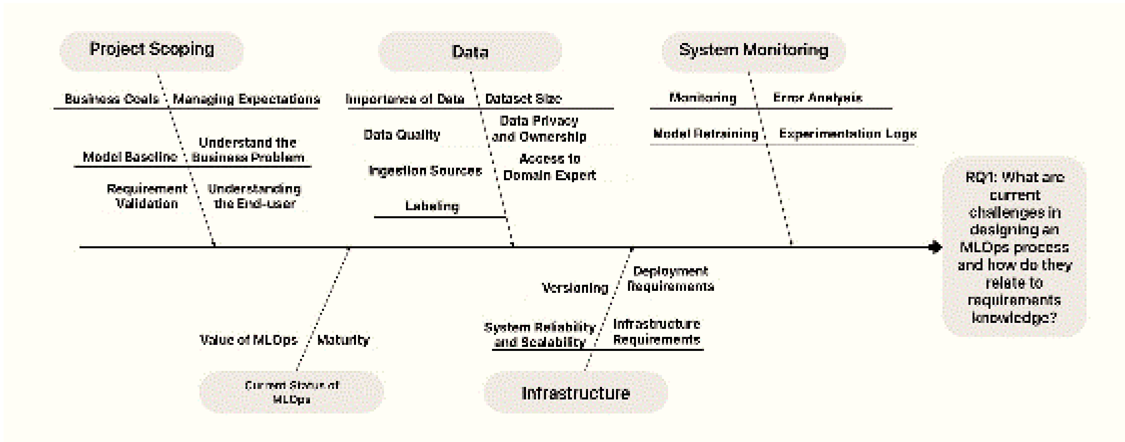


Figure 5 1:A Fishbone diagram was designed to illustrate thematically analyzed topics during an investigation of initial interrogations in semi-structured form.

5.2.1.1 Project Scoping

Under this theme, a “Project Scoping” category comprises some codes about the scoping phase of MLOps projects. For instance, business goals, understanding the business problem, model baseline, managing expectations, and requirement validation are identified codes under this theme.

“Project Scoping” was a focal point in conversations with 7 out of 20 respondents. This theme emphasizes the value of integrating various views and demands during development. It underlines the necessity for continuous customer understanding and engagement to enable successful ML products.

Business Goals: Several interviewees strongly emphasized integrating different perspectives and requirements at every stage of development to guarantee thriving ML products. In addition, one respondent talked about how challenging it is to capture business needs, integrate them into ML research, and map them seamlessly.

"This is a major challenge in machine learning development especially with regards to requirements engineering for ML. Over the last three or four years, I have been looking into capturing requirements from a business perspective and smoothly integrating them into a researcher's workflow. The idea is that researchers will be able to make direct links between their results and business requirements while conducting research, developing candidates, or doing related tasks." ID 7.1

If you look at the interviewee, he has mentioned some of the general problems in requirements engineering as a discipline. They stress that software engineers involved in machine learning must think beyond the customer's articulation. However, ML is different from other fields due to its high complexity.

"…trying to get what they want for their products but also…what they know they need however a good product should provide customers with something they didn't even know they needed. Therefore, understanding one's own needs and finding how to deliver on them…"ID6.1

Requirement Validation: A complete understanding of the requirements must be required for the form to be implemented and aligned with business objectives successfully. It aims to ensure clarity in their documentation.

"The challenging issue here is if the requirement has been correctly understood or not. Important aspects should not be overlooked since it is crucial to avoid interpreting too much or misinterpreting during implementation. Moreover, lack of sufficient detail may lead to overlooking important issues relating to the requirement." ID8

Understand the Business Problem: One of the challenges of machine learning (ML) is identifying appropriate problems for its use. Because of the complexities and resource requirements of ML and MLOps, detailed assessments must be done to determine whether a business problem necessitates an ML solution or traditional optimization methods. It is essential to critically examine the features and requirements of the issue when selecting the best ML solutions, such as choosing appropriate ML algorithms and data sets that match these specific questions.

" I regard identifying the kind of problem you want to solve as a very vital aspect during development. This encompasses identifying whether machine learning is suitable or classical estimation techniques should be used. Next, when analyzing data, you check whether it can be modeled using factors like structure or characteristics." ID5

"Being able to decipher what qualifies as an appropriate problem for machine learning presents one of the challenges. As a result, it is important to distinguish between cases where machine learning can lead to huge gains and others in which traditional optimization or coding a solution might be better alternatives. A major consideration is whether the problem has scalability advantages and whether it aligns with those of machine learning or not."ID7.1

In discussing how to recommend ML solutions to clients, one respondent noted how difficult it was for them to explain the concepts behind ML.

“Machine learning is often wanted by customers as they find it interesting and cool. However, they might not know how best to strategically implement and still prefer a much more traditional predictive approach. Behind-the-scenes integration of machine learning comes with its own set of problems. How then, will the intricacies around machine learning be effectively conveyed to such customers?”ID6.1

Model Baseline: In the scoping stage of an MLOps project, a critical consideration is defining the baseline for the final model. It is imperative to assess previous solutions and establish performance requirements.

"Typically, business stakeholders express the expectation for the new ML solution to yield improved results compared to the previous one."ID1.1

Nevertheless, we also recognize that it is sometimes crucial to consider other factors, such as the comparison between machine learning and manual labor and their impact on business value.

"... if you introduce machine learning in the right spot, like say you have a current problem and the only way to solve it requires a lot of manual labor and a lot of time. And maybe it's brittle and it must be reworked every six months or something. That's the bar you must meet with machine learning. ... if that's a 90% accurate solution during this mechanical way, discrete way, you can maybe get an 85% (with ML) ... fully automated (solution) with no manual labor and maybe even works better with scale. So, improves with scale. And if you get to that point then you've reset the expectations on the actual quality to the way you produce it and the cost." ID7.1

Managing Expectations: Five interviewees stressed the significance of making clear expectations regarding end-users in implementing ML solutions. Many people have unrealistic hopes for what AI can do, yet they know little about it, which is a big gap. It must be done at the outset so that no such problems occur, and the project succeeds.

Five respondents emphasized the significance of establishing the ML system's users' expectations when implementing an ML solution, citing a knowledge gap between ML's actual abilities and people's perceptions about what it can achieve. Avoiding any misunderstandings by outlining these possibilities early on could help this project succeed.

"However, many individuals who are aware of machine learning possess misconceptions concerning its potential. These beliefs must therefore be clarified early enough to understand clearly what the technology can and cannot do" ID7

Understanding the End-user: The interviewees recognized the significance of meticulously determining who their end-users are and how they will use their ML product or service. This work is a significant challenge due to the inherent complexity of identifying the end-user; however, it significantly determines the nature of eventual ML solutions. As a result, before developing effective ML solutions, it is necessary to investigate end-user demographics and expected model applications.

5.2.1.2 Current Status of MLOps

Covering the "Current Status of MLOps" topic resonated with 7/20 interviewees who made up a smaller thematic category grouping with two separate codes: "Value of MLOps" and "Maturity." This theme is well named because it investigates what we know about MLOps, including their justification and several roadblocks that limit their widespread deployment in business. Notably, concepts from this subject were deemed irrelevant to the emphasis of Research Question 1 (RQ1) and hence pulled out from the subsequent Design Cycle outlined in Section 6.2.2.

Value of MLOps: Interviewees explained why they thought an architecture such as MLOps was beneficial to implement. One person among those interviewed made a point regarding integrations between DevOps practices and designing or supporting machine learning products. It was only possible if he mentioned what he said about Maintenance and Development procedures. The diversity of ML products, including models, datasets, and configurations, necessitates versioning, testing, and stringent monitoring to maintain their quality and reliability.

"Highlighting the significance of DevOps in the realm of machine learning products, this integration becomes crucial for effectively organizing a variety of software artifacts inherent in software development. Under the context of machine learning, different types of artifacts need strong practices such as versioning, testing, tracking, and maintenance to ensure the robustness and reliability of deployed models." ID3

Moreover, another interviewee went deeper into the subject matter concerning bug tracking in ML projects, highlighting how a well-crafted pipeline can help deal with issues related to bug tracking within ML systems.

" I will argue that MLOps pipeline is suited for bug tracking problems though it may be costly at times. It's an appropriate way to make an effective bug-tracking system for machine learning." ID 6.1

Maturity: Repeated concern concerns regarding the level of maturity in MLOps and specific discussions about the topic were at the center of several discourses. This thesis shows that out of twenty interviewees, three were concerned about MLOps immaturity. It was observed that many technologies in this sector need to be more ambitious. These articles discuss some of these challenges, hinting at possible solutions either by narrowing down those selecting architectures for MLOps through specific demands or admitting that these challenges could be because this industry is still growing.

"Whereas software development has seen decades of maturity with standardized tools; however, such maturity does not shield it from major changes like the adoption of the Cloud or embracing DevOps regularly. On the other hand, machine learning has not undergone such a level of standardization and evolution. As a result, numerous tools exist in the field which often take highly opinionated approaches. For example, they will claim to manage all aspects starting from development up to deployment following a one-size-fits-all strategy used by many other tools in its category. Nonetheless, this approach is ineffective in software development overall and particularly inadequate for larger machine learning projects" ID7.1

Another interviewee pointed out that this industry has yet to receive the necessary funding for MLOps implementation because it is still nascent, and its value proposition is poorly defined.

5.2.1.3 Data

Several respondents said that data is at the heart of the MLOps process. Its thematic scope covers a range of issues and challenges connected with managing information, encompassing technical dimensions like ingestion sources, quality of data, who owns it, or privacy issues. Additionally, domain knowledge remains a critical element within this subject matter because it helps one better understand these complexities when dealing with labeling processes during the preparation stages. Eight interviews mentioned the title “Data” five times.

Importance of Data:

About MLOps, a few people pointed out that data is at the heart of everything. The scope in question ranges from several considerations and challenges regarding data management, hence stretching from technical aspects such as ingestion sources and data quality. In addition, this topic establishes domain knowledge and expertise as vital in understanding and navigating the intricacies of data well, always emphasizing the need for it during labeling data. This recurring theme, "Data," was found in 5 out of 20 interviews, illustrating its prominence in MLOps discourse.

“The code is mostly what makes up products in traditional software projects; however, with ML projects, they stand as small portions used to obtain an outcome as results depend heavily on training done on them.” ID4

Furthermore, one more interviewee stressed that within the machine learning area, it is essential to prioritize data over models.

“It is not only about picking preferred models and refining them in machine learning success, but it also depends on data preprocessing. In natural language processing, for instance, image processing as well as object recognition different models can provide comparable results, but some models are best in certain situations. Meanwhile, effective image preprocessing, measured signals, and information storage are the core issues at stake. This includes successful aggregation of data for accomplishing this aspect of efficiency is fundamentally crucial in making sure that there is overall success in machine learning" ID8

Dataset Size: According to the interviewees, data availability was the most crucial consideration. Inadequate existing data makes it challenging to come up with meaningful insights or predictions.

"Usually, machine learning should be better than prior approaches based on previous projects I have been involved in. This happened because the ML model outputted worse/identical results, unlike other cases. However, this was due to lack of enough data thus making it impossible to come up with a robust ML model." ID1.1

Data Quality: The success of ML projects is dependent on the data. Hence, before commencing the development of a machine learning application, we should meticulously evaluate the availability and quality of our data, as these factors have a substantial impact on the outcome of the product.

"Most machine learning projects depend on having the right data. Getting appropriate data can be difficult especially if it entails private information that may not be necessarily accessible. In such instances, limitations on gathering the needed information can affect the creation of an effective model." ID6.1

Data Privacy and Ownership: During the scoping phase, it is best to place privacy and data ownership discussions. When working within one company, individuals must still handle sensitive information prudently, even if these issues only arise occasionally.

“For instance, most times, debates about privacy and data ownership are better suited for the earlier stages of a project cycle like “scoping.” However, when treating with separate groups particularly those from different organizations then this becomes very challenging. Even within one organization caution must be taken while handling some discrete types of information thereby safeguarding them reasonably.” ID1.1

Ingestion Sources: Understanding the data source and associated information is critical for assessing potential data bias. The metadata should fully explain the data origin and identify the collection procedure, whether automated or human-driven, which is essential. By thoroughly reviewing both sourcing and collection methods, we can ensure that the information serves its intended purpose and can be trusted.

“Essentially, it is important to have a full understanding of the data source, with metadata characterizing it effectively. This will help in identifying any biases that may be present in the resulting data. The emphasis should not only be on origin but also collection methodology — whether a person-to-person conversation over the phone or an automated process happening every Tuesday night. This kind of knowledge is vital to assess and correct for biases.”ID7.1

Another participant emphasized the need to use a constant ingestion source that fits the model training. Suppose the food consumed does not match the information given during development. In that case, it might cause the model to lose its ability to understand any meaning, creating difficulties in the data pipeline. Furthermore, fixing these discrepancies later in the process may be difficult due to their complexity, which costs more resources and increases the likelihood of errors.

“Indeed, if you train AI models to understand language, they will fail once you input just part numbers.” To the AI or the data pipeline though, such data means nothing despite how important it may appear to the user who entered it. What’s more, trying to bypass reading this data and then figuring out that it is a part number for translation can be difficult. This becomes hard and necessitates additional work causing potential error” ID1.1

Data Ingestion Cycle: The field of research and the nature of the data being gathered determine the data collection methodology. When considering time series analysis or data windowing across a given timeframe, it can happen to select an inappropriate window that does not align with a natural cycle. It has the potential to result in incomplete or erroneous outcomes. Hence, neglecting to consider the uniqueness of data and employing suitable methodologies to address potential biases and limitations throughout the data-gathering process is a significant error.

“Surely, it’s an approach, which varies from one area to another. For example, can these windows collect all periods or cycles correctly when collecting time series data or information within some time windows?”ID7.1

Labeling:

Proper labeling is essential in supervised learning, as it establishes the basis for the model to comprehend its designated job. Developing appropriate labeling standards and norms requires the involvement of domain-specialist label experts and technical teams comprising data scientists. To ensure accurate categorization, these groups must communicate effectively with each other, thus mitigating biases as well as errors. Ultimately, this collaboration enhances the data quality used to train the model.

“The importance of accurate labeling in supervision cannot be overstated. There must be clear standards and guidelines for this. Regarding technical issues, it is advisable to refer to data scientists as well as the labeling staff, who are often content experts but unassociated with the matter under discussion most importantly.” ID 7.1

Access to Domain Expert: Several interviewees emphasized the crucial role that domain specialists play in ensuring accurate and efficient data analysis. The research established that even though data analysts have many skills, they may need to gain knowledge about a particular domain required for complete comprehension of their evaluated data. Thus, it is crucial to have access to experts to place data into context or interpret findings to shape hypotheses and guide analysis. In this way, they can improve the quality of their work while using other people's skills and thus enhance the utilization of the information they save.

"Knowledge is essential to have access. Nevertheless, the job of an analyst is broad and multifaceted, and she/he might not always be well-informed about some matters. Before starting any work, the process of data analysis requires consulting with specialists who will inquire into facts and develop plausible hypotheses. Analysts can be more knowledgeable and effective if they have access to professionals in that field."ID1.1

One response emphasized at length how important it was to involve subject matter experts and technical specialists in constructing rules for data labeling. However, domain experts bring unique knowledge and contextual insights essential to delivering quality work. In that case, guidance from technical personnel should help ensure these instructions are doable and understandable. With such teams as those working together closely, they can produce precise, comprehensive recommendations that render data usable all over again.

