

# A Sociotechnical Framework for Operationalizing Machine Learning in the Banking sector

Yash Singh



# A Sociotechnical Framework for Operationalizing Machine Learning in the Banking sector

By

Yash Singh  
Student number: 5237491

**Master of Science**  
in Management of Technology

Faculty of Technology Policy and Management  
Delft University of Technology,

## **Graduation Committee**

Chairperson:	Prof.dr.ir. M.F.W.H.A Janssen, TU Delft
First Supervisor:	Prof.dr.ir. M.F.W.H.A Janssen, TU Delft
Second Supervisor:	Dr. Ben Wagner, TU Delft
External Supervisors:	Mr. Boris Smits, Deloitte
External Supervisors:	Mr. Pascal Lagerweij, Deloitte

# Acknowledgements

If you could travel back in time to the late 18<sup>th</sup> century, and tell people about mass electrification in the years to come, wouldn't that be an interesting discussion? To many people of that age, this transition might seem unimaginable and sometimes even frightening. However, owing to a number of scientific discoveries and engineering marvels most parts of the world today have safe and reliable electricity. While working as a machine learning engineer, I realized that as we transition to a future where artificial intelligence is widely used, a lot of questions remain unasked, a lot of findings remain undiscovered, and a lot of work is yet to be done. This has been my main motivation to return back to academic life. I am fortunate enough to have had the chance to conduct research on a topic that I have a deep interest in, and I would like to finally present my thesis on a 'socio-technical framework for operationalizing machine learning in the banking sector'. I hope this research, can in some way contribute to the development of safe and reliable adoption of artificial intelligence applications in the future to come.

I would like to start by thanking Prof. Marijn who has been my first supervisor and the chair of my graduation committee. Your guidance has been extremely insightful and has played an important role not only in shaping this research but also in shaping my thinking as an independent researcher. I would also like to thank my second supervisor Ben, who was able to join this project on such short notice. You have introduced a lot of energy into the research process and I have benefited greatly both from your critique and your ideas. I would also take this opportunity to thank Seda, who unfortunately could not continue with this project, but has provided a lot of support during the initial phases of the research. Next, I would like to extend my thanks to my external supervisors Boris and Pascal for being so actively involved in the research process. Your feedback along with our discussions and brainstorming has been very valuable. It has also helped me keep a check on what is feasible and what is not. I would also like to appreciate Deloitte as an organization, in particular, the Artificial Intelligence and Data team for being open to a new and challenging research topic, and providing me with a great environment to learn and grow.

Over the course of this research, I have had the chance to discuss my ideas with many of my colleagues at TU Delft and at Deloitte. Especially Sem, Agathe, and Sebastiaan. Thank you all for taking out the time to engage in critical discussions and helping me challenge my thoughts and biases. I would also like to extend my gratitude to all the anonymous expert participants who have taken out the time and effort to participate in this study and contributed to the research with their valuable knowledge.

This journey would have been extremely difficult and perhaps incomplete without the support of my family and close friends. Especially my parents and Krishma. Thank you to everyone who has been there for me and has supported me through this journey. Lastly, with a humble intention, I would like to thank myself, for hanging in there through the tough times and working hard to complete this thesis.

Yash Singh  
29<sup>th</sup> June 2022

# Executive Summary

Artificial intelligence techniques such as machine learning are playing a pivotal role in shaping the future of the banking industry. Given their enormous potential, banks have started experimenting with machine learning algorithms to reinvent their core business functions. However, banks have been struggling to convert machine learning experiments into enterprise-grade production applications, which limits them from realizing the true business potential of machine learning.

While machine learning algorithms have received significant attention both in academia and in the industry, there remains a need to develop a comprehensive understanding of how machine learning algorithms are operationalized in enterprises. In the context of this research, machine learning operationalization is explicitly defined as the process of converting machine learning experiments into consumable production services, and the sustainably managing the machine learning service throughout its life cycle. Machine learning operationalization is a multi-faceted process that involves a complex intertwining of social and technical systems within an organization. One of the reasons why organizations fail to realize sustainable machine learning applications in production can be attributed to the fact that they envision the ML operationalization process as a predominantly technical phenomenon, rather than a socio-technical phenomenon. *The main objective of this research is to develop a socio-technical framework, that supports the understanding and implementation ML operationalization process in the banking industry.* The research uses a theoretical lens of the socio-technical system theory which analyses the ML operationalization process based on four variables: technology, tasks, people, and their relationship structures.

By combining an extensive literature study and 15 expert interviews the framework identifies nine socio-technical factors and their relationship to create an integrated understanding of the machine learning operationalization process. These factors include *business user adoption, machine learning decision quality, and compliance & audibility* which are categorized as successful outcomes; *data quality, adaptability, enterprise integration, and risk management* which are categorized as challenges; *shared knowledge, and controls* which are categorized as drivers. To provide practical significance of the factors, the research conducts a case study to analyze the operationalization of a machine learning application for anti-money laundering, in a large Dutch bank. The findings from the expert interviews and the case study suggest that risk management is one of the most challenging yet crucial aspects of machine learning operationalization. From a socio-technical standpoint, it also remains an unexplored area in scientific and gray literature. Therefore the research narrows the focus by diving deep into the socio-technical challenges of managing risks during the operationalization process. To address these challenges, the research proposes four strategic guidelines: *moving from explainable algorithms to explainable ML operationalization, de-risking machine learning by design, integrating risk management with the agile workflow, and red team and blue team gamification.* Further, a conceptual model is presented for managing risks during the ML operationalization process.