"You wouldn't ask a data engineer whether a cell is cancerous; instead, you'd need to ask a doctor who needs detailed instructions. It may be necessary for the doctor to annotate an image, therefore specific instructions catered to their area of expertise are required. The data specialist can also need technical information at the same time, such as making sure that any markings remain within predetermined parameters. The domain expert and technical specialist work together in writing comprehensive guidelines that take care of both viewpoints." ID7.1

5.2.1.4 Infrastructure:

On MLOps MLOp's infrastructure side, "Infrastructure" is quite broad regarding aspects and challenges (Dorny, 2020). All twenty respondents extensively discussed this theme, including codes such as versioning system reliability and scalability, infrastructure, and deployment requirements. These codes emphasize the importance of building and maintaining solid infrastructures for efficient data management (Dorny, 2020). Good versioning is required to preserve the scalability and dependability of the infrastructure; consequently, storage standards, task automation, and test procedures are being established (Dorny, 2020). Moreover, the theme also depicts the challenges of implementing and relocating infrastructure, emphasizing the high technical know-how required for successful implementation.

Versioning: It is clear from inputs by ML and MLOps practitioners that having access to earlier code versions is crucial in locating and fixing bugs, thus enhancing development efficiency. While version control is common in traditional software development, a machine learning (ML) environment underscores the importance of versioning across different components such as data, ML models, ML code, hyper-parameters, and outcomes.

"When there's a bug in conventional software, you revert to an already released version, configure it properly, and then run tests to find out what went wrong. But when it comes to model development in machine learning, detecting errors often involves repeating much of its initial construction. This becomes hard if the pipeline is not systematically constructed with instructions, data, labeling techniques, preprocessing stages, or even alterations on test sets or metric calculations. Besides this, the software itself which makes use of the model introduces more complex parts involving careful tracing." ID6.1

“Depending on only a version control system like Git is not enough for this. It doesn’t go far in terms of revealing the consequences of the changes you have committed and pushed. “ ID3

System Reliability and Scalability: Using well-known DevOps techniques like CI/CD (Continuous Integration/Continuous Deployment), predefined test suites, and automation made the importance of reliable and expandable infrastructures in MLOps transparent. Furthermore, these procedures are critical in MLOps. It enables developers to resolve unforeseen issues consistently and maintain control via CI/CD. In addition, it is possible to implement automated processes that adhere to predefined test patterns to mitigate the potential for inadvertent flaw introduction during code integration or resolution.

“CI/CD speeds up deployment hence enabling quick release of new features to clients or fixing issues without need for tedious manual processes. Thus, facilitating a smooth distribution process that users can consume safely and uninterrupted.” ID6.1
“It makes tests mandatory and becomes a safeguard against accidental blunders, making it harder to make errors because of the CI/CD pipeline.”ID4

This is seen through the importance of automation in ML training or other routine jobs that streamline processes and improve efficiency.

“In machine learning, however, automating the process is important as there are many runs required to comprehensively assess and understand obtained results. Thus, it would be very useful if these processes could be automated to increase efficiency and accuracy with which repetitive tasks can be handled.” ID2

“Essential for continuous automation in the development process is the same as eliminating manual regression testing together with other such repetitive activities. Certain areas may require manual attention while the objective is to automate everything possible for better efficiency.” ID6.1

Deployment Requirements: Respondents identified several challenges associated with deploying the developed ML system as the MLOps life cycle neared its conclusion. The utilization of MLOps for developing and operating the ultimate ML system is inconsequential, given the obstacles that arise during its deployment in clients' environments.

“One aspect of creating an effective machine learning model with favorable results is that challenges often arise in the process of implementation into the system, especially when it comes to integrating with edge devices. Despite its success as a model, establishing seamless integration with machines proved to be extraordinarily difficult.” ID1.1

“... The challenge occurred due to server requirements that were too strict; we performed training and testing on different servers while the actual deployment ended up on others. Appreciating running in a varied environment which could potentially have suboptimal packaging and led to unsatisfactory comprehensive testing in production settings was problematic. Additionally, there were problems linked to libraries compatibility, which directly affected the final model..” ID6.1

Infrastructure Requirements: Respondents emphasized the significance of careful thinking during the MLOps infrastructure-building process. It entails carefully selecting tools and hardware capacities and deciding whether cloud hosting is suitable.

“Essential for determining computer capacity required for training. Does that mean one needs the most powerful GPU? On the other hand constraints like starting without a large cluster of GPUs means that they can even influence model choice.” ID4

The interviewee pointed out the importance of dependable MLOps infrastructures, especially in real-time data.

“When it comes to real-time data, the infrastructure must be particularly reliable. Any disruption could lead to loss of valuable information and subsequent damage to the system.”ID1.1

Failure by practitioners in MLOps to consider necessary infrastructures automatically makes infrastructure migration challenges very pertinent and costly.

“My preference would have been using this artifact as a guide for choosing my architecture or setting up my environment including decisions on what database and data storage methods are appropriate. It is explicit and changing them may result in a costly process” ID8

5.2.1.5 System Monitoring

Regarding the “System Monitoring theme,” 6 out of 8 respondents highlighted the need for monitoring deployed system systems to maintain the excellent performance of an ML system throughout an MLOps process. Four codes under this theme include Monitoring, Model Retraining, Experimentation Logs, and Error Analysis. These codes mainly underline why system monitoring is essential for ML systems, retraining models for continued accuracy, logging experiments towards better reproducibility, and error analysis to enhance system reliability.

Monitoring: Most interviewees saw the significance of monitoring a deployed system closely; this would help gain valuable insights into its performance over time. This practice strengthens the team’s ability to optimize the model’s predictive capabilities.

“To monitor our function in real-time, we had established an initial metric system dashboard. It was a simplified way of distinguishing correct and incorrect predictions without going into further details. By using this real-time monitoring, we could see changes in how well or poorly the model performed" ID8

Error Analysis: Efficiently observing and interpreting models is challenging, particularly in cases with many issues and multiple classes on board. However, it is vital to understand why a model makes a particular prediction if it deviates significantly from what it learns from training data.

"Monitoring is difficult because there are just too many cases for anyone to sit down and go through all 1000 that come up every week. The other challenge is that there are about 100 classes so understanding the model is intricate, and sometimes predictions differ remarkably from the training dataset." ID1.1

"...Thus, the point is not just checking the predictions but also understanding why this model chooses one prediction over another."ID1.1

Model Retraining: The structure of input-output distributions across time for a model helps determine when to commence retraining and which parts should take precedence during retraining. Comparing the current input data distribution with past ones helps to identify changes that may necessitate retraining. In addition, an inspection of the output distribution detects alterations in a model's performance, which determines the retreat.

"Many statistical techniques are available for recognizing when to retrain especially if there is a classifier involved. Monitoring the input data's distribution over time and detecting variations based on measures such as mean and standard deviation provide an easy method of looking at these changes. The output distribution can be measured too to measure how much change has occurred in terms of how well they perform."ID7.1

"The alignment between the current classification by AI and actual classifications, I prefer monitoring. If there is a significant deviation, I would like to focus the model's learning on these different points."ID7.1

Experimentation Logs: Those who spoke pointed out the need to keep track of various parameters or configurations during model retraining for comparison analysis. This method helps pinpoint parameter values or settings that give the best results. A more reproducible model improves troubleshooting and problem identification during the development and deployment stages.

"I think that machine learning models need to see how outcomes differ with different parameters. Variance in response is likely to be much larger when a parameter gets adjusted. Developing a way to follow up and compare results across time, learning when exactly a certain outcome occurred as well as why it differed from others would be invaluable." ID3

Results from the Pipeline-Focused Interview

The ID1.1 interview supplement focused on the team's existing data pipelines (see Figure B.2 for the interview script). Unfortunately, there is no new information about MLOps issues. Instead, it reinforced some of the critical issues and best practices discussed earlier in general discussions that come before this section. Even though the interview yielded no new insights, the results were considered during the artifact's development, as indicated in subsection 5.2.2.

5.2.2 Solution Candidates

Following on from the problem investigation findings in cycle two (refer to subsection 5.2.1) and considering evaluation feedback received throughout cycle one (see subsection 5.1.3), we have introduced a second version of our artifact described above, which is available in Appendix Table A.3 and Table A.4". An overview of these changes is provided hereunder in Table 5.2 between the first and second artifacts made. The artifact also comes with an introductory page that tells how to use it and what its contents are (see Appendix).

Table 5. 2:Traceability matrix showing the changes made to the artifact after evaluation in the first cycle and problem analysis in the second cycle.

Requirements Question	**Change**	**Interviewee ID**
Scoping:		
What are the business problems and can they be solved with AI?	Deleted	ID4
What are the business problems?	Added	ID4
Can the business problems be solved with ML, and how?	Added	ID4
What are the metrics for success?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
What are the resources needed?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Who is the end user?	Added	ID1.1, ID6.1
How will the users interact with the model, what interface will they need?	Added	ID1.1, ID6.1
Who is the domain expert and can we access them?	Added	ID6.1, ID7.1
Data:		
Where does the data come from?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
What is the data labeling standard?	Roles changed	ID6.1, ID7.1

What meta-data should be collected?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Are there any privacy concerns regarding the data?	Added	ID1.1, ID6.1
Are there any necessary data ownership considerations?	Added	ID1.1, ID6.1, ID8
How much data is expected to be stored?	Added	ID8
When does the data become irrelevant?	Added	ID8
Are there any cyclic behaviors in the data?	Added	ID8
What is the minimum amount of data that is necessary to train the model?	Added	ID8
How will the data be acquired?	Added	ID7.1
Modeling:		
What is the model baseline?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Is it necessary to audit the model? Who should audit the model? What is the audit focus?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Which potential risks for bias exists?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
How is the input data served to the model?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
What are important business goal metrics the ML model should consider?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
What deployment constraints exist?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Deployment:		

How should the deployment process be handled?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Where should the prediction device be located?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Which software metrics are important to monitor?	Roles changed	Refined by us
Which input metrics are important to monitor?	Roles changed	Refined by us
Which output metrics are important to monitor?	Roles changed	Refined by us
How often should the model be retrained on the data gathered from deployment?	Roles changed	ID1.1, ID6.1, ID7.1, ID8
Are there any specific performance requirements?	Roles changed	ID1.1, ID6.1, ID7.1, ID8

5.2.3 Evaluation

This sub-section presents findings of the evaluation phase based on semi-structured interviews conducted during the second round. The primary purpose of these interviews was to gather input on how practical or useable the redesigned artifact was considering earlier interviews and relevant research. In this regard, suggestions were solicited regarding what interviewees found helpful in the antique, potential fixes, issues associated with its translation into architectural design, practical usage, its capacity to function well put forward as a baseline requirement, and applicability outside its context. Subsection 5.3.2 summarizes all measures performed in response to the interviewees' comments. This evaluation's findings will help shape future revisions of the artifact. In addition, Appendix Table A.17 introduces several new participants who began participating after three others were already involved in the interview process at this stage of the development cycle.

Appreciation and Concerns

All the fifteen people interviewed for this appraisal round recognized and liked the artifact. They unanimously agreed on its usefulness when asked how it could be beneficial or limited in using it at the start of an MLOps project. They highlighted that expectations are well managed through this artifact, and communication is enhanced towards requirements clarification and explaining what the MLOps is supposed to bring. Additionally, the document has been commended for being concise, lucid, and to the point.

" I think it gives great detail and covers a lot of topics very well." ID 7.2

However, all respondents raised concerns about some aspects of the artifact's overall value. A notable problem was its extensiveness as an artifact. Some participants mentioned that it could become irrelevant when producing models that do not aim at. For example, if its ML project is small, doing MLOps might seem unnecessary to them. Still, others said they wanted only a high-level overview to pitch this idea of MLOps to potential clients. MLOps is an example, and one interviewee emphasized that deploying it for application on experimental ML projects seems overkill.

" When experimenting is simple, a problem can develop when this condition occurs. If the sole objective is to perform basic experiments to address distribution drift and serve no model, then having a full-fledged MLOps development environment might be viewed as too much. Nonetheless, flexibility would still matter. The drawback lies in the risk of over-engineering, but the value of framework has to be considered about specific project demands." ID10

Suggestions

The interviews also provided some suggestions for improving the functionality of the artifact. The first suggestion was to include dependencies among Requirements

Questions, which will clarify how changes to one Requirement Question Answer (RQA) will affect other related questions further down in the tree structure. Furthermore, role-based filtering for Requirements Questions should exist so each participant can locate relevant questions at any stage of the artifact's development. Another suggestion included incorporating scalable input boxes that can accommodate detailed requirement documentation. First and foremost, it provided functionality allowing businesses to add or delete questions and customize the form format based on their preferences. Finally, there are specifics on what could happen if certain issues regarding probable system impacts are adapted. Also, in terms of functionality, there were specific recommendations for this stage of data.

Further, one interviewee proposed asking about data leakage, where test data seeps into training data, thus overestimating model performance. Furthermore, another respondent suggested the inclusion of a query regarding the frequency of streaming across models, whether it should occur at every time of high resource stress or by collecting and, after that, transmitting data in batches. Additionally, the respondent proposed the importance of determining the appropriate preprocessing position, whether on an edge device or within a pipeline. Further, another interviewee suggested asking how often streaming should occur between models, every point (high resource load), or aggregating and then sending it via batches. Moreover, this interviewee also suggested we need to know where our location should be in preprocessing – whether edge device or pipeline. Two participants recommended asking who would label their data to determine whether this process would be done internally or outsourced using a crowdsourcing approach. They both mentioned how important it is for companies like them to hire new employees and ensure they have time after school to complete assignments given at home, too!

Answers Can Translate to An Architectural Design:

During the interview, all interviewees had to answer if a fully documented version of this artifact would have enough information for a responsible person to decide which tools and technologies to utilize in MLOps architecture. All of the interviewees responded positively by saying that the effectiveness of choosing tools for the design depended on the answers to the artifact's questions.

Using the Artifact

All fifteen interviewees stated that they would use it. One respondent said they would use it as a checklist for critical things during their early days, which will become an onboarding tool for new team members.

“Of course, especially at first, I would make sure nothing important is missed while using it as a checklist. Afterward, when onboarding new team members, it becomes ideal for documentation purposes since such cases are common. It gives them a complete picture of things. Moreover, even as the project advances, it is good to have something that can refresh one’s memory. In general, having extensive documentation is always good.” ID11

Three additional interviewees said to use the artifact as a checklist throughout the earliest stages. Besides, two others recommended that it be employed as a planning tool to define critical project roles and have a more accurate estimation of project expenditures.

“Especially during commencement times in the development of product road maps. Unfortunately, when outlining features for the product team before deciding to incorporate machine learning there is an ignorance around the scale of work involved. They may think it’s as simple as bringing on board a researcher to build and deploy a model. Planning and costing are helped by using this artifact upfront for teams and their collaborations. It is always beneficial to clarify these matters early.” ID7.2

Lead to Requirements

Interviewees had to answer if the artifact, namely any replies recorded while using it, could be used to produce requirements for MLOps. Fifteen out of six participants unanimously agreed that answers could lead to conditions. Half of them emphasized that detailedness was vital in recording responses because less detailed ones would not qualify such:

...For instance, ‘’Why should we answer these questions?’ might be asked by some members of our team. That they may come up with the kind of answers that are useless. However, it is good when people are on the same page and know that all these queries have to be answered in detail! Therefore, one can use it effectively if he/she provides answers at a level of detail that is appropriate." ID6.2

The Requirement Questions in the shorter version receive praise for their clarity and conciseness. Nevertheless, some interviewees opined on using an iterative approach when interacting with the artifact, particularly for dealing with more complicated Requirement Questions.

"Most requirements are not vague but need a concrete answer as possible. Maybe an iterative method would help here. The first version should start with an Interview; evaluate its outcome and then, in another Interview, refine and address questions whose specificity was initially low to run through this circle whenever required.” ID1.2

Generalizability:

To illustrate this point further, all fifteen respondents agreed that the artifact could apply to many other areas of science apart from data science. According to one respondent, the following view about how useful it is:

"Yeah, right." In my mind, I hadn't even thought of Accenture. Without even pausing to think for a second, I was thinking about what I am working on right now, other jobs that I have done in the past, and some other projects. It is relevant. So I also didn't see any car's focus here so it is useful otherwise…" ID9

5.3 Cycle III Findings

I want to present the results achieved from this third DSR cycle, which ended up being the last one. Furthermore, this section has the same three subsections found in the previous chapter: Problem investigation, Solution candidates, and Evaluation. This cycle was primarily concerned with analyzing and discussing the artifact, as mentioned in section 3.3, and also expressed why such an approach occurred throughout this study endeavor. As a result, the subsequent Problem investigation subsection will be narrower than those included in subsections 5.1.1 and 5.2.1. Furthermore, it also implies that this cycle's Solution candidates and Evaluation sections are now more significant than the ones contained within the former processes.

5.3.1 Problem Investigation

Interviewees considered many themes to be the focus of their attention during the problem investigation in cycle 2 (see subsection 5.2.1). In this cycle's problem investigation, during the first part of the second set of semi-structured interviews, it became more apparent that some of these themes were important as they were emphasized again by several interviewees. Subsection 5.3.1.1 lists the articles confirmed in this DSR cycle. The remaining new pieces come as subsubsections:

5.3.1.1 Confirmed Themes

Previous problem investigations elicited a lot of discussion about "Project Scoping" among interviewees; this theme became even more apparent in the second round of interviews, where many respondents emphasized similar or identical codes, such as Understanding the Business Problem and Business Goals. Furthermore, another new code highlights the importance of understanding ML. Though it might seem like something granted, given the need to work with ML, MLOps involves multidisciplinary team collaboration, which makes it challenging to ensure that everyone understands how ML works, mainly when working with an ambiguous group. In this regard, the topic “Project Scoping” becomes more relevant since various interviewees have acknowledged the importance of Understanding the Business Problem and Business Goals, which dovetails with the need to understand ML models and development procedures beneath. Additionally, “System Monitoring” was a theme identified during the previous cycle (see subsubsection 5.2.1.5). This theme did not involve any new codes as far as this problem investigation cycle was concerned. Nonetheless, the data provided more evidence of its significance as interviewees repeated and emphasized the same topics.

5.3.1.2 Developing a Model

The theme “Developing a model” encapsulates difficulties and best practices related to model development. For instance, two out of fifteen respondents discussed it during this cycle, focusing on Data leakage, Versioning, Development environments, and CI/CD, among other points. In an earlier problem investigation cycle (5.2), Versioning and CI/CD were under “Infrastructure.” At the same time, these codes did not provide any new insights into current findings other than to validate what had already been across. Therefore, this subsection introduces new evidence for codes associated with data leakage and development environments.

Data Leakage: One interviewee highlighted data leakage as one aspect of machine learning model development, particularly in research. Using the data from the test set to evaluate how well the model performed on the training set is an integral aspect of building a model. Therefore, this can lead to overfitting and make a model look better than it is when it comes to new and unseen information. It means that this concern, according to the interviewee, goes far beyond just research, where you may end up having substandard model deployment due to an organization's overestimation of its performance.

"The interviewee agreed saying "Yes I think that because I am a researcher, and I would like to use my results for presentation in conferences and writing papers." For me, I want you not to leak test data through your models. So, if there's one thing that I always try as much as possible to avoid it would be not having any evidence of data getting leaked out through my models. So, you don't overestimate the performance of your models."ID10

Development Environments: Another interviewee also emphasized the importance of having the development environment closely mimic a production environment when creating a model for production use. It is crucial because different factors, such as software dependencies, hardware specifications, and operating systems, can affect an ML model's performance. Consequently, it is essential to ensure that a machine learning model's training and development environment is similar to its final production environment. It will help identify and fix any issues before they impact the model's functionality and efficiency.

"Certainly, once you have a model ready for production deployment, it is necessary to maintain consistency from the development stage onwards. It would be best if the system or service used for developing this model was like what will appear at runtime. As much as possible, this similarity ensures that not only does the development environment but also that utilized for training models approximate those that will prevail in a real production setting. A uniform environment during its growth ensures an orderly transition and smooth performance." ID9

5.3.1.3 Requirement Management

Three out of fifteen participants discussed the theme "Requirement Management" during the interviews. This theme entails vital elements such as non-functional requirements, adaptability to dynamic environments, and constant engagement with requirements. It discusses the importance of including major non-functional requirements when building ML models for production, differentiating traditional ML projects from those with MLOps, and highlighting the continuous need to handle requirements throughout the project lifecycle.

5.3.2 Solution Candidates

According to the problem investigation findings in this cycle and feedback received during the second evaluation cycle (see subsection 5.2.3), I developed my artifact's third and final version. This artifact is available in Chapter Three only. Look at Table 5.3 to see how it changed between artifacts two and three, shown below, or compare them side by side if you like; otherwise, use it as an exemplification of all modifications made on the second version of this one. A picture on the first page also showcased MLOps stages and their continuous iterative nature, as illustrated in Figure A.1.

Table 5. 3: The session reviewed artifact modifications post-second cycle evaluation, attributing specific changes to interviewees.

Requirements Question	**Change**	**Interviewee ID**
Scoping:		
Is there any budget limit for the computation necessary to train the model?	Added	ID10

Data:		
Who will label the data?	Added	ID6.2, ID10
For streaming data, what is the minimum frequency of data necessary to meet business goals?	Added	ID1.2
Modeling:		
Deployment:		

5.3.3 Evaluation

The following subsection will present findings from the final evaluation via interviews; within this section, feedback received during the discussion will be discussed, with future potential actions outlined in Chapter 7; the organization and conduct of the interview, including its objectives, are presented in Section 3.3. The appendix contains a complete description of the hypothetical case instances given to participants throughout the debate.

The comments in the Appendix show that all participants used a learned artifact to scope a hypothetical project case, indicating that the interview was successful. Participants in the interview survey uniformly agreed that using an artifact facilitated extensive discussion of all key features of the imagined situation. According to participants, the artifact itself served as a guiding tool during the various ML stages, preventing them from overlooking any critical aspects that would have been missed without it, thus highlighting that this artifact is helpful as a checklist and can streamline required thought processes for MLOps processes, as evidenced by active discussions during the interview.

It takes effort to determine the precise period for employing an artifact. Over two hours, participants examined and discussed the entire artifact in fair detail. Results show that complicated situations necessitate more time than two hours.

Guests suggested reminding users that they do not need to answer questions in a specific order while using Requirements Questions. Although there are sequential MLOps processes, several issues in the scope scoping stage were challenging to answer without dealing with later-stage needs problems. For example, "What resources are essential?" caused a problem. Digitizing this item might be the solution, as participants have noted and have previously mentioned in review cycles (Silva et al., 2018).

The interview and questionnaire participants indicated that the artifact should be up-to-date with new requirements or that existing ones must change. It may expand the question "Can the business challenges be solved using ML, and how?" to ask whether ML is an appropriate solution to these specific business problems. Instead, request a different requirement question that takes this into account. It is an important topic because there may be more straightforward solutions than developing a whole ML system.

CHAPTER VI:

DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

Imagine you are a data scientist at ABC company, where you have developed a deep learning model for price prediction to automate the pricing setup based on users' previous purchase history. You have tested this model very diligently and done hyperparameter tuning countless times. It has an outstanding outcome that will generate an additional 2 million dollars for your company annually. You deploy this model in the system and get the terrible news that your deep learning model has mispriced all items and gone uneven. You have pushback pushed back this model, and your company has lost 4 million dollars in this process. You have started looking at every step of model building and retested the model by building it from scratch, and the model was looking fine. Are there any changes occurring in price distribution? Or did data preparation methods go wrong?

Or has data quality gone down because of changes in process changes? You have used your brain wildly, but nothing was coming out. As a result, technical skills help tackle MLOps/DLOps, but we must provide a flexible matrix that will be visible over time. How will we assess the technology threat in the system, and what is the payout for this threat? In general, noting that the team can move fast does not constitute proof of a low technology threat or preferred procedures. Still, the overall cost of technological threats may increase as time passes. (D. Sculley,2015; Morgenthaler,2012)

For companies aiming to expand their machine learning and deep learning models into a cloud environment, ML-Ops (Machine Learning Operations) and DL-Ops (Deep Learning Operations) are the emerging areas of focus. For business problems and company growth in a cloud environment, these are the key areas:

Model Deployment and Serving: ML-Ops / DL-Ops comprises deploying and serving machine learning models in a cloud environment. In particular, this involves incorporating models into existing systems and workflows, managing infrastructure and resources, and ensuring high availability and performance.

Model Monitoring and Management: When it comes to monitoring machine learning models, ML-Ops / DL-Ops is about life cycle management. It involves observing model performance in production, drift detection, and retraining to maintain accuracy or consistency.

Data Management and Governance: This includes overseeing data pipelines while ensuring quality security and privacy, among other aspects under the ML-Ops/DL-ops umbrella. In addition, it covers the build-up/maintenance of data catalogs besides version control auditing.

Model Version Control: On the part of ML-Ops / DL-Ops, it requires handling multiple versions of the model and tracking their evolution over time, which encompasses version control, including comparing models' rollback options for stability purposes.

Collaboration and Teamwork: Data scientists, engineers, and business stakeholders must collaborate to create and deploy ML models, which include communication, coordination, and data sharing, among others.

Compliance and Regulatory Requirements: The company has to adhere to some rules and regulations/standards set by the concerned authorities regarding ML-Ops / DL-Ops. Among these are model and process auditability, data privacy, and security.

Scalability and Performance: Scaling machine learning models and infrastructure involved in ML-Ops / DL-Ops help them cope with rising data workloads. It consists in optimizing resources such as performance, cost, etc.

In conclusion, for companies seeking to scale their machine learning and deep learning models in a cloud environment, the MLOps/DLops area must be considered critical. It is technically challenging because it creates multidisciplinary challenges due to its interdisciplinary nature, which involves collaboration among diverse groups within the organization. In the whole life cycle of machine learning approaches, MLOps is valuable beyond many other stakeholders, including those who address the specific needs of data scientists. It will then give an insight into how they can realize this through MLOPs by data scientists:

6.1 What Data Scientists Seek to Accomplish: -

Model Development:

Goal: To come up with accurate machine learning models that work based on business requirements and data analysis.

Challenges: Making models that generalize well beyond training data, handling hyperparameters properly, and quickly running through experiments.

Experimentation and Prototyping:

Goal: Data scientists experiment with different algorithms, features, and model architectures to obtain the best solution.

Challenges: Experiment management/tracking, model performance comparison, rapid prototyping for iterative development.

Collaboration:

Goal: Work together with other teams like data engineers, operations, and business stakeholders across functions.

Challenges: Communication gaps; versioning; seamless sharing of models and code among collaborators.

Scalability and Production Deployment:

Goal: Turn experimental models into production-ready solutions that can handle real-world data and user demands.

Challenges: Scaling out model deployments; ensuring reliability; considering infrastructure resources.

Monitoring and Maintenance:

Goal: To maintain the continuous performance and relevance of deployed models.

Challenges: Protecting against model drift, handling changing data, and putting in place strategies for ongoing model improvement.

Reproducibility and Documentation:

Goal: Making sure that experiments and models can be repeated for confirmation as well as audibility basis.

Challenges: Documenting experiments, tracking dependencies, and maintaining a transparent history of model configurations plus versions.

6.2 How MLOps Can Help Data Scientists: -

Automated Pipelines:

How MLOps Helps: Implementing automated pipelines for model training, testing, and deployment reduces manual intervention, streamlines processes, and minimizes errors.

Collaborative Platforms:

How MLOps Helps: Team members can use platforms that allow communication, and version control systems among others to facilitate their work with other stakeholders which enhances joint research endeavors.

Experiment Tracking and Management:

How MLOps Helps: By providing tools that track experiments discretely and manage them in a unified dashboard where data scientists sign parameters or metrics or results logged across models making it easier to compare models or reproduce the same at later times.

Model Versioning and Deployment:

Why MLOps is Helpful: One of the issues that can be addressed by MLOps is model versioning and consistency in different environments, including those that deal with scalability, repeatability, and deployment.

Monitoring and Maintenance Tools:

How MLOps Helps: The other thing about MLOps is that it comes with monitoring tools for tracking the deployed models' performance, drift detection among many others as well as automated retraining processes to support the continued good performance of the model.

Infrastructure Scalability:

How MLOps Helps*:* This also allows for scalable infrastructures such as cloud-native solutions like containerization which are critical for handling increased workloads and deployment environments that are diverse.

Security and Compliance Integration:

How MLOps Helps: Additionally, certain features within these platforms make it possible to integrate security measures and help achieve data privacy plus regulatory compliances to eliminate concerns about the protection of sensitive information.

Documentation and Model Registry:

How MLOps Helps: On a separate note, documentation tools together with model registry functionalities allow data scientists to keep proper records regarding their experiments carried out, models developed as well and configurations used hence promoting reproducibility and auditability purposes.

In summary, MLOps provides a set of practices and tools that enable data scientists to streamline their workflows, work together effectively, and shift from experimental models to reliable, scalable, and maintainable production solutions. Data scientists leading machine learning projects benefit from integrating MLOps practices in terms of efficiency and success.

This study has examined the findings comprehensively. Structurally, it is organized based on the questions outlined above, with each section addressing a specific research question hypothesized at the beginning of this study. Furthermore, an additional area is reserved to explore issues relating to validity risks that could have influenced the reliability and generalizability of our results. The discussions presented in this chapter aim to give a detailed and critical analysis of the findings while shedding light on what the study means and pointing out areas for further investigation.

6.3 Research Questions: -

RQ1: What are the current challenges in designing an MLOps process and how do they relate to requirements knowledge?

A literature-based framework was developed after a thorough examination of design cycle challenges and the knowledge requirements problem. Challenges identified as **P1 Data Drift and P2 Concept Drift** are closely associated with non-functional requirements relating to performance. Addressing these concerns with effectiveness necessitates adding monitoring capabilities into the MLOps architecture, making it a functional condition.

Through subsequent problem investigations conducted during the design cycles, several interviewees highlighted the need for monitoring. This consensus by academic research and expert interviews amplifies the crucial role played by this aspect in the operation of MLOps systems. By showing proof from various sources, monitoring is an essential part of dealing with non-functional requirements issues, making MLOps architectures more reliable and efficient.

It indicates that the **P4 performance** during the serving issue is also a non-functional performance requirement, increasing scalability. It shows that at no point should merely meeting performance benchmarks suffice for a system. It has to scale up, handle vast volumes of traffic, and maintain high-performance levels under heavy loads. These insights appear in various interviews, with interviewees emphasizing the need for a resilient and scalable infrastructure (Maiya et al., 2020). Additionally, it aligns with MLOps' core philosophy, stressing its adoption of established DevOps methodologies and a reliable and scalable operational framework in addressing performance and scalability challenges associated with MLOps systems.

Known as **P5 Disorganized Data**, this challenge defines using unprocessed data for model training, especially when collected from different sources. It closely reflects against BP3 Data Quality and Labeling – an essential step in ensuring unbiased data quality and labeling guidelines. Establishing data and labeling criteria is inextricably linked to the system's needs. Such requirements may specify the data types required for model building and the labeling method to comply with the formats.

The cohabitation of data-labeling criteria and system requirements demonstrates how well-defined protocols influence model training data. In addition, insights from expert interviews corroborate the need for capturing massive volumes of high-quality data coming from ingestion sources, which are consistently reliable consistently reliable ingestion sources. Furthermore, involving domain experts in developing or refining labeling guidelines is another recurring theme that highlights the collaborative nature and multi-disciplinary approach to addressing problems occasioned by disordered MLOps data.