The framework has both practical and scientific relevance. It can be used by organizations as a reference to develop a theory-driven understanding of the machine learning operationalization process and apply strategic guidelines to transform their risk management activities to complement their machine learning initiatives. The framework can also be used by scholars to understand the socio-technical aspects of ML operationalization, and transition from an algorithmic-centric approach of machine learning to a process-driven approach of operationalizing machine learning. Further, the limitations of the study are addressed to promote a transparent interpretation of the results and to provide an impetus for future research. This is followed by two directions to build on the current work. The first direction emphasizes increasing the diversity of expert participants and case studies to improve the internal and external validity of the research. The second direction focuses on refining the research with an in-depth analysis of certain socio-technical factors and empirically testing the proposed conceptual model to develop an established theory.

# Abbreviations

AFM: Autoriteit Financiële Markten (Authority for financial markets)

AI: Artificial Intelligence

CI: Continuous Integration

CD: Continuous Delivery

DNB: De Nederlandsche Bank (Central Bank of the Netherlands)

GDPR: General Data Protection Regulation

ML: Machine Learning

MLOps: Machine learning operations

Ops: Operations

Wwft: witwassen en financieren van terrorisme (Money Laundering and Terrorist Financing Prevention Act )

# List of Figures

Figure 1: ML Operationalization .....	12
Figure 2: Socio-technical perspective for ML Operationalization.....	13
Figure 3: Research Structure .....	16
Figure 4: Data Analysis.....	22
Figure 5: Topology of technical teams.....	30
Figure 6: ML operationalization workflow .....	31
Figure 7: Experimentation.....	32
Figure 8: Deploy and Monitor .....	34
Figure 9: Network representing the relationship between socio-technical factors .....	39
Figure 10: Diversity of Stakeholders based on knowledge, interests, goals and cognitive states .....	49
Figure 11: Inter-dependencies between stakeholders .....	50
Figure 12: Variations in accuracy vs explainability trade-off based on experiments of Herm et al, (2022) ...	52
Figure 13: Power-Interest in ML operationalization.....	53
Figure 14: Power vs Interest in risk assessment.....	54
Figure 15: A Conceptual model for managing risks during ML operationalization.....	57
Figure 16: Consent Form Page 1 .....	73
Figure 17: Consent Form Page 2.....	74

# List of Tables

Table 1: Literature Review: Keywords and Synonyms .....	17
Table 2: Literature Review- Inclusion and Exclusion criteria .....	17
Table 3: Expert Interview Participants.....	19
Table 4: Case Study Interview .....	21
Table 5: Summary - Literature Review.....	26
Table 6: Internal Stakeholders in ML operationalization .....	29
Table 7: External Stakeholders in ML operationalization .....	30
Table 8: Socio-technical factors - Successful outcomes .....	36
Table 9: Socio-technical factors- Challenges.....	37
Table 10: Socio-technical factors - Drivers .....	38

# Table of Contents

<b>1 INTRODUCTION.....</b>	<b>11</b>
1.1 BACKGROUND .....	11
1.2 ML OPERATIONALIZATION.....	11
1.3 SOCIO-TECHNICAL PERSPECTIVE FOR ML OPERATIONALIZATION .....	12
1.4 PROBLEM STATEMENT .....	13
1.5 RESEARCH OBJECTIVE AND RESEARCH QUESTIONS .....	14
<b>2 RESEARCH METHODOLOGY .....</b>	<b>15</b>
2.1 RESEARCH APPROACH .....	15
2.2 RESEARCH STRUCTURE.....	15
2.3 LITERATURE REVIEW .....	16
2.4 EXPERT INTERVIEWS.....	18
2.4.1 Interviewing Strategy .....	18
2.4.2 Interview Recruitment .....	18
2.4.3 Research Contribution based on different expert roles: .....	20
2.5 CASE STUDY ANALYSIS.....	20
2.5.1 Case Selection and Recruitment .....	21
2.5.2 Data collection .....	21
2.6 DATA ANALYSIS .....	21
<b>3 LITERATURE REVIEW .....</b>	<b>23</b>
3.1 ML PROCESS AND CHALLENGES .....	23
3.1.1 Data Dependencies .....	23
3.1.2 Hidden Technical Debt .....	24
3.1.3 Governance of ML models .....	24
3.2 MLOps PARADIGM .....	24
3.3 SOCIO-TECHNICAL SYSTEMS THEORY .....	25
3.4 RISKS AND TRUSTWORTHINESS OF ML APPLICATIONS .....	25
3.5 CONCLUSION OF LITERATURE REVIEW:.....	26
<b>4 SOCIO-TECHNICAL ANALYSIS .....</b>	<b>27</b>
4.1 STAKEHOLDER ANALYSIS .....	27
4.1.1 Identifying Relevant Stakeholders .....	27
4.1.2 Structure and Topologies of technical teams .....	30
4.2 ML OPERATIONALIZATION WORKFLOW.....	31
4.2.1 Experimentation.....	31
4.2.2 Model risk assessment.....	32
4.2.3 Application Engineering .....	32
4.2.4 Overall risk assessment.....	33
4.2.5 Deploy and Monitor .....	34

4.2.6 Consumption of ML Results .....	35
<b>4.3 CONCLUSION OF THE SOCIO-TECHNICAL ANALYSIS .....</b>	<b>35</b>
<b>5 SOCIO-TECHNICAL FACTORS .....</b>	<b>36</b>
<b>5.1 SOCIO-TECHNICAL FACTORS INFLUENCING ML OPERATIONALIZATION .....</b>	<b>36</b>
5.1.1 Successful Outcomes .....	36
5.1.2 Challenges .....	37
5.1.3 Drivers .....	38
<b>5.2 INTER-RELATIONSHIPS BETWEEN SOCIO-TECHNICAL FACTORS .....</b>	<b>39</b>
<b>5.3 CONCLUSION OF THE SOCIO-TECHNICAL FACTORS .....</b>	<b>40</b>
<b>6 CASE STUDY ANALYSIS: ANTI MONEY LAUNDERING .....</b>	<b>41</b>
<b>6.1 CASE BACKGROUND .....</b>	<b>41</b>
<b>6.2 THE SUCCESSFUL OUTCOMES: DECISION QUALITY, BUSINESS USER ADOPTION, AND COMPLIANCE. ....</b>	<b>42</b>
<b>6.3 THE CHALLENGES: DATA QUALITY, ADAPTABILITY, ENTERPRISE INTEGRATION, AND RISK MANAGEMENT .....</b>	<b>42</b>
<b>6.4 THE DRIVERS: CONTROLS AND SHARED KNOWLEDGE .....</b>	<b>45</b>
<b>6.5 CONCLUSION OF THE CASE STUDY .....</b>	<b>46</b>
<b>6.6 INTERMEZZO: NARROWING THE FOCUS .....</b>	<b>46</b>
<b>7 RISK MANAGEMENT IN ML OPERATIONALIZATION .....</b>	<b>47</b>
<b>7.1 TECHNICAL CHALLENGES OF RISK MANAGEMENT .....</b>	<b>47</b>
7.1.1 Model Opaqueness and Biases .....	47
7.1.2 Risks in the Operationalization process .....	47
<b>7.2 RISK MANAGEMENT IN A NETWORK OF STAKEHOLDERS .....</b>	<b>48</b>
7.2.1 Diversity of Stakeholders .....	48
7.2.2 Resources and inter-dependencies.....	49
7.2.3 Trade-offs and Negotiations .....	50
<b>7.3 SOCIAL CHALLENGES IN RISK MANAGEMENT: STAKEHOLDER CONFLICTS .....</b>	<b>52</b>
7.3.1 Knowledge Asymmetries .....	52
7.3.2 Power-Interest Asymmetries .....	53
<b>8 CONCLUSIONS.....</b>	<b>58</b>
<b>8.1 MAIN FINDINGS.....</b>	<b>58</b>
<b>8.2 PRACTICAL RELEVANCE .....</b>	<b>59</b>
<b>8.3 SCIENTIFIC RELEVANCE.....</b>	<b>59</b>
<b>8.4 LIMITATIONS.....</b>	<b>60</b>
<b>8.5 FUTURE RESEARCH DIRECTIONS .....</b>	<b>60</b>
<b>8.6 RELEVANCE TO THE MOT PROGRAM .....</b>	<b>61</b>
<b>8.7 PERSONAL REFLECTION .....</b>	<b>61</b>
<b>REFERENCES .....</b>	<b>63</b>
<b>APPENDIX.....</b>	<b>69</b>

<b>A. CODEBOOK.....</b>	<b>69</b>
<b>B. INTERVIEW PROTOCOL.....</b>	<b>70</b>
<b>C. RESEARCH ETHICS.....</b>	<b>72</b>
<b>D. DESCRIPTION OF SOCIO-TECHNICAL FACTORS.....</b>	<b>75</b>
D.1 Successful Outcomes .....	75
D.2 Data Quality .....	75
D.3 Adaptability.....	76
D.4 Enterprise Integration.....	77
D.5 Risk Management .....	77
D.6 Shared Knowledge .....	78
D.7 Controls.....	79

# 1 Introduction

## 1.1 Background

Innovation in advanced analytics has promised businesses to no longer depend on gut instincts but make data driven decisions and forecasts. Machine learning (ML) is one such advanced analytics technique, that uses past information to make accurate predictions. ML is considered to be a branch of artificial intelligence (AI) based on the idea that systems can ‘learn from data, identify hidden patterns, and make decisions with minimal human intervention’ (Al-Sahaf et al., 2019). Machine learning algorithms are capable of performing tasks, that human beings perform routinely without conscious introspection, for example, speech recognition, image understanding, driving, etc. At the same time, they are capable of performing tasks that exceed human capability by analyzing large datasets for genomic interpretation, weather prediction, and fraud detection (Shalev-Shwartz & Ben-David, 2014).