The **P6 Sustainable** MLOps task entails three core aspects conflicting with sustainable MLOps: explainability, fairness, and accountability- non-functional requirements. Even though there were no explicit comments about equality during the interviews, traceability, explainability, and accountability came up frequently. MLOps runs on this sort of thinking.

Nevertheless, there is a need for a more comprehensive examination of non-functional requirements particularly associated with MLOps. Nonetheless, it is essential to note that there has yet to be an explicit agreement on fairness regarding the MLOps. It argues for a deeper understanding of why integrating fairness into practices within the domain of MLOps may be so challenging and complex that it requires further investigation. However, using the artifact in this thesis, we identified difficulties and best practices for traceability. The identification also indicates how useful it can be for practitioners who want a structured approach while addressing hurdles and trying to implement the best strategies for traceability when applied in MLOps.

In contrast, within the specific field of MLOps, there is little or no literature on such requirements, whereas non-functional machine learning (ML) requirements have attracted significant research interest. This lack of research aligns with the argument that non-functional requirements mainly apply to ML models. However, there is a good reason for this constraint, as MLOps helps practitioners to fulfill these needs. For instance, MLOPs involve systems monitoring, data versioning, and model versioning.

MLOps enables practitioners to effectively address and meet multiple non-functional ML requirements while appreciating its ability to make the operationalization and deployment of ML models more practical. This view supports using MLOps to improve the reliability, performance, and interpretability of these models in real-world applications. Studying how non-functional ML requirements connect to MLOps practices is more critical.

RQ2: Which potential solution exists to mitigate the challenges of developing an MLOps process grounded in requirements engineering?

The MLOps Requirements Form relies on a solid foundation, including a thorough literature study and information gathered from interviews with industry experts. This meticulous procedure ensures that the Requirement Questions reflect the most recent best practices in MLOps and are specific to address issues that all MLOps implementations confront. Furthermore, the Requirements Questions are adaptable to any project or business. It transforms the artifact into a flexible solution, providing teams with a broad and versatile foundation for collaboration.

No machine learning model is specific to a project or industry. Instead, it is a versatile instrument with multiple applications. Hence, the MLOps Requirements Form becomes an invaluable tool for eliciting and documenting requirements for MLOps processes, offering a nuanced and adaptive basis for teams participating in MLOps activities.

Furthermore, our assessment of related literature revealed that this particular form differed from similar instruments. Although given the existing conditions to support the selection of technologies for MLOps, our solution is unique because it does not have a dedicated form for eliciting MLOps requirements. Per our understanding, this MLOps

Requirements Form is an original contribution to the discipline aimed at addressing a gap and offering fresh ways of capturing and recording MLOps requirements.

While one could think that filling out the MLOps Requirements Form must occur sequentially, this strictness is not required. It aims to be compatible with several workspaces, such as Waterfall or Agile processes, so its application may be specific to any context. In Agile environments, digitizing the artifact might be particularly useful as it simplifies versioning activities and enhances efficiency. This suggestion follows our forthcoming work recommendations covered in section 6.2.

Thus, this form has been designed intentionally with flexibility so that different project management methodologies currently in the industry are well-catered to. Without a fixed sequence for responding to the form, it becomes adaptable, allowing teams to customize what they do based on their project specifics and dynamics. The intended digitalization of the artifact in an Agile context is meant to exploit the iterative and collaborative nature of Agile methodologies, thus serving as a potential avenue for improved traceability and ease of version control.

Maintain the MLOps Requirements Form in good condition; a critical challenge was to ascertain the appropriate level of specificity for the Requirement Questions, ensuring a manageable growth of this artifact. This problem arises from differences between MLOps requirements, such as the kind of ML model used, data sources employed, or computing infrastructure selected for deployment and uses case design. For this reason, some non-generic Requirement Questions are deliberately omitted from the MLOps Requirements Form so that teams can customize their MLOps according to their unique and peculiar needs.

This intentional omission of more detailed non-general questions is a conscious design decision to achieve completeness without excessiveness. When teams are allowed to set their own MLOps needs, this artifact remains adaptable across various projects; hence, it retains its versatile nature of requirements elicitation and documentation in the changing landscape of MLOps implementations.

It is essential to understand that besides these inherent limitations, the MLOps Requirements Form artifact serves as a handy tool for addressing another critical challenge – ensuring that MLOps processes align with stakeholders' and developers' purposes and expectations. Through systematic use of this form, the team gets an opportunity to identify and prioritize critical MLOps requirements, leading to precise, unambiguous requirements concerning the overall project goals. Thus, the importance of the MLOps Requirements Form lies in enabling the development of more Stakeholder-developer-friendly MLOps processes. In other words, this mismatch fosters lower chances of successful implementation of Machine Learning Operations (MLOps), leading to poor efficiency and poor outcomes for any initiative implemented under such circumstances.

It is crucial because it must be stated explicitly from the start that despite being grounded on RE principles, there needs to be more than a one-size-fits-all approach to guarantee success when designing a process for MLOps using an RE methodology. Alternative potential solutions are critical to be aware of while still being promising tools to alleviate such problems. Finally, given their specific needs and operations, the artifact may only apply to some organizations.

Including Requirements Engineering in the scoping phase of an MLOps project is a critical development, even though it has received little attention thus far. Its importance lies in its ability to systematically confront challenges and offer a structured means by which requirements unique to the MLOps sector can be elicited and documented. It is essential to acknowledge that the form does not provide a universal solution but continues to be a valuable asset in the realm of MLOps Project scoping and implementation that has been largely unknown and left out thus far.

In conclusion, the MLOps Requirements Form artifact also shows excellent promise as a solution for establishing an MLOps process founded on RE (Requirements Engineering). This artifact provides an efficient method for gathering and documenting MLOps needs in ways that have considerable promise in promoting the implementation of such processes. However, it is essential to acknowledge that the MLOps Requirements Form does not represent a comprehensive resolution. Instead, it acts as one of many tools that organizations can use to deal with the complexities and subtleties of making MLOps work. The artifact's structured format helps companies deal with the intricacies involved in MLOps projects, thereby leading to improved project scoping, requirement gathering, and documentation processes. Although the MLOps Requirements Form can be beneficial, its integration into a larger strategy necessitates knowledge of numerous other tools and methodologies that can work in tandem to ensure a comprehensive MLOps implementation.

RQ3: How well does the potential solution mitigate the requirements-related problems with developing an MLOps process?

In my thesis, an evaluation iteration of the artifact came after every design cycle in our Design Science Research (DSR) approach. Participants in the first round of evaluation interviews had a positive impression of the participants' artifacts. We chose to keep improving the original artifact because many interviewees gave us good feedback. Moreover, subsection 6.2.3 provides further evidence that it did not go without being commended during another set of evaluative interviews.

This iterative evaluation process was a crucial component of the DSR methodology, allowing for validating and refining the artifact using feedback collected from real-world situations and industry specialists' insights. Positive feedback from multiple reviews indicates that this artifact is well-received and helps with issues with developing RE-based MLOps processes.

Interviews highlighted a visible gap in industrial needs for a tool like our artifact, offering some justification for its highly positive assessment results. In addition, further research into the literature suggests the possibility of a tool that can help with early requirement capture for MLOps.

Identifying this need in the market aligns with the overarching objective of our Design Science Research (DSR) approach, which is to provide novel answers to real-world problems. The positive feedback indicates that the artifact was deemed valuable and relevant within the domain of MLOps and highlights its potential for meeting an unmet need. It combines empirical information from interviews and academic studies to stress how vital the artifact is for closing a critical gap in the industry by providing practical tools earlier on during MLOps project development.

At the same time, these two initial evaluation loops acted as opportunities to detect flaws, errors, repetitions, or vagueness contained in those first two artifact versions (see Appendix A for prior versions of artifacts). An apparent decrease followed these assessments in comments that could be identified as unfavorable in the second artifact assessment compared to the first. This trend demonstrates a level of finality and saturation, especially concerning how well it addresses Requirement Questions at all.

Refinement occurs iteratively in this process, with each cycle bringing new feedback through criticisms and comments. With the client's expectations for the MLOps requirements adequately documented, we are heading towards a more complete and older artifact.

Phase two of the evaluation interviews described in section 6.2.3 aimed to glean further suggestions for improving the artifact's adaptability and usefulness. All indications pointed in the same direction: the digitization bias to make it more flexible through such things as writing cells can be adjusted automatically, use of user and scenario Requirement Questions, and general dynamism and customization of the artifact. This switch makes it clear that the first feedback was content-oriented. In contrast, the second cycle feedback related to usability and use cases shows that the artifact is very suitable for various applications.

The evaluation has, therefore, established that it's usable in many scenarios. These evaluators recommend steps to digitize this artifact in the later stages of this thesis. Due to time constraints, it is not possible to digitally implement these suggestions in this production. Notwithstanding this, the recommendations for further investigation encompass these concepts (section 5.2).

With two participants requesting the artifact's introduction to other organizations for practical usage, the evaluation feedback showed overwhelming support. Moreover, the unanimous agreement among all participants showed the artifact's capability to build a collaborative environment. As such, active engagement with discussions led to a better understanding of both the Machine Learning (ML) problem and corresponding solutions. Notably, the interview yielded positive results for a more straightforward case involving three individuals; nevertheless, it is crucial to comprehend that the outcomes may differ in more intricate cases featuring a large number of expert participants.

The responses obtained from the questionnaires completed by the participants were consistent with prior opinions. The artifact garnered recognition for its capacity to bring attention to frequently disregarded issues. It emphasized its potential to make a substantial contribution during the scoping stage of an MLOps system. Nonetheless, this recurring theme hints at the fact that digitization could enhance the flexibility of this component. Thus, it becomes evident from this recurrent view that the artifact is almost complete in terms of content and saturation level.

6.4 Conclusion

The present study’s main objective was to enhance knowledge of the relationship between MLOps and Requirements Engineering (RE). It is essential because it will help MLOps practitioners have benefits similar to those in RE applied to traditional software development. Its specific purpose was facilitating, determining, and documenting goals and objectives for Machine Learning (ML) projects. In this regard, clearly understanding desired outcomes from the ML system helps align stakeholders’ expectations. Moreover, our principal objective was to produce an artifact that may be a road map for examining, via design science research methods, how organizations can implement MLOps to establish more standardized and predictable maintenance of machine learning models, with a focus on requirements engineering.

This study followed principles of Design Science Research (DSR), which allowed for identifying extant approaches towards implementing MLOps processes, exploring available best practices, and investigating challenges faced while creating an MLOps process. Additionally, this study sought to assess the efficiency of different measures to deal with the above challenges, establish new knowledge, and provide insights that may guide and enhance MLOps and RE integration.

Our findings have highlighted the perceived importance of including Requirements Engineering (RE) into in Machine Learning Operations (MLOps) practices for successfully implementing and operating ML models. Our exploration identified various challenges and best practices linked to MLOps implementation. This scan investigates strategies for mitigating these challenges while effectively integrating these best practices. The outcome was the development of the MLOps Requirements Form, an artifact intended primarily for practitioners. It is considered a blueprint that will aid practitioners in constructing secure MLOps processes from an RE perspective, thus offering a systematic approach to addressing intricacies arising from combining these two fields.

Our study produced a pragmatic guide for executive managers and technical leaders in MLOps implementation by refining an MLOps requirements form. It is worth noting that the feedback from respected experts in this area has been overwhelmingly positive, thereby validating its effectiveness and possible use in practical environments. Our findings show that practitioners across various fields can benefit from these insights, enabling them to navigate through and surmount challenges encountered while implementing Machine Learning (ML) models into production. Therefore, the outcome of research efforts like this one adds to theoretical understanding and provides a real-life tool for bettering MLOps processes within live applications.

In summary, our report presents new viewpoints on MLOps at the crossroads with RE, thus contributing to the growing knowledge about RE for MLOps. This contribution has far-reaching implications for academia and practice: academia predominantly benefits from mutual complementation between these concepts – MLOps and RE, whereas industry derives utility value from this artifact itself. We have been using a largely theoretical approach in our evaluation of the artifact in this study, suggesting that it would be valuable for future researchers to apply it through a case study and evaluate its performance and implications in the real world. Such an approach will enable us to understand better the artifact’s practicality and how it may influence MLOps processes across various application scenarios.

6.5 Future Work

Even though there are some merits regarding the artifact's current state, it is still possible to make some changes to add to its performance. We propose digitizing this artifact as a means to raise its quality. By doing so, more elements will be brought in, including filtering capabilities and dependencies management, as well as scalable input boxes and information that gives insight into what could happen if one ignores specific questions within this artifact.

Digitization may suggest integrating filter mechanisms to enable users to customize the artifact depending on their roles and identify specific questions for every stage of the MLOps process. It is possible to establish question dependence that allows the reveal of downstream requirements prone to changes in the initial set of requirements. Furthermore, users could create more detailed requirements within the same document instead of having separate documentation by integrating scalable input boxes. Moreover, this information should include the consequences of leaving some other things unaddressed, making it clear where there is a massive need for attention.

These recommended improvements reflect current technological advances, showing the significance of an adaptive and complex artifact that can meet the changing demands of MLOps practitioners.

Furthermore, there might be other challenges and best practices relating to MLOps not fully explored in this study. For this reason, further exploration into these under-researched domains has the potential for valuable insights with broader implications. Furthermore, upcoming tasks include performing a case study to test the practicality of this tool, especially in building MLOps structures. A meaningful understanding of the artifact's practical feasibility and identification of potential areas for improvement necessitate undertaking an empirical analysis.

From such an outlook, it becomes clear that research is an ever-evolving process that requires continuous improvement of techniques and methodologies due to changing industry practices and challenges for knowledge acquisition.

REFERENCES

AirbnbEng (2015). How Airbnb uses Machine Learning to Detect Host Preferences. [online] Medium. Available at: https://medium.com/Airbnb-engineering/how-Airbnb-uses-machine-learning-to-detect-host-preferences-18ce07150fa3.

A. B. Kolltveit and J. Li, "Operationalizing machine learning models - a systematic literature review", in *2022 IEEE/ACM 1st International Workshop on Software Engineering for Responsible Artificial Intelligence (SE4RAI)*, 2022, pp. 1–8. doi: 10.1145/3526073.3527584

Adadi, A. (2021). A survey on data-efficient algorithms in big data era. Journal of Big Data, 8(1). doi:https://doi.org/10.1186/s40537-021-00419-9.

Alla, S.; Adari, S.K. What Is MLOps? In Beginning MLOps with MLFlow: Deploy Models in AWS SageMaker, Google Cloud, and Microsoft Azure; Apress: Berkeley, CA, USA, 2021; pp. 79–124.

A. Ng, Machine learning engineering for production (mlops) specialization, Coursera, 2023 [Online]. [Online]. Available: https://www.coursera.org/specializations/machine-learning-engineering-for-productionmlops.

A.. Paleyes, R.-G. Urma, and N. D. Lawrence, "Challenges in deploying machine learning: A survey of case studies", vol. 55, no. 6, 2022, issn: 0360-0300. doi: 10.1145/3533378. [Online]. Available: https://doi.org/10.1145/3533378.

Ashmore, R., Calinescu, R. and Paterson, C. (2021). Assuring the Machine Learning Lifecycle. ACM Computing Surveys, 54(5), pp.1-39. doi:10.1145/3453444.

A. R. Hevner, "A three cycle view of design science research", Scandinavian journal of information systems, vol. 19, no. 2, p. 4, 2007.

A. Vogelsang and M. Borg, "Requirements engineering for machine learning: Perspectives from data scientists", in 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW), IEEE, 2019, pp. 245–251

Capizzi, A., Distefano, S. and Mazzara, M. (2020). From DevOps to DevDataOps: Data Management in DevOps Processes. Software Engineering Aspects of Continuous Development and New Paradigms of Software Production and Deployment, pp.52-62. doi:10.1007/978-3-030-39306-9_4.

Challenges with ML in Production. 2022. Available online: https://docs.cloudera.com/machine-learning/1.1/product/topics/ml-challenges-in-prod.html

dotscience.com. (n.d.). Dotscience Blog. Why do Data Scientists Need DevOps for Machine Learning (MLOps)? [online] Available at: https://dotscience.com/blog/2019-10-21-devops-for-ml/ [Accessed 29 Aug. 2022].

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Franc¸ois Crespo, Dan Dennison, "Hidden Technical Debt in Machine Learning Systems" – 2015

E. Knauss, "Constructive master's thesis work in industry: Guidelines for applying design science research", in 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET), 2021, pp. 110–121. doi: 10.1109/ICSE-SEET52601.2021.

Guru99 (2019). Supervised vs Unsupervised Learning: Key Differences. [online] Guru99.com. Available at: https://www.guru99.com/supervised-vs-unsupervised-learning.html.

H. Miao, A. Li, L. S. Davis, A. Deshpande. "Towards Unified Data and Lifecycle Management for Deep Learning". In: 2017 IEEE 33rd International Conference on Data Engineering (ICDE). IEEE, Apr. 2017, pp. 571–582. isbn: 978-1-5090-6543-1.doi: 10.1109/ICDE.2017.112 (cit. on pp. 22, 23).

H. Villamizar, T. Escovedo, and M. Kalinowski, “Requirements engineering for machine learning: A systematic mapping study”, in 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), IEEE, 2021, pp. 29–36.

I. C. Weber, P. Hirmer, P. Reimann, H. Schwarz. “A New Process Model for the Comprehensive Management of Machine Learning Models”. In: Proceedings of the 21st International Conference on Enterprise Information Systems Iceis (2019), pp. 415– 422. doi: 10.5220/0007725304150422 (cit. on pp. 17, 22, 23).

J. D. Morgenthaler, M. Gridnev, R. Sauciuc, and S. Bhansali. Searching for build debt: Experiences managing technical debt at google. In Proceedings of the Third International Workshop on Managing Technical Debt, 2012.

J. Saldaña, “The coding manual for qualitative researchers”, The coding manual for qualitative researchers, pp. 1–440, 2021.

Jones, J.; Ionita, A.; Mihai, I.C. AI and IoT Mapping and the Transition to an Interconnected Cyber Defence and Intelligence Capabilities. Int. Conf. Cybersecur. Cybercrime 2022, 9, 5–22.

Karlaš, B., Interlandi, M., Renggli, C., Wu, W., Zhang, C., Mukunthu Iyappan Babu, D., Edwards, J., Lauren, C., Xu, A. and Weimer, M. (2020). Building Continuous Integration Services for Machine Learning. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. doi:10.1145/3394486.3403290.

Letouzey, J.-L. and Whelan, D. (n.d.). Introduction to the Technical Debt Concept What is Technical Debt? Where does it comes from? [online] Available at: https://www.agilealliance.org/wp-content/uploads/2016/05/IntroductiontotheTechnicalDebtConcept-V-02.pdf [Accessed 29 Aug. 2022].

Leite, L.; Rocha, C.; Kon, F.; Milojicic, D.; Meirelles, P. A Survey of DevOps Concepts and Challenges. ACM Comput. Surv. 2019, 52.

L. Baier, N. Kühl, and G. Satzger, "How to cope with change?-preserving validity of predictive services over time", 2019

Mayr, A., Kißkalt, D., Meiners, M., Lutz, B., Schäfer, F., Seidel, R., Selmaier, A., Fuchs, J., Metzner, M., Blank, A. and Franke, J. (2019). Machine Learning in Production - Potentials, Challenges and Exemplary Applications. Procedia CIRP, 86, pp.49-54. doi:10.1016/j.procir.2020.01.035.

Miao, H., Li, A., Davis, L.S. and Deshpande, A. (2017). Towards Unified Data and Lifecycle Management for Deep Learning. 2017 IEEE 33rd International Conference on Data Engineering (ICDE). doi:10.1109/icde.2017.112.

Microsoft, Machine learning operations (mlops) framework to upscale machine learning lifecycle with Azure machine learning, Microsoft Azure, blog, 2023 [Online]. [Online]. Available: https://learn.microsoft.com/en-us/azure/ architecture/example-scenario/mlops/mlops-technical-paper.

N . Polyzotis, S. Roy, S. Whang, and M. Zinkevich, "Data lifecycle challenges in production machine learning: A survey", ACM SIGMOD Record, vol. 47, pp. 17–28, Dec. 2018. doi: 10.1145/3299887.3299891.

O. Doody and M. Noonan, "Preparing and conducting interviews to collect data", en, Nurse Researcher, vol. 20, no. 5, pp. 28–32, May 2013, issn: 13515578, 2047-8992. doi: 10.7748/nr2013.05.20.5.28.e327. [Online]. Available: http://rcnpublishing.com/doi/abs/10.7748/nr2013.05.20.5.28. e327 (visited on 01/26/2023).

R. Ashmore, R. Calinescu, C. Paterson. "Assuring the Machine Learning Lifecycle: Desiderata, Methods, and Challenges". In: (2019). url: http://arxiv.org/abs/1905.04223 (cit. on pp. 22, 24).

Romero, O.; Wrembel, R.; Song, I.Y. An Alternative View on Data Processing Pipelines from the DOLAP 2019 Perspective. J. Inf. Syst. 2020, 92, 101489.

Schelter, S., Biessmann, F., Januschowski, T., Salinas, D., Seufert, S. and Szarvas, G. (n.d.). On Challenges in Machine Learning Model Management. [online] Available at: http://sites.computer.org/debull/A18dec/p5.pdf [Accessed 11 Jan. 2022].

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F. and Dennison, D. (n.d.). Hidden Technical Debt in Machine Learning Systems. [online] Available at: https://proceedings.neurips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf.

Seif, G. (2022). Don't make this big machine learning mistake: research vs application. [online] Medium. Available at: https://towardsdatascience.com/dont-make-this-big-machine-learningmistake-research-vs-application-bd52d5a9a8b9 [Accessed 29 Aug. 2022].

shagunsodhani.com. (n.d.). Searching for Build Debt - Experiences Managing Technical Debt at Google · Papers I Read. [online] Available at: https://shagunsodhani.com/papers-I-read/Searching-for-Build-Debt-Experiences-Managing-Technical-Debt-at-Google [Accessed 29 Aug. 2022].

Tao, C. (2021). Machine Learning in Academic Research v.s. Practical. [online] Medium. Available at: https://towardsdatascience.com/machine-learning-in-academic-research-vs-practical-5e7b3642fc06 [Accessed 29 Aug. 2022].

Weber, C., Hirmer, P., Reimann, P. and Schwarz, H. (2019). A New Process Model for the Comprehensive Management of Machine Learning Models. Proceedings of the 21st International Conference on Enterprise Information Systems. doi:10.5220/0007725304150422.

Yuri Demchenko, "From DevOps to DataOps: Cloud based Software Development and Deployment", in Proceedings of the Proc. The International Conference on High Performance Computing and Simulation (HPCS 2020)", 2020

Zoabi, Y., Deri-Rozov, S. and Shomron, N. (2021). Machine learning-based prediction of COVID-19 diagnosis based on symptoms. npj Digital Medicine, [online] 4(1), pp.1-5. doi:10.1038/s41746-020-00372-6.

Hans Van Vliet, JC Van Vliet, and Hans Van Vliet (2008). Software engineering: principles and practice.

Kraska, T. (2018). Northstar. Proceedings of the VLDB Endowment, 11(12), pp.2150–2164. doi:10.14778/3229863.3240493.

Nakandala, S., Zhang, Y. and Kumar, A. (2020). Cerebro. Proceedings of the VLDB Endowment, 13(12), pp.2159–2173. doi:10.14778/3407790.3407816.

Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Xie, F. and Zumar, C. (n.d.). Accelerating the Machine Learning Lifecycle with MLflow. [online] Available at: https://cs.stanford.edu/~matei/papers/2018/ieee_mlflow.pdf.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D.B., Amde, M., Owen, S., Xin, D., Xin, R., Franklin, M.J., Zadeh, R., Zaharia, M. and Talwalkar, A. (2016). MLlib: Machine Learning in Apache Spark. Journal of Machine Learning Research, [online] 17(34), pp.1–7. Available at: https://jmlr.org/beta/papers/v17/15-237.html [Accessed 17 Sep. 2022].

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A. and Talwalkar, A. (2018). Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. arXiv:1603.06560 [cs, stat]. [online] Available at: https://arxiv.org/abs/1603.06560 [Accessed 16 Mar. 2021].

Modi, A.N., Koo, C.Y., Foo, C.Y., Mewald, C., Baylor, D.M., Breck, E., Cheng, H.-T., Wilkiewicz, J., Koc, L., Lew, L., Zinkevich, M.A., Wicke, M., Ispir, M., Polyzotis, N., Fiedel, N., Haykal, S.E., Whang, S., Roy, S., Ramesh, S. and Jain, V. (2017). TFX: A TensorFlow-Based Production-Scale Machine Learning Platform. [online] Google Research. Available at: https://research.google/pubs/pub46484/ [Accessed 17 Sep. 2022].

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D. and Cox, D.D. (2015). Hyperopt: a Python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, 8(1), p.014008. doi:10.1088/1749-4699/8/1/014008.

Polyzotis, N., Zinkevich, M., Roy, S., Breck, E. and Whang, S. (2019). Data Validation for Machine Learning. *Proceedings of Machine Learning and Systems*, [online] 1, pp.334–347. Available at: https://proceedings.mlsys.org/paper/2019/hash/5878a7ab84fb43402106c575658472fa-Abstract.html [Accessed 9 Apr. 2022].

Jessica Davis, J.D., AppsJune 26, E. and 2019 (2019). Getting Machine Learning into Production: MLOps. [online] InformationWeek. Available at: https://www.informationweek.com/ai-or-machine-learning/getting-machine-learning-into-production-mlops [Accessed 17 Sep. 2022].

Markowitz, D. (n.d.). Google Cloud BrandVoice: Why MLOps Is Critical To The Future Of Your Business. [online] Forbes. Available at: https://www.forbes.com/sites/googlecloud/2021/05/19/why-mlops-is-critical-to-the-future-of-your-business/?sh=6d98c9284537 [Accessed 17 Sep. 2022].

Paleyes, A., Urma, R.-G. and Lawrence, N.D. (2021). Challenges in Deploying Machine Learning: a Survey of Case Studies. arXiv:2011.09926 [cs]. [online] Available at: https://arxiv.org/abs/2011.09926.

APPENDIX A:

MACHINE LEARNING REQUIREMENT FORM

Part of the ML Lifecycle	Roles to ask	Requirement Question	REQ Answers	Examples
Scoping:				
	Business stakeholder	What are the business problems and can they be solved with AI?		Has it been done before, research proves it possible, still unclear
	Business stakeholder	What are the metrics for success?		ROI, customer wishes,
	Business stakeholder	What are the resources needed?		Data, time, people
Data:				

	Business stakeholder, Data scientist, Data engineer	Where does the data come from?		Owned data, crowdsourced, purchase data, purchase labels
	Data scientist, Data engineer	What data format will be used?		Structured, unstructured
	Data scientist, Data engineer	How should the data be preprocessed?		Remove data, remove duplicates
	Data scientist, Data engineer	What is the data labeling standard?		On images: Label each scratch independently on the screen, and label each animal separately in the field
	Data scientist, Data engineer, Business stakeholder	What meta-data should be collected?		Time, system model, factory, device type
Modeling:				

	Business stakeholder, Data scientist	What is the model baseline?		Human-level performance, an earlier system's performance, Dummy model
	Business stakeholder	Is it necessary to audit the model? Who should audit the model? What is the audit focus?		Yes/No. Business stakeholder, Third party, Data scientists. Transparency, Equality, Fairness, and Accountability.
	Business stakeholder, Data scientist, Data engineer	Which potential risks for bias exists?		Gender bias, Brand bias, Ethnicity bias
	Business stakeholder, Data scientist	How is the input data served to the model?		Batch data, Real-time data
	Data scientist, IT Architect	Where should the experimental		Database, Excel document, JSON-file

		data result be stored?		
	Business stakeholder	What are important business goal metrics the ML model should consider?		Business needed classifications performance, different from general ML model performance
	Data scientist	What experimental data should be tracked?		Dataset used, Hyperparameters, Results, Results with metric summary/analysis, Training resources, Training time),
	Data scientist, Software engineer, DevOps engineer, Business stakeholder	What deployment constraints exist?		None, Edge device's hardware capabilities
Deployment:				

	Business stakeholder, MLOps engineer, DevOps Engineer	How should the deployment process be handled?		Canary releases, A/B releases, Shadow releases
	Business stakeholder, MLOps engineer, DevOps engineer, Software engineer	Where should the prediction device be found?		Cloud or edge device
	DevOps engineer, MLOps engineer	Which **software metrics** are important to monitor?		Memory, computing power, latency, throughput, server load
	Data scientist, Data engineer, MLOps engineer, Software engineer	Which **input metrics** are important to monitor?		feature types (INT or String), feature range, Data schema validation
	Data scientist,	Which **output metrics** are		# times users redo the search, avg. prediction accuracy

	Software engineer, DevOps engineer	important to monitor?		
	Business stakeholder, MLOps engineer, Data scientist	How often should the model be retrained on the data gathered from deployment?		Every Monday, once a month, based on deployed input/output metric triggers
	Business stakeholder, DevOps engineer, Data scientist	Are there any specific performance requirements?		Latency requirements, Query per seconds requirements

Table A.1: The first version of the artifact was developed based on the literature review findings from cycle one's problem investigation.

Machine Learning Operations Requirements Form

Form utilization and purpose

Depending on the company's structure, the usage of this form may vary in practice. For example, during the scoping phase of a project, a meeting could be arranged with the relevant parties and roles, where questions are posed and answers are recorded. Alternatively, a designated individual might ask the questions in a one-on-one setting. Nevertheless, the purpose of this form remains the same: to inquire about specific requirements, derived from common challenges, to relevant roles within the organization. The responses may then be documented within this form and shared with relevant parties, such as the implementation team or business stakeholders.

Column descriptions

Part of ML Lifecycle: Displays the stages of a machine learning life cycle

Roles to Ask: Indicates which role should be asked a specific requirements question

Requirement Questions: Specifies a requirements question to ask

Requirement Answers: Blank field to record the answers to a requirements question

Examples: Provides common answers to a requirements question

Figure A.1: Front page of the second version artifact created. Supposed to serve as an introduction to the MLOps Requirements Form (the artifact).