Organizations have turned to machine learning (ML) for diverse objectives such as increasing processes efficiency, creating new products and services, and improving decision-making, however, only a few have achieved substantial returns from their investments (Benbya et al., 2020). In a survey conducted by Gartner, 79% of the organizations said they were piloting AI and ML projects, while only 21% had applications in production (Gartner, 2020). In another survey conducted by Mckinsey, only 15% of the companies had been able to scale ML applications across multiple parts of the business, and only 36% had progressed beyond pilots (Panikkar et al., 2021). This highlights the paradoxical relationship between pilots and production, wherein launching ML pilots is deceptively easy, but deploying them as production applications can be extremely challenging.

Large financial institutions such as banks have shown a strong incentive to adopt ML to transform their businesses and remain competitive in the rapidly growing financial technology landscape. ML can introduce substantial economic and social benefits when applied to use cases such as fraud detection, anti-money laundering, market surveillance, prediction of credit risks, and automated loan approvals. However, realizing machine learning pilots into production applications involves several complexities in banks. From a technical perspective, data and system-level dependencies along with existing life-cycle management practices result in long lead times, low levels of reproducibility, and difficulties in updating the model. At an institutional level, ML operationalization requires collaboration between multiple stakeholders, effective management of the overall process, and compliance with regulations. The application of machine learning in the banking sector also involves high levels of risks which could not only lead to the failure of an individual institution but could also lead to profound consequences at a national and global level (Karimi & Yahyazade, 2021). ML applications must also undergo assessments before they can be deployed to production. Therefore, in order to convert machine learning projects into real-world applications, a multi-perspective view is required that acknowledges the social and technical factors involved in the processes.

## 1.2 ML Operationalization

Before defining ML operationalization, this section describes the process of developing and deploying ML models. This process can be broadly categorized into the experimentation phase and the production phase.

**Experimentation:** In the experimentation phase, data scientists develop an ML model by conducting multiple experiments with different algorithms, datasets, and tuning parameters to optimize a certain business metric. The phase begins with data extraction to integrate data from different relevant sources. This is followed by exploratory data analysis, to understand the schema and distribution of data. Data is then prepared for the ML

task by performing data cleaning and feature engineering. In the next, step different algorithms and hyperparameters are used to train an ML model. The trained model is evaluated on a set of metrics to gauge its quality and performance. Depending on the results of the evaluation, data scientists and engineers may choose to go back to one or more of the previous steps.

**Production:** In its final or the desired state, the ML application must run in a production environment to produce results for its purpose. Mature production deployments require that the model is recreated through an automated ML pipeline to ensure consistency, and to facilitate retraining of the model. The ML pipeline serves as an end-to-end construct that orchestrates data extraction, validation, preparation, model training, and model evaluation. Given the evaluation of the model yields expected results, the model is served or deployed in one of the following formats: - A microservice with a REST API to serve online predictions. - A model that executes on a batch dataset. The performance of the model is continuously monitored, based on which a new iteration in the ML process can be initiated to retrain or redevelop the model.

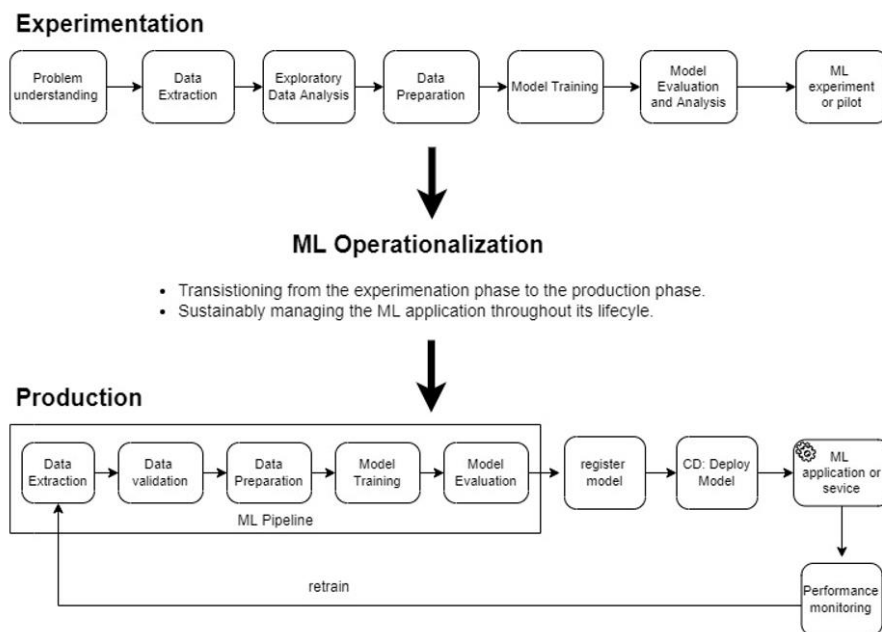


Figure 1: ML Operationalization

**Defining ML Operationalization:** ML operationalization refers to the process of transitioning from the experimentation phase to the production phase, and sustainably managing the ML application throughout its life-cycle. This transition requires a number of other activities which include development of other software components to host the ML model, risk assessment of the ML application, application deployment and operational support. In addition to the existing complexities of the ML model development, the operationalization also includes complexities relating to the management of important artifacts (code, data, and models) and the interaction between various stakeholders that influence and contribute to the process. The overall process including the experimentation phase, production phase, and ML operationalization is illustrated in Figure 1.

### 1.3 Socio-technical Perspective for ML operationalization

A socio-technical perspective is a research outlook that considers the interdependence between social systems and technical systems. This perspective stems from the socio-technical systems theory, as per which the technical systems consist of the technology and the tasks, whereas the social systems are concerned with the attributes of the people, and the structure of relationships within which they interact (Bostrom & Heinen, 1977a). These scholars further argue that the absence of a socio-technical perspective can lead to the failure

of technology design and implementation. This research therefore builds on the logical belief that a socio-technical perspective can support the operationalization of ML applications in the banking industry.

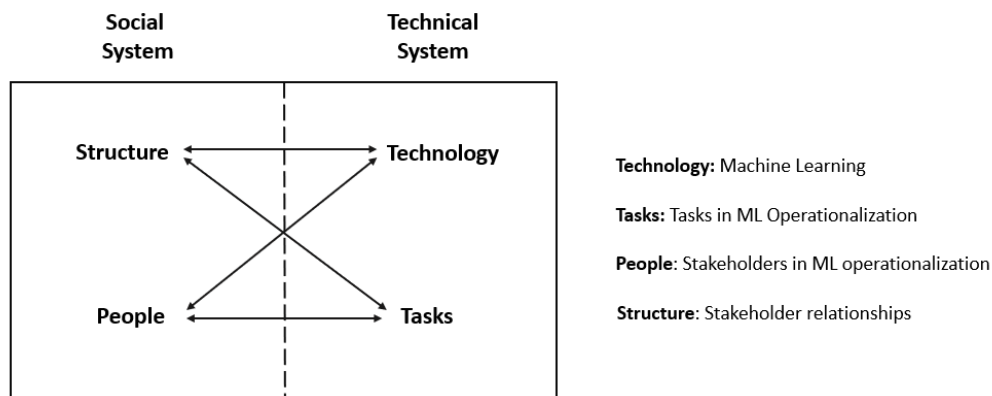


Figure 2: Socio-technical perspective for ML Operationalization

With respect to the technical system, the technology under consideration is ML which analyzed in the context of the ML operationalization process. This draws the focus to tasks which correspond to the development, deployment and maintenance of ML applications. The analysis of the social system focuses on the stakeholders involved in the process of ML operationalization. Further, the stakeholders must continuously collaborate, interact and negotiate with each other during the ML operationalization process which highlights their relationships or the structure of the socio-technical system. This approach is encapsulated in Figure 2.

## 1.4 Problem Statement

The banking industry has started adopting advanced analytics techniques like machine learning in the past few years. However, banks struggle to convert ML experiments into sustainable ML production services which limits them from realizing the full potential of machine learning. Even though, a number of paradigms such as MLOps, have been introduced, they are predominantly technical and practice oriented. There is therefore a need for an theory driven socio-technical approach that compliments the practice and also motivates research in this area. Therefore, the problem statement can be formulated as: *Organization such as banks are faced with several socio-technical complexities in operationalizing ML, which limits the banks from realizing the full potential of ML.*

This problem statement serves relevance from both an industrial and academic perspective. From an industrial perspective, it relates to a problem that widely exists in the technology landscape of the banking sector. The problem statement further addresses a complex organizational challenge, in the sense that 1) It is unstructured: there is no single right solution 2) It encompasses uncertainties: the behavior of stakeholders and the outcomes of the process cannot be easily predicted, 3) Has a dynamic nature: technology development in the banking industry is dynamically affected by new regulations and technology. The investigation of the problem can therefore lead to pragmatic insights for the development and deployment ML projects.

From an academic perspective, while a number of studies have been done to investigate the challenges in machine learning, the knowledge is not integrated and the inter-relationship between different socio-technical factors is rarely highlighted. Further, little knowledge is available on how ML projects are operationalized in the banking sector. Lastly, addressing such a problem statement bolsters conventional machine learning research to take into account the practical real-world challenges in the development of algorithms and other technical solutions.

## 1.5 Research Objective and Research Questions

The main objective of this research is to develop a socio-technical framework that supports the understanding and implementation of the ML operationalization process. To achieve this objective the research develops a comprehensive understanding of the socio-technical system within which the process of ML operationalization takes place, and identifies the socio-technical factors that influence the ML operationalization process. The framework lays an important focus on risk management by delving into the socio-technical challenges of managing risks during the ML operationalization process and provides guidelines for strategically improving the same. The framework can be used by organizations and practitioners to strengthen the analysis, implementation, and execution of the ML operationalization process. It can also be used by scholars for critically analyzing the concepts of the ML operationalization process and augmenting the same with further research. To achieve the objective the main research question is formulated as follows:

***RQ: How can banks use a socio-technical approach to support the ML operationalization process?***

The main research can be broken down into the following sub-research questions:

*Srq1: What are the socio-technical complexities in the ML operationalization process?*

The goal of this sub-research question is to identify and analyze the complexities involved in the socio-technical system within which the ML operationalization takes place. This includes the analysis of the key stakeholders and the ML operationalization workflow to develop a comprehensive understanding of the process and its intricacies.