Part of the ML Lifecycle	**Roles to ask**	**Requirement Question**	**Requirement Question Answer**	**Examples**
Scoping:				
	Business stakeholder	What are the business problems?		Battery optimization, Fraud detection, Demand forecasting
	Data Scientist	Can the business problems be solved with ML, and how?		Has it been done before, research proves it possible, still unclear
	Product owner	What are the metrics for success?		ROI, customer wishes
	Product owner	What are the resources needed?		Data, time, people
	Business stakeholder	Who is the end user?		Demographical information, Internal company users, Customers
	Business stakeholder	How will the users interact with the model, what interface will they need?		App, Voice-activated feature, Web page, API
	Business stakeholder, Product Owner, Data scientist, Data engineer	Who is the domain expert and can we access them?		Doctors, Lawyers, Domain-specific researcher
Data:				

	Business stakeholder, Product Owner, Data scientist, Data engineer	Where does the data come from?		Owned data, crowdsourced, purchase data, purchase labels
	Data scientist, Data engineer	What data format will be used?		Structured, unstructured
	Data scientist, Data engineer	How should the data be preprocessed?		Remove data, remove duplicates
	Data scientist, Data engineer, Domain expert	What is the data labeling standard?		On images: Label each scratch independently on the screen, and label each animal separately in the field
	Data scientist, Data engineer, Product owner	What meta-data should be collected?		Time, system model, factory, device type
	Data Engineer, Legal team, Business stakeholder, Product owner	Are there any privacy concerns regarding the data?		Names, Emails, Addresses, Phone numbers, and general GDPR concerns
	Data Engineer, Legal team, Business stakeholder, Product owner	Are there any necessary data ownership considerations?		Data is owned by us, it's open source, and another party owns all data
	Product owner, Data scientist, Data engineer	How much data is expected to be stored?		~10TB
	Product owner, Data	When does the data become irrelevant?		Never, new product versions

	scientist, Data engineer, Domain expert			are released, annually
	Data engineer, Domain expert	Are there any cyclic behaviors in the data?		Seasonal sales cycle, full-day cycle
	Data scientist	What is the minimum amount of data that is necessary to train the model?		10k images, 100 GB worth of 1080p mp3 video recordings
	Product owner, Data scientist, Data Engineer	How will the data be acquired?		Automated tool, manually collected, purchased

Table A.2: Part one of artifact version two.

see Table A.2 for the second part. This is the form without the front page, the front page can be found in the Appendix. This artifact is the result of cycle two's problem investigation and the evaluation from cycle one.

Modeling:				
	Product owner, Data scientist	What is the model baseline?		Human-level performance, A previous system's performance, Dummy model
	Product owner, Legal	Is it necessary to audit the model? Who should audit the model? What is the audit focus?		Yes/No. Business stakeholder, Third party, Data scientists. Transparency, Equality, Fairness, and Accountability...

	Data scientist, Data engineer	Which potential risks for bias exists?		Gender bias, Brand bias, Ethnicity bias
	Product owner, Data scientist	How is the input data served to the model?		Batch data, Real-time data
	Data scientist, IT Architect	Where should the experimental data result be stored?		Database, Excel document, JSON-file
	Product owner	What are important business goal metrics the ML model should consider?		Business required classifications performance, different from general ML model performance
	Data scientist	What experimental data should be tracked?		Dataset used, Hyperparameters, Results, Results with metric summary/analysis, Training resources, Training time),
	Data scientist, Software engineer,	What deployment constraints exist?		None, Edge device's hardware capabilities

	DevOps engineer, MLOps engineer			
Deployment:				
	Product owner, MLOps engineer, DevOps engineer	How should the deployment process be handled?		Canary releases, A/B releases, Shadow releases
	Product owner, MLOps engineer, DevOps engineer	Where should the prediction device be located?		Cloud or edge device
	DevOps engineer, MLOps engineer, Software engineer	Which software metrics are important to monitor?		Memory, computing power, latency, throughput, server load

	Data scientist, Data engineer, MLOps engineer	Which input metrics are important to monitor?		feature types (INT or String), feature range, Data schema validation
	Data scientist, Software engineer, MLOps engineer	Which output metrics are important to monitor?		# times users redo the search, avg. prediction accuracy
	Product owner, MLOps engineer, Data scientist	How often should the model be retrained on the data gathered from deployment?		Every Monday, once a month, based on deployed input/output metric triggers
	Product owner, DevOps engineer, Data scientist	Are there any specific performance requirements?		Latency requirements, Query per seconds requirements

Table A.3: Part two of artifact version two.

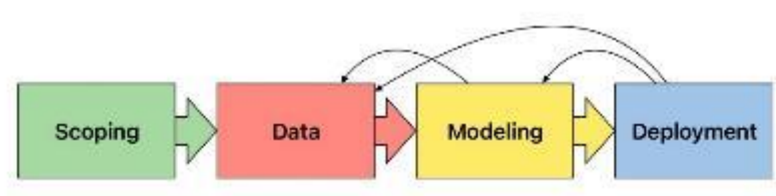
see Table A.3 for the first part. This is the form without the front page, the front page can be found in Appendix A. This artifact is the result of cycle two's problem investigation and the evaluation from cycle one.

Form utilization and purpose

Depending on the company's structure, the usage of this form may vary in practice. For example, during the scoping phase of a project, a meeting could be arranged with the relevant parties and roles, where questions are posed, and answers are recorded. Alternatively, a designated individual might ask the questions in a one-on-one setting. Nevertheless, the purpose of this form remains the same: to inquire about specific requirements, derived from common challenges and best practices to relevant roles within the organization. The responses may then be documented within this form and shared with relevant parties, such as the implementation team or business stakeholders. The documented answers can be interpreted as informal MLOps requirements. Therefore, they can be used as they are or as the foundation for creating more formal requirements depending on the implementation context.

The MLOps stages

MLOps is an iterative framework that requires constant maintenance and monitoring. Therefore, it is common for the MLOps requirements to evolve and change iteratively in parallel. The figure below gives a visual representation of how the changes and



information from one stage feed into another.

Part of ML Lifecycle: Displays the stages of a machine learning life cycle.

Roles to Ask: Indicates which role should be asked a specific requirement question.

Requirement Questions: Specifies requirements question to ask.

Requirement Answers: Blank field to record the answers to a requirement question.

Examples: Provides common answers to a requirements question

Figure A.2: Front page of the final artifact created. Supposed to serve as an introduction to the MLOps Requirements Form (the artifact).

Part of the ML Lifecycle	Roles to ask	Requirement Question	Requirement Question Answer	Examples
Scoping:				
	Business stakeholder	What specific challenges is the business facing?		Battery optimization, Fraud detection, Demand forecasting
	Data Scientist	Is machine learning a viable solution for addressing these business problems, and if so, how?		Has it been done before, research proves it possible, still unclear
	Product owner	What metrics will be used to measure the success of the solution?		ROI, customer wishes
	Product owner	What resources are required to implement the proposed solution?		Data, time, people

	Product owner, Business stakeholder, Data scientist	What is the budgetary constraint for the computation required to train the model?		If on-premise: 100h allowed, 50h, Unlimited If on cloud: Budget is $1,000, $5,000, $500
	Business stakeholder	Who constitutes the end user in this context?		Demographical information, Internal company users, Customers
	Business stakeholder	How will users interact with the model, and what interface is necessary for their interaction?		App, Voice-activated feature, Web page, API
	Business stakeholder, Product Owner, Data scientist, Data engineer	Who serves as the domain expert, and is there accessibility to them for consultation?		Doctors, Lawyers, Domain-specific researcher

Table A.4: Part one of the final artifact, includes requirement questions regarding the scoping stage of an ML system.

Part of the ML Lifecycle	Roles to ask	Requirement Question	Requirement Question Answer	Examples
Data:				
	Business stakeholder, Product Owner, Data scientist, Data engineer	Where is the data sourced from?		Owned data, crowdsourced, purchase data, purchase labels
	Data scientist, Data engineer	What format is designated for the data?		Structured, unstructured
	Data scientist, Data engineer	What preprocessing steps are required for the data?		Remove data, remove duplicates
	Data scientist, Data engineer, Domain expert	What guidelines exist for labeling the data?		On images: Label each scratch independently on the screen, label each animal separately in the field
	Product owner,	By whom will the data be labeled?		In-house resources, Crowdsourced,

	Business stakeholder			Outsourced, Mixture of resources
	Data scientist, Data engineer, Product owner	What meta-data needs to be gathered alongside the data?		Time, system model, factory, device type
	Data Engineer, Legal team, Business stakeholder, Product owner	Are there privacy considerations related to the data?		Names, Emails, Addresses, Phone numbers, general GDPR concerns
	Data Engineer, Legal team, Business stakeholder, Product owner	Are there specific ownership considerations for the data?		Data is owned by us, it's open source, and another party owns all data
	Product owner, Data scientist, Data engineer	What is the anticipated volume of stored data?		~10TB

	Product owner, Data scientist, Data engineer, Domain expert	When does the data reach a point of irrelevance?		Never, new product versions are released, annually
	Data engineer, Domain expert	Are there recurring patterns or cycles in the data?		Seasonal sales cycle, full-day cycle
	Data scientist	What is the minimum data quantity required for model training?		10k images, 100 GB worth of 1080p mp3 video recordings
	Data scientist	What is the minimal data point frequency required for streaming data to align with the business goals?		Every 5ms, Every 1s, Every data point
	Product owner, Data	How is the data acquisition		Automated tool, manually collected, purchased

	scientist, Data Engineer	process structured?		

Table A.5: Part two of the final artifact, includes requirement questions regarding the data stage of an ML system.

Modeling:				
	Product owner, Data scientist	What constitutes the baseline for the model?		Human-level performance, A previous system's performance, Dummy model
	Product owner, Legal	Is model auditing required, and if so, who is responsible for conducting it, and what is the focal point of the audit?		Yes/No. Business stakeholder, Third party, Data scientists. Transparency, Equality, Fairness, and Accountability...
	Data scientist, Data engineer	What potential biases should be acknowledged and addressed in the model?		Gender bias, Brand bias, Ethnicity bias

	Product owner, Data scientist	How is the input data presented to the model?		Batch data, Real-time data
	Data scientist, IT Architect	Where is the appropriate storage location for the outcomes of experimental data?		Database, Excel document, JSON-file
	Product owner	What key business metrics should the ML model prioritize?		Business required classifications performance, different from general ML model performance
	Data scientist	Which experimental data points should be monitored?		Dataset used, Hyperparameters, Results, Results with metric summary/analysis, Training resources, Training time),
	Data scientist, Software engineer, DevOps	What constraints are in place for model deployment?		None, Edge device's hardware capabilities

	engineer, MLOps engineer			

Table A.6: Part three of the final artifact, includes requirement questions regarding the modeling stage of an ML system

Part of the ML Lifecycle	**Roles to ask**	**Requirement Question**	**Requirement Question Answer**	**Examples**
Deployment:				
	Product owner, MLOps engineer, DevOps engineer	What is the recommended approach for managing the deployment process?		Canary releases, A/B releases, Shadow releases
	Product owner, MLOps engineer, DevOps engineer	Where is the designated location for the prediction device?		Cloud or edge device
	DevOps engineer, MLOps engineer,	Which software metrics should be closely monitored?		Memory, computing power, latency, throughput, server load

	Software engineer			
	Data scientist, Data engineer, MLOps engineer	What input metrics are considered crucial for monitoring?		feature types (INT or String), feature range, Data schema validation
	Data scientist, Software engineer, MLOps engineer	What output metrics are essential for ongoing monitoring?		# times users redo the search, avg. prediction accuracy
	Product owner, MLOps engineer, Data scientist	At what frequency should the model undergo retraining using the data collected during deployment?		Every Monday, once a month, based on deployed input/output metric triggers
	Product owner, DevOps engineer, Data scientist	Are there any explicit performance requirements that need to be met?		Latency requirements, Query per seconds requirements

Table A.7: Part four of the final artifact, includes requirement questions regarding the deployment stage of an ML system.

APPENDIX: B

INTERVIEW SCRIPT

1st set of semi-structured interviews.

Introduction and explanation of the study (Max 5 minutes)

Demographics and Professional Experiences (Max 10 minutes)

1. Can you shortly introduce yourself and tell me which team you work with?
2. How long have you worked in the industry?
3. What is your current role in the company?
4. What does this role entail? (What usually is your part of projects?) 5. Does your education match your current role and what is your education?

Open interview (Max 25 minutes)

General Questions:

1. Have you worked with requirements for software projects? Describe your experiences briefly.
2. Have you worked with ML projects?
    a. Was the project deemed successful?
        i. If the project was successful, how did you know that the project was successful?
        ii. If it was not successful, what indicated its failure?
3. Have you worked with DevOps on projects? Describe your experiences briefly.
4. Have you worked with or thought about using DevOps for ML projects (MLOps), what are or would be the challenges?
5. Have you worked with or thought of specifying requirements for ML projects? What were or could be the challenges?
    a. If worked with: what worked/works well?
6. Would the MLOps process need further requirements than typically available for ML projects?

Last question:

1. Is there anything you would like to add? Any factors you think we have missed, or something else you want to add to the interview?

Evaluation of prototype artifact from literature (Max 20 min)

1. The following are MLOps requirement questions elicited from the literature. Do you consider them helpful, please motivate why or why not?
2. Can you spot anything that is missing, wrong, or redundant?
3. Do you believe that you and your colleagues would benefit from using this form/guide when scoping for a new MLOps process/infrastructure?
    a. If that is the case, in which way?
4. Do you have any suggestions for improvements to the artifact? These could be related to the structure, UX, content, or anything else.

Figure B.1: The script used for the first set of semi-structured interviews.

Script for the extra interview

1. Are there any common, general, concerns when collecting and working with data for ML models?
2. If you have experience working with data that is stored on the cloud, how does your workflow with that data look like?
3. Which are some general questions you ask yourself when setting up a new data pipeline?
4. How do you handle data pre-processing, what are key things to consider?
    a. Is the process automated in your pipeline or is it done manually?

Figure B.2: The script used for the extra interview held during iteration 2 which focused on the participants' team's current data pipelines and workflows.

APPENDIX: C

PARTICIPANTS REQUIREMENTS QUESTIONS ANSWERS

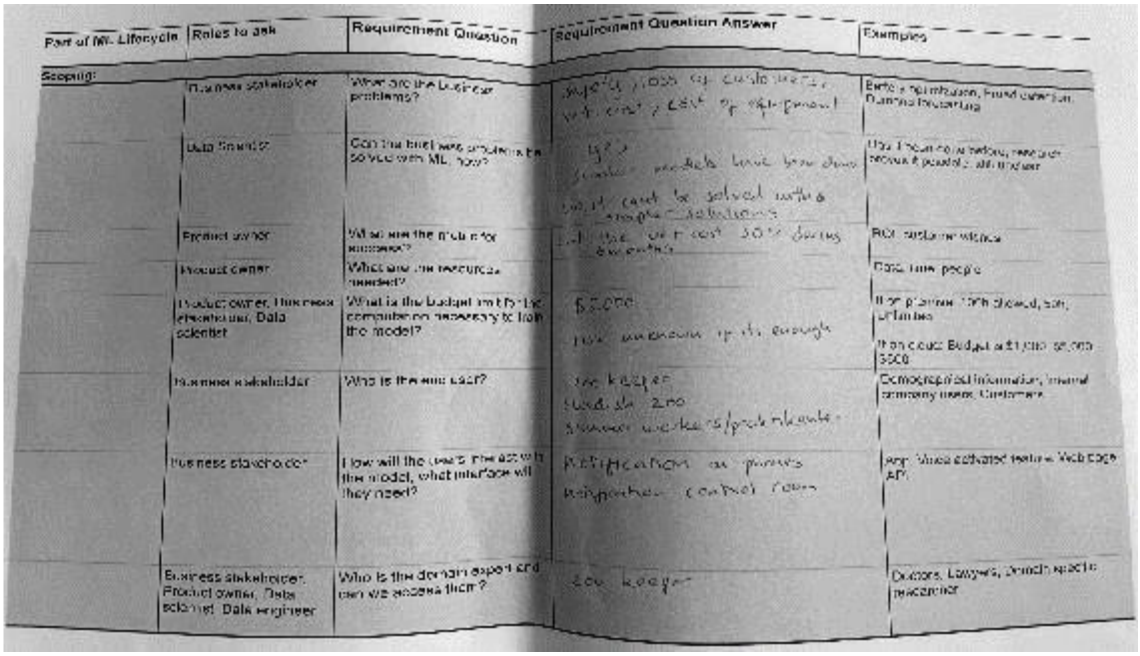


Figure C.1: Participants Requirements Questions Answers regarding the scoping stage.