*Srq2: What socio-technical factors influence the ML operationalization process?*

This sub-research question aims to integrate the various theoretical concepts that enable, hinder, or represent the desired outcomes of the ML operationalization process. This helps in developing a theoretical structure that holds the understanding of the ML operationalization process based on empirical findings.

*Srq3: Risk management was identified as a key challenge during ML operationalization. What are the socio-technical challenges specific to managing risks during the ML operationalization process?*

The previous sub-research question yielded a comprehensive list of socio-technical factors. To ensure the parsimony and specificity of the research, a decision was made to narrow the focus and analyze risk management which represents one of the most crucial yet challenging factors in the ML operationalization process. Section 6.6 further justifies this choice in more detail. The goal of this sub-research question is to conduct an in-depth investigation into the socio-technical challenges of managing risks during the ML operationalization process.

*Srq4: How can banks strategically address the challenges of risk management?*

This sub-research question aims to develop strategic guidelines and a conceptual model to address the challenges of risk management and thereby support the ML operationalization process. The findings of this sub-research question can be used by both industry practitioners for improving their risk functions and by academic scholars for evaluating and augmenting related research.

# 2 Research Methodology

## 2.1 Research Approach

Studies can be exploratory, descriptive, or explanatory in their nature. This study has a prominent exploratory nature since very few studies have been conducted in the area of ML operationalization, and the knowledge available is both scant and scattered. The exploratory research approach combines both extensive analyses of literature and expert interviews to explore the socio-technical factors that influence ML operationalization. In addition to this, the study also aims to explain and justify the cause-effect relationship between factors, which adds an explanatory component to the research.

In terms of the data being collected and analyzed, the research follows a qualitative approach. A qualitative approach can be useful in 1) supporting the researcher to understand the nature and complexity of the phenomenon of interest, 2) enabling research in relatively new areas of research, and 3) supporting the investigation of a phenomenon in its natural environment (Basias & Pollalis, 2018). Given the nascent state of ML operationalization, a qualitative approach enables the understanding of the complexities of the process through multiple perspectives, concepts, and logical beliefs in the contextual environment of the banking sector.

Lastly, with respect to reasoning and drawing conclusions, studies can follow a theory-driven deductive approach or an observation-driven inductive approach. However, studies such as this also combine both approaches by the means of abduction. In abduction, the researcher examines how the data collected supports the existing knowledge, as well as how the data collected may call for new findings and modifications to existing understandings (Flick, 2017).

## 2.2 Research Structure

The overall research can be structured in four phases: preliminary literature review, data collection and analysis, conceptualization, and research conclusion. Figure 2 highlights these phases along with the data sources used and the research activities and sub-research questions that constitute each of the phases.

Depending on the source from which the data is collected during the research, the data is classified as primary or secondary data. Primary data refer to information obtained first-hand by the researcher on the variables of interest whereas secondary data refers to information gathered from sources that already exist (Sekaran & Bougie, 2016). Literature which includes peer-reviewed journals, books, conference proceedings, and company documentation has been the main source of secondary data. While the literature has played a crucial role in understanding various concepts, challenges, and opportunities relating to ML operationalizations, the existing literature in the field is rather scattered and the number of scientific studies that analyze ML operationalization in the context of the banking sector is extremely scant. Therefore, the literature study is complemented with semi-structured expert interviews, which have aided a more in-depth and pragmatic analysis of the contextual environment. Data collection and analysis has been an interactive and iterative process, with both the literature study and expert interviews taking place simultaneously. The literature study has aided the formulation of expert interview questions and the creation of predefined qualitative codes, while the expert interviews have provided new search criteria for the literature study.