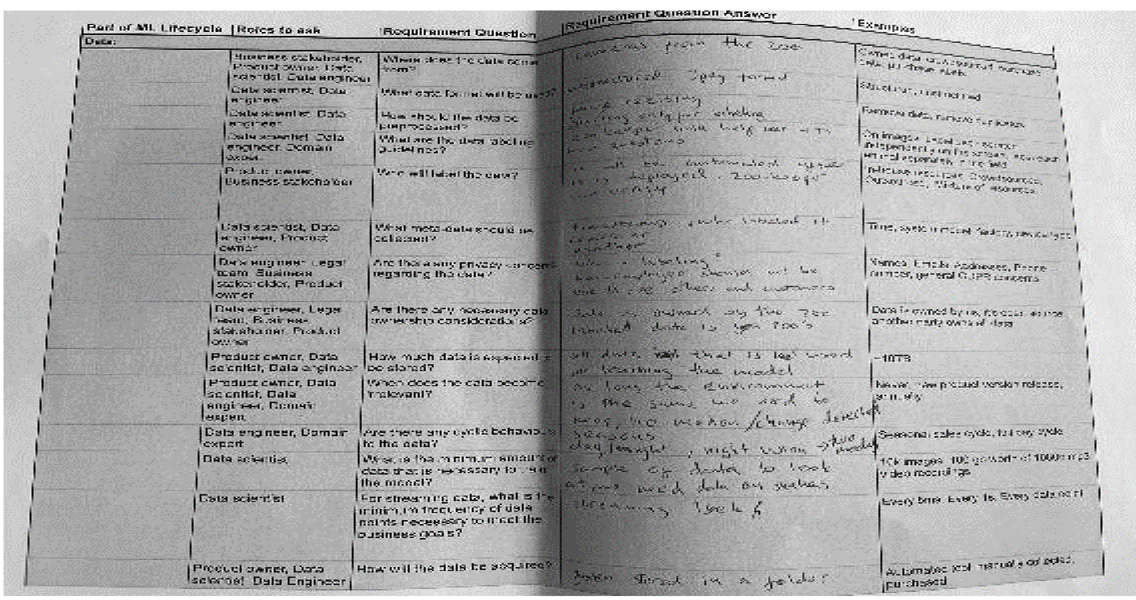
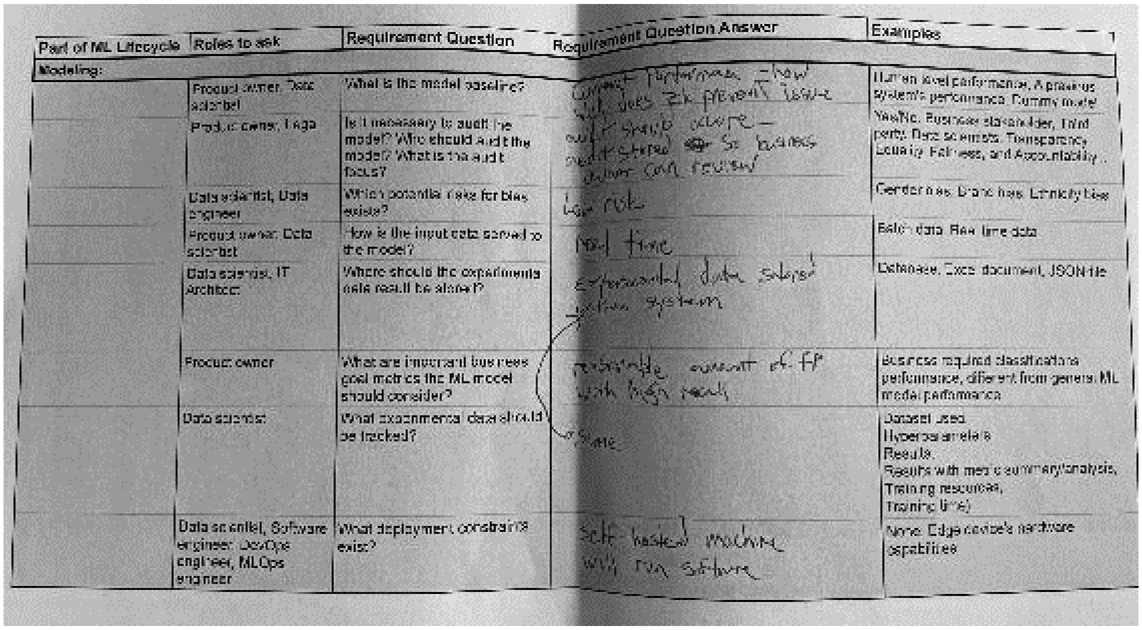
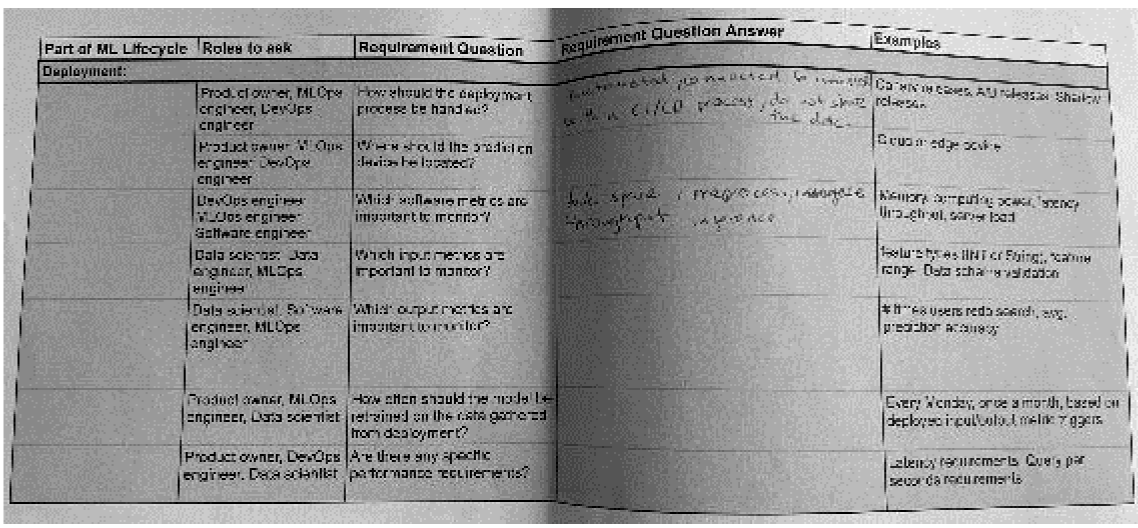


Figure C.2: Participants Requirements Questions Answers regarding the data stage.



Part of ML Lifecycle	Roles to ask	Requirement Question	Requirement Question Answer	Examples
Modeling:				
	Product owner, Data scientist	What is the model baseline?	Current performance – how well does ZK prevent issue	Human level performance, A previous system's performance, Dummy model
	Product owner, Legal	Is it necessary to audit the model? Who should audit the model? What is the audit focus?	audit should occur – audit should [illegible] so business owner can review	Yes/No. Business stakeholder, third party, Data scientists. Transparency, Equality, Fairness, and Accountability.
	Data scientist, Data engineer	Which potential risks for bias exists?	Low risk	Gender bias, Brand bias, Ethnicity bias
	Product owner, Data scientist	How is the input data served to the model?	Real time	Batch data, Real time data
	Data scientist, IT Architect	Where should the experiments data result be stored?	experimental data stored within system	Database, Excel document, JSON file
	Product owner	What are important business goal metrics the ML model should consider?	reasonable amount of FP with high recall	Business required classifications performance, different from general ML model performance
	Data scientist	What experimental data should be tracked?	Same	Dataset used, Hyperparameters, Results, Results with metric summary/analysis, Training resources, Training time)
	Data scientist, Software engineer, DevOps engineer, MLOps engineer	What deployment constraints exist?	Self-hosted machine will run software	None, Edge device's hardware capabilities

Figure C.3: Participants Requirements Questions Answers regarding the modeling stage.



Part of ML Lifecycle	Roles to ask	Requirement Question	Requirement Question Answer	Examples
Deployment:				
	Product owner, MLOps engineer, DevOps engineer	How should the deployment process be handled?	[illegible] connected [illegible] with CI/CD process, do not share the data	Canary releases, A/B releases, Shadow release.
	Product owner, MLOps engineer, DevOps engineer	Where should the prediction device be located?		Cloud or edge device
	DevOps engineer, MLOps engineer, Software engineer	Which software metrics are important to monitor?	[illegible] speed / [illegible] throughput [illegible]	Memory, computing power, latency, throughput, server load
	Data scientist, Data engineer, MLOps engineer	Which input metrics are important to monitor?		Feature types (INT or String), feature range, Data schema validation
	Data scientist, Software engineer, MLOps engineer	Which output metrics are important to monitor?		# times users redo search, avg. prediction accuracy
	Product owner, MLOps engineer, Data scientist	How often should the model be retrained on the data gathered from deployment?		Every Monday, once a month, based on deployed input/output metric triggers
	Product owner, DevOps engineer, Data scientist	Are there any specific performance requirements?		Latency requirements, Query per seconds requirements

Figure C.4: Participants Requirements Questions Answers regarding the deployment stage.

APPENDIX: D

CODEBOOK

Table D.1: Codebook displaying the collection of inductive and deductive codes and their description used during the thematic analysis.

Codes:	**Description:**
Access to a domain expert	The ability of data scientists to consult with subject matter experts who have deep knowledge of the domain or industry in which the machine learning model is being applied.
Automation	The use of software tools and algorithms to automate the process of building, training, and deploying models.
Avoid development requirements	Requirements specifying how development should be done are advised
Business goals	The specific objectives that a company or organization is trying to achieve through the use of machine learning technology.
CI/CD	Set of best practices and tools for automating the development, testing, and deployment of models.
Customer feedback loop	Process of continuously gathering feedback from users or customers of a machine learning product or service, and using that feedback to improve the performance and usability of the product or service.

Data ingestion cycle	Collection of a complete data cycle
Data leakage	Information from the training dataset is inadvertently included in the test dataset or otherwise used to inform model development.
Data Ownership	The legal and ethical ownership and control of the data used in models. This includes considerations such as who owns the data, who has the right to access and use the data, and how the data can be used.
Data privacy	The protection of sensitive and personal data used in ML models.
Data quality	The accuracy, completeness, and consistency of the data used to train and test ML models.
Dataset size	The amount of data necessary to train and test a machine learning model.
Deployment requirements	The specific needs and constraints for deploying a machine learning model in a real-world production environment.
Development environment	The set of tools, software, and hardware used to develop, test, and refine machine learning models
Difficult to evaluate if requirements are met	Difficult to pinpoint when a requirement in an ML environment is accomplished
Documenting	The process of creating and maintaining comprehensive and accurate documentation for an ML project.

Dynamic environment	The environment in which the data or conditions may change over time.
Error analysis	The process of analyzing the errors made by a model to identify patterns or trends that can be used to improve its performance.
Experimentation logs	A systematic and comprehensive record of experiments that have been conducted during the development and optimization of machine learning models.
Given requirement	Requirements that are elicited by another person and then given for implementation
Hard to implement infrastructure	Difficult to implement MLOps infrastructure
Importance of Data	Recognition of the impact data has on the success of an ML model
Infrastructure migration	The process of transferring ML models and associated workflows from one infrastructure environment to another.
Infrastructure requirements	Configuration decisions regarding the infrastructure
Ingestion sources	The various types of data sources that can be used to feed data into ML models
Inter-team communications,	Communication within a single team or between multiple different teams

Labeling	The process of assigning a categorical or numerical value to a data point or sample.
Manage expectations	The process of setting realistic goals and outcomes for an ML project, communicating those goals to stakeholders, and regularly evaluating and adjusting those expectations as the project progresses.
Maturity	The level of sophistication and effectiveness of organizations' ML operations processes and practices.
Model baseline	A simple, minimal, or trivial model that is used as a benchmark to evaluate the performance of more complex models.
Model requirements	The specifications and expectations that a model must meet to be considered successful and useful for its intended purpose.
Model retraining	The process of updating or refining a model using new data or updated parameters.
Monitoring	Continuously observing the performance of a model over time to ensure that it is still accurate and relevant for its intended task.
Non-functional requirements	Characteristics and qualities of a system that are not related to its primary function or task, but rather to its overall performance, scalability, reliability, and maintainability.

Production model	The process of deploying a machine learning model into a production environment, where it can be used to make predictions on new data in real time.
Self-made requirement	Requirements that are elicited by the implementation team
Testing	The process of evaluating ML models to ensure that they are working as intended and producing accurate results.
Understand ML	Understanding of how machine learning works
Understand the data	Understanding the data needed to train a model
Understand the end-user	Understanding the end-users of a model
Understand the problem	Understanding the problem being solved with ML
Value of MLOps	The value MLOps brings to an organization
Versioning	The practice of tracking and managing changes to ML models and associated artifacts over time. This includes tracking changes to the model code, data sets, model configurations, and other related resources.
Work continuously with requirements	A dynamic system requires continuous work on the requirements

APPENDIX: E

PARTICIPANT DATA

Interviewee ID	**Current Role**	**Company**	**Experience (Years)**
1st set semi-structured interviews			
ID1.1	Data Analyst	Company 1	8
ID2	Software Engineer	Company 1	3*
ID3	Software Engineer	Company 3	2
ID4	Software Engineer	Company 2	1
ID5	Software Developer	Company 2	2*
ID6.1	Product Owner	Company 3	15
ID7.1	Sr Manager, ML Engineering and Research	Company 2	20*
ID8	Software Engineering Manager	Company 1	8
ID9.1	Sr. Data scientist	Company 3	8
ID10	ML Researcher	Company 3	3*
ID11	Software Engineer	Company 1	2
ID12	Software Engineer	Company 1	1
ID13	Software Developer	Company 1	2*
ID14.1	Product Owner	Company 1	15
ID15.1	Data Analyst	Company 2	2
ID16.1	Software Engineering Manager	Company 2	10
ID17.1	Data Analyst	Company 2	6

ID18	Software Engineer	Company 3	2*
ID19.1	Software Engineer	Company 3	4
ID20.1	Software Engineer	Company 3	2
2nd set semi-structured interviews			
ID1.2	Data Analyst	Company 1	8
ID2	Software Engineer	Company 1	3*
ID3	Software Engineer	Company 3	2
ID4	Software Engineer	Company 2	1
ID6.2	Product Owner	Company 1	15
ID7.2	Sr. Manager, ML Engineering and Research	Company 2	20*
ID9	Sr. Data scientist	Company 3	15
ID10	ML Researcher	Company 2	3
ID11	Software Engineer	Company 1	2
ID13	Software Developer	Company 1	2*
ID14.1	Product Owner	Company 1	15
ID15.1	Data Analyst	Company 2	2
ID18	Software Engineer	Company 3	2*
ID19.1	Software Engineer	Company 3	4
ID20.1	Software Engineer	Company 3	2
3rd set semi-structured interviews			
ID1.2	Data Analyst	Company 1	8
ID2	Software Engineer	Company 1	3*
ID3	Software Engineer	Company 3	2

ID4	Software Engineer	Company 2	1
ID6.2	Product Owner	Company 1	15
ID7.2	Sr. Manager, ML Engineering and Research	Company 2	20*
ID9	Sr. Data scientist	Company 3	15
ID10	ML Researcher	Company 2	3
ID15.1	Data Analyst	Company 2	2
ID18	Software Engineer	Company 3	2*

Table E.1: Interviewee participant traceability matrix

The trace of the interviewee is available on their identification (the first number represents ID and the second number indicates repeated interviews). In addition, details related to current position as well as organization are given below. anonymization is made by replacing names with numbers and company names using figures. *Minimum possible experience that was used in cases, where it was not clear.

APPENDIX F:

INFORMATION SHEET

Title	*WHAT ARE THE KEY AREAS OF ML-OPS / DL-OPS IN BUSINESS PROBLEMS FOR COMPANY GROWTH USING CLOUD ENVIRONMENT?*
Coordinating Principal Investigator/	*Gokul Talele*
Location	*India*

Part 1 What does my participation involve?

1 Introduction

You are invited to take part in this research project, which is **what are the key areas of ml-ops / dl-ops in business problems for company growth using cloud environment.** You have been invited because you had indicated in the survey that you represent a profitable company in the data driven business in India. Your contact details were obtained from the Google survey, or from LinkedIn.

This Participant Information Sheet/Consent Form tells you about the research project. It explains the processes involved with taking part. Knowing what is involved will help you decide if you want to take part in the research.

Please read this information carefully. Ask questions about anything that you don't understand or want to know more about. Before deciding whether to take part, you might want to talk about it with a relative, friend or local health worker.

Participation in this research is voluntary. If you don't wish to take part, you don't have to.

If you decide you want to take part in the research project, you will be asked to sign the consent section. By signing it you are telling us that you:

- Understand what you have read.
- Consent to take part in the research project.
- Consent to be involved in the research described.
- Consent to the use of your personal and health information as described.

You will be given a copy of this Participant Information and Consent Form to keep.

2 What is the purpose of this research?

This research is being conducted to gain more knowledge that impedes machine learning projects in cloud environment for company growth.

The aim of the research is to create a framework for MLOps to follow to achieve what the company you represent has.

The results of this research will be published, and will be used the researcher, Gokul Talele, to obtain a Doctorate in Business Administration degree.

3 What does participation in this research involve?

If you decide to take part in the research project, you will first be given consent form to sign, and a questionnaire asking about yourself and the company you represent; this will determine if you are eligible to take part. Completing the questionnaire will take approximately 10 – 15 minutes.

If the screening questionnaire shows that you meet the requirements, then you will be able to start the research project. If the screening questionnaire shows that you cannot be in the research project, the researcher will discuss other options with you.

The interview with the researcher will be conducted over a video conferencing software like Microsoft teams. The interview will last between 45 to 60 minutes at a time and date of your choosing. The interview will be conducted in English. If you are not comfortable speaking English, the researcher will arrange for a translator.
The interview will be recorded by the researcher with your consent and will be stored securely.
At the end of the interview, the researcher may request for additional time to ask any follow up questions or cover any unanswered questions.

During the interview, the researcher will ask you a series of questions about yourself, the business you represent and ask you to share your thoughts.
The questions posed will be open ended, with no right or wrong answers.

This research project has been designed to make sure the researcher interprets the results in a fair and appropriate way and avoids study doctors or participants jumping to conclusions.

There are no costs associated with participating in this research project, nor will you be paid.

4 Other relevant information about the research project

During the course of this research, the researcher will be speaking to around 10 people like yourself.

Each interview will be conducted separately, and the interview, their details, and results will be kept completely confidential.

This research has Gokul Talele as the primary researcher and no assistant researchers.

5 Do I have to take part in this research project?

Participation in any research project is voluntary. If you do not wish to take part, you do not have to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage.

If you do decide to take part, you will be given this Participant Information and Consent Form to sign and you will be given a copy to keep.

6 What are the possible benefits of taking part?

The researcher cannot guarantee or promise that you will receive any benefits from this research; however, after the research is published, you will have access to the paper, and may be able to derive additional insights that may help improve your business.

7 What are the possible risks and disadvantages of taking part?

You may feel that your interview and answers may be accessed by third parties; The researcher will store the interview on a secured local drive, and a backup copy of the interview on a cloud drive with two-factor authentication enabled. The interview recording can only be accessed by the researcher, their supervisor or yourself.

You may be averse to sharing information about the company you represent; The researcher will not discuss or reveal any information about you or your participation in this research (save for naming the company you represent) to other participants. Any data shared during the interview process will be aggregated in the research and all identities will be anonymized.

You may feel that some questions are stressful or upsetting; If you do not wish to answer a question, you may skip it and move to the next, or stop the interview immediately.

8 What if I withdraw from this research project?

If you do consent to participate, you may withdraw at any time. If you decide to withdraw from the project, please notify the researcher before you withdraw. The researcher will inform you if there are any special requirements linked to withdrawing. If you do withdraw, you will be asked to complete and sign a '**Withdrawal of Consent**' form; this will be provided to you by the researcher.

If you decide to leave the research project, the researcher will not collect additional personal information from you, although personal information already collected will be retained to ensure that the results of the research project can be measured properly and to comply with law. You should be aware that data collected up to the time you withdraw will form part of the research project results. If you do not want your data to be included, you must tell the researcher when you withdraw from the research project.

9 Could this research project be stopped unexpectedly?

The risk of this research project stopping is very low. However, some reasons may include:

- A lack of participants
- The researcher concludes that the work is unnecessary or invalid
- The research supervisor deems that the work is unnecessary or invalid
- Unforeseen circumstances

10 What happens when the research project ends?

After the research is concluded, the researcher will contact you via your preferred mode of communication and share a summary of the result.

You will also be given a chance to ask any follow up questions or request a copy of the dissertation from the researcher.

The research is scheduled to conclude around March 2023, and the researcher will complete the dissertation around January 2024.

Part 2 How is the research project being conducted?

11 What will happen to information about me?

By signing the consent form, you consent to the researcher collecting and using personal information about you for the research project. Any information obtained in connection with this research project that can identify you will remain confidential. The data collected is for the research in question ONLY and will not be shared or used in any future or parallel research. Your information will only be used for the purpose of this research project and it will only be disclosed with your permission, except as required by law.

All personally identifiable data such as your name, title, contact information etc. that is shared with the researcher will be stored securely on a local drive and a backup copy of the same will be stored on a cloud drive with two factor authentication. The only people with access to the cloud drive will be the researcher and the research supervisor.

All information shared during the interview will be anonymized (if identifiable data) or presented as aggregates or ranges (if figures) in the dissertation.

It is anticipated that the results of this research project will be published and/or presented in a variety of forums. In any publication and/or presentation, information will be provided in such a way that you cannot be identified, except with your express permission. Your confidentiality will be maintained by anonymizing your identity and the company you represent.

In accordance with the privacy laws of the EU and other relevant laws, you have the right to request access to the information about you that is collected and stored by the researcher. You also have the right to request that any information with which you disagree be corrected. Please inform the researcher named at the end of this document if you would like to access your information. The researcher will then provide a copy of the interview transcript for your perusal.

All data will be stored for a period of 1 (ONE) year from the date of the interview, or the date you wish to withdraw from the research, whichever is earlier.

At the end of the study, and the publication of the dissertation, the researcher will purge all local copies and securely erase the storage drive. The cloud backup will be permanently destroyed with no way to recover the data.

12 Complaints and compensation

If you have any concerns or complaints about the research or interview process, you may contact the research supervisor or the institute directly. This information is provided in the subsequent section

13 Who is organizing and funding the research?

This research is being self-funded by the researcher as part of the requirement toward a Doctorate in Business Administration. There are no financial benefits applicable for any parties involved in the research.

14 Who has reviewed the research project?

The ethical aspects of this research project have been approved by the Supervisor/Mentor of *SSBM Geneva*.
This statement has been developed to protect the interests of people who agree to participate in human research studies.

15 Further information and who to contact

The person you may need to contact will depend on the nature of your query. If you want any further information concerning this project or if you have any problems which may be related to your involvement in the project, you can contact the researcher on **+xx-xxx-xxx-xxx** or any of the following people:

Research contact person

Name	Gokul Talele
Position	Primary Researcher
Telephone	
Email	

For matters relating to research at the site at which you are participating, the details of the local site complaints person are:

Complaints contact person

Name	
Position	
Telephone	
Email	

If you have any complaints about any aspect of the project, the way it is being conducted or any questions about being a research participant in general, then you may contact:

Reviewing Supervisor/Mentor name	
HREC Executive Officer	
Email	

Reviewing HREC approving this research and HREC Executive Officer details

APPENDIX G:

INTERVIEW CONSENT FORM

Research Participant name:

The interview will take between 45 – 60 minutes. We don't anticipate any risks associated with your participation, but you can stop the discussion or withdraw from the research anytime.

Thank you for agreeing to be interviewed for the above research project. Ethical procedures for academic research require that interviewees explicitly agree to be interviewed and how the information in their interview will be used. This consent form is necessary for us to ensure that you understand the purpose of your involvement and agree to the conditions of your participation. Would you, therefore, read the accompanying **information sheet** and then sign this form to certify that you approve the following:

1. The interview will be recorded, and a transcript will be produced.
2. You will be sent the transcript and allowed to correct any factual errors.
3. As a research investigator, Gokul Talele will analyze the interview transcript.
4. Access to the interview transcript will be limited to Gokul Talele and academic colleagues and researchers with whom he might collaborate as part of the research process.
5. Any summary interview content or direct quotations from the interview made available through academic publications or other academic outlets will be anonymized so that you cannot be identified, and care will be taken to ensure that additional information in the interview that could identify you is not revealed.
6. The actual recording will be kept.
7. Any variation of the conditions above will only occur with your explicit approval.

Optional consent for direct quotation

If you wish to give explicit consent to the researcher to allow them to quote you directly, please initial it next to any of the statements below. If all comments below are unchecked, clause (5) from the previous section will apply.

	I wish to review the notes, transcripts, or other data collected during the research pertaining to my participation.
	I agree to be quoted directly.
	I agree to be quoted directly if my name is not published and a made-up name (pseudonym) is used.
	I agree that the researchers may publish documents that contain quotations by me.

All or part of the content of your interview may be used: -

1. In academic papers, policy papers, or news articles
2. On our website and in other media that we may produce, such as spoken presentations.
3. On other feedback events
4. In an archive of the project, as noted above.

I agree with signing this form: -

1. I am voluntarily participating in this project. I understand that I don't have to, and I can stop the interview at any time.
2. I have read the accompanying **information sheet**.
3. The transcribed interview or extracts from it may be used as described above.
4. I don't expect to receive any benefit or payment for my participation.
5. I can request a copy of my interview transcript and may make any necessary edits to ensure the effectiveness of any confidentiality agreement.
6. I have been able to ask any questions, and I am free to contact the researcher with any questions I may have in the future.

APPENDIX H:

FORM FOR WITHDRAWAL OF PARTICIPATION

Title	*WHAT ARE THE KEY AREAS OF ML-OPS / DL-OPS IN BUSINESS PROBLEMS FOR COMPANY GROWTH USING CLOUD ENVIRONMENT?*
Coordinating Principal Investigator/	*Gokul Talele*
Location	*India*

Declaration by Participant

I want to withdraw from the above research project. Doing so will not impact my regular medical care, my interactions with the researchers, or my affiliation with the Swiss School of Business and Management.

	______________________________
Signature ______________________	Date ______________________

If the participant vocally communicates their intention to withdraw, the Senior Researcher is required to describe the following circumstances.

Declaration by Researcher†

I think the participant has received the verbal explanation I provided about the consequences of withdrawing from the research project.

Name of Researcher (please print)	______________________		
Signature	______________________	Date	______________________

† If a research team member wants to leave from the project, they must inform the team member with the necessary qualifications.

Note: Each signer must date their own signature in the consent area.

APPENDIX I:

ETHICAL REVIEW APPLICATION FORM

Section 1: Candidate Details	
First Name	Gokul
Last Name	Talele
Faculty	Choose an item.
Names of co-researchers, both internal and external Kindly provide names, responsibilities, and institutions. Please indicate N/A if there are no co-researchers.	N/A
Staff / Student	Student
Degree	Postgraduate Research
Name of Mentor / Supervisor	Dr. Mario Silic, PhD

Remarks from the supervisor or mentor

Supervisors should confirm the following for student applications before the study starts:

- The research issue is worthy of additional investigation.
- the student possesses the necessary abilities to conduct the study.
- the participant information sheet is suitable.

The protocols for enlisting research participants and obtaining informed permission are appropriate.

This is where the supervisor must add remarks. If this is not done, the application will be returned.

Click or tap here to enter text.

APPENDIX J:

SURVEY COVER LETTER

All prospective participants received a link to the Google survey along with this letter.

The correspondence was transmitted via email or social media sites like Discord or LinkedIn, contingent upon the researcher's original communication channel with the subject.

" I'm Gokul, a Team Lead at Accenture and a DBA scholar. For my doctoral thesis, I'm investigating data-driven business in India. I'm speaking with several industry experts as part of my research to get their perspectives. The method of doing a research interview is not complicated and should take 45 to 60 minutes. Do you want to take part in this?"

If, following this introduction, the potential participant agreed to participate in the study, the researcher then sent the following email.

Thank you very much for consenting to take part in this study.

I'll give a little introduction of myself and my motivations before we get in.

Right now, I work at Accenture as a Team Lead. I have worked in India's data science field for more than ten years.

I've witnessed several incredible triumphs throughout this time, as well as other firms that appeared promising but ultimately failed.

My driving force is the fact that India hasn't yet produced a data-driven, multimillion dollar company. This comes from a nation where the workforce for data-driven development is the largest.

I think that the industry as a whole as well as the participants will gain from my research.

If you feel comfortable answering these questions, kindly fill out the Google survey by clicking this link.

I've also included two more documents for you to check over and sign.

- ***An information sheet***
- ***A consent form***

Read the information leaflet in its entirety, please. This booklet includes information on the alternatives available to you, the nature of the research, and what to expect from the interview.

Please fill out the last page with your name, signature, and date after reading the information sheet.

You have choices about what happens to the information you disclose on the interview consent form.

Please read the interview consent form, mark the appropriate level of consent for direct quotation, sign and date the second page, and input your name.

Once you have read and signed both documents, kindly submit a soft copy. Another option is to save a copy for your documentation.

The consent form and information page also include my doctorate guide's contact details, which you can use to get in touch with her personally if you have any questions or need any explanations.

I appreciate it and hope to speak with you soon.

APPENDIX H:

GOOGLE SURVEY

The participants were instructed to finish the Google survey after receiving the email. There were twelve questions total, divided into three sections, on the Google Forms survey. It was necessary to finish each part before moving on to the following question.

SECTION 1 – General

S. No.	Type	Question
1	Checkbox	This survey is entirely optional. The researcher will record your responses. Your answers will be anonymized and aggregated when used in quotes. By filling out this survey, you indicate that you can represent your company. If you agree, please indicate this by selecting the checkbox below. If you disagree, please do not continue filling out the survey.
2	Radio Button	The researcher may want to interview you for 45 - 60 minutes to discuss your responses. Do you consent? • Yes • No

SECTION 2 – About the participant

S. No.	Type	Question
3	Text	Name
4	Text	Email Address
5	Radio Button	How many years of experience you have in the data science domain? • < 1 Year • 1 – 5 Years • 5 – 8 Years • 8 – 12 Years • 12 – 15 Years • 15+ Years
6	Multiple Choice	Which role do you play among these? Please select all that apply. • Executive (Any CXO) • Project Management • Data Scientist • Machine Learning Engineer • Data Analyst • Senior Data Analyst • Senior Data Scientist • Other (Please fill)

SECTION 3 – About the business

S. No.	Type	Question
7	Text	You are currently working in which company?
8	Text	What is your role/position in current company?
9	Radio Button	How long have you been working with the company you currently represent? • < 1 Year • 1 – 5 Years • 5 – 8 Years • 8 – 12 Years • 12 – 15 Years • 15+ Years
10	Radio Button	Are you using cloud services for ML model development or deployment? • Yes • No
11	Radio Button	How many ML Model, you have deployed using MLOps? • < 2 model • 3 - 5 model • 6 - 8 model • 9 - 12 model • 13 - 15 model • > 15 model

12	Multiple Choice	Which of these ML models are developed or deployed by you in your current company? Please select all that apply. • Classification • Regression • Clustering • Deep learning • Natural language processing • Computer Vision • Other (Please fill)