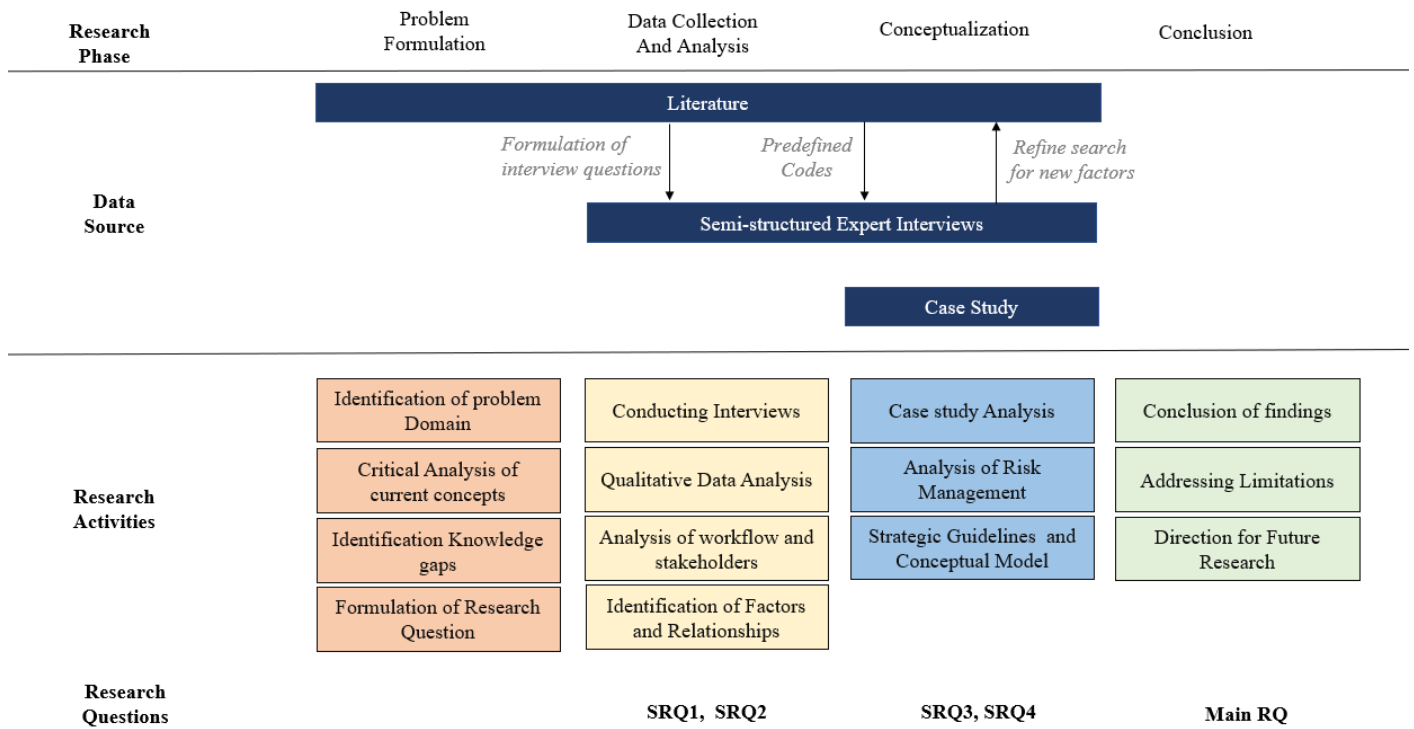


Figure 3: Research Structure

The preliminary literature review has contributed to the main inspiration of the research along with an in-depth understanding of the problem domain, analysis of the current knowledge in the field, the identification of knowledge gaps, and the formulation of research questions. The Data collection and analysis mainly consist of conducting semi-structured expert where audio recordings were made and transcripts were generated. Further, the qualitative data was reduced by the means of coding and interview summarization. In the conceptualization phase, the socio-technical factors were then extracted based on the analysis of the literature and interviews. The results are then triangulated and the framework is finalized. In the last phase, the main aim is to provide a set of strategic guidelines which aim to aid banks in operationalization ML applications along with a consolidated conclusion of the overall research. Lastly, scientific research is an imperfect process, it is, however, imperative that the limitations of the research are reflected on and addressed to provide future research directions.

## 2.3 Literature Review

A comprehensible literature review has served the purpose of understanding the existing knowledge in the area of ML operationalization. The literature review has been instrumental from the beginning of the study, for an in-depth understanding of the problem and the formulation of the problem statement. Further, the literature review has contributed to the critical analyses of the existing concept, thereby leading to the comprehension of the commonalities and differences of views proposed by different scholars, and the identification of research gaps that have inspired the research question for this study.

In addition to this, the study leverages the existing knowledge to design questions for the expert interview and for the determination of predefined codes to guide the qualitative analysis. The search process started with the selection of one primary database which was Scopus. The next step was to identify all the keywords and core concepts related to the central theme of ML operationalization. In order to store and organize the citations, Mendeley was used. Table 1 shows all the main keywords used in the search process along with their corresponding synonyms.

Main Keyword	Synonyms
Machine learning:	ML, Learning systems, learning algorithms, Artificial Intelligence, data sciences.
ML Operationalization	ML productionization, ML processes, ML deployment, ML life-cycle.
Organization	Company, institution, firm, enterprise
Banking	Banks, Financial institutions
MLOps	Machine learning operationalization, DevOps for ML, ML practices for production.
Socio-technical	Sociotechnical

Table 1: Literature Review: Keywords and Synonyms

The keywords were used in various combinations on the Scopus database, to identify and study the relevant publications. Google Scholar was used as a secondary database to supplement the search process with books, journal articles, and grey literature. Certain inclusion and exclusion criteria were applied in order to narrow down the results of the overall search process. These criteria were based on the language, document type, subject area, and relevance to the topic as explained in Table 2.

Criterion	Inclusion	Exclusion
Language	English	Other than English
Document Type	Peer-reviewed journals, books, book chapters, grey literature: conference proceedings, reports, and white papers.	Newspaper articles, Blogs, Patents
Subject Area	Information Systems, Social Sciences, Decision Sciences, Business & Management.	Computer science, Mathematics, Philosophy, Psychology.
Relevance	<ul style="list-style-type: none"> <li>- Papers related to ML operationalization.</li> <li>- Papers related to a socio-technical perspective of ML and ML processes.</li> <li>- Papers discussing various organizational and institutional aspects of ML and ML operationalization.</li> </ul>	<ul style="list-style-type: none"> <li>- Purely mathematical and computational</li> <li>- Purely philosophical</li> <li>- Focused solely on algorithm selection and performance.</li> <li>- Focused on using ML for improving a process.</li> </ul>

Table 2: Literature Review- Inclusion and Exclusion criteria

The literature review itself has been an iterative process, it required frequent narrowing and broadening of the scope to study different concepts. The goal of the search was to look at ML from a process viewpoint, therefore most of the articles selected initially were related to ML processes and operationalizations practices. To further narrow down, socio-technical aspects and organizational aspects of ML processes and operationalization were made relevant. When such criteria led to dead ends, broadening of relevance criteria was needed. Articles that look at ML as a whole were also selected to understand the inherent principle of ML and its relation to different social and technical aspects. This lead to new head-starts and helped in understanding the possible research gaps. All articles that were purely algorithmic, mathematical, and philosophical were excluded from the selection. Search with keyword combinations of “machine learning” and “process” resulted in several articles

that used ML techniques to improve a specific process, which were not considered relevant for this literature survey. Most of the publications in ML and ML operationalization were dated after 2015, therefore the publication date by itself was not a very important criterion that was applied to the search results. However, the publishing date along with the number of citations helped to cherry-pick articles when google scholar returned a plethora of results, or a broader concept of machine learning was being studied. Limited grey literature was included in the review in order to incorporate the perspective of the practitioners and eliminate publication bias. This mainly included conference proceedings, technical documentation, and white papers of companies that are front leaders in operationalizing machine learning.

## **2.4 Expert Interviews**

### **2.4.1 Interviewing Strategy**

Qualitative semi-structured interviews are one of the most widely used methods of data collection within social sciences and can be valuable in exploring subjective viewpoints and in gathering in-depth accounts of people's experiences (Flick, 2017). This study conducts semi-structured interviews in which the participant is asked a combination of open-ended and close-ended questions allowing the participant to share broader undiscovered perspectives, at the same time ensuring that the conversation is guided and aligned with the topic. The nature of semi-structured interviews further allows for the customization of the questions depending on the background of the interviewee. Further, the interviews employed a funneling approach which starts with broader questions followed by progressively asking reflective, and probing questions. The interview protocol used in this research can be found in Appendix B.

### **2.4.2 Interview Recruitment**

Since the research requires knowledge that is specific to a particular technology and industry sector, a non-probabilistic sampling approach is used which ensures that the participants of the interview are selected in a purposeful and systematic manner. Therefore study selected interviewees who have expertise in machine learning operationalization, preferably but not limited to the banking sector.

To ensure that the research finding incorporates diverse perspectives, the participants of the interview were recruited from different organizations and had different roles. Participants from three different Dutch banks were included in the study. The participants also included experts belonging to a large consultancy firm, who had expertise in the area of ML operationalization and had substantial experience working with clients belonging to the banking sector. Further, diversity in roles of the participant elucidates how different actors perceive the ML operationalization process and the challenges associated with it. Table 1 provides an overview of the experts included in the study.

<b>Expert ID</b>	<b>Role / Function</b>	<b>Role Description</b>	<b>Organization</b>
<b>E1</b>	Senior Data Scientist	Data scientist in the credit risk domain. Focused on developing ML models and working with platform teams to operationalize models in production.	Bank1
<b>E2</b>	Product Owner, ML Platform	Product Owner focused on developing internal products to enable end-to-end advanced analytics within the bank.	Bank 1
<b>E3</b>	Product Owner, Data Science	Product Owner focused on developing data science products for enhancing the information security of the bank.	Bank 1
<b>E4</b>	Senior Solution Architect, Data and AI	Senior Solutions architect specializing Data and AI, and integration of advanced analytics services (including ML) with enterprise architecture.	Bank 2
<b>E5</b>	Enterprise Advisor, Advanced Analytics.	Enterprise Advisor focused on developing a target operating model for advanced analytics within the bank. Also the head of the data science community and ML CoE within the bank.	Bank 2
<b>E6</b>	Senior Solutions	Solution engineer focused on developing platform capabilities for the experimentation and production of ML applications.	Bank 2
<b>E7</b>	Data Scientist	Data Scientist, having worked on multiple in the ML CoE, having worked on multiple data science use cases.	Bank 2
<b>E8</b>	ML Engineer	ML engineer working in a data science team, mainly responsible for developing software components of the ML application.	Bank 2
<b>E9</b>	Specialist Lead, AI	Manager and specialist lead in a leading consultancy with a strong focus on MLOps. Experienced in working with clients in the FSI sector.	Consultancy
<b>E10</b>	Specialist Lead, AI	Manager and specialist lead in a leading consultancy with a strong focus on MLOps. Experienced in working with clients in the FSI sector.	Consultancy
<b>E11</b>	Manager, Data and AI	Manager focused on strategic transformation and service delivery of data and analytics projects for clients in the Banking sector.	Consultancy,
<b>E12</b>	Product Owner, Data Science	Product Owner in the credit risk modeling team. Responsible for developing product vision and roadmap for the end-to-end delivery of ML applications.	Bank 3
<b>E13</b>	Team Lead Advanced Analytics	Focused on the technical implementation of advanced analytics solutions, and managing the overall process to ensure along with supporting end users to bolster the adoption ML.	Bank 3
<b>E14</b>	PhD Candidate	PhD candidate having a research focus on the design process of public AI systems from a socio-technical perspective.	Public University
<b>E15</b>	Senior Risk Consultant & PhD Candidate	A senior risk consultant working on model risk management of ML applications. The participant is also a part time PhD candidate researching in the area of Ethical AI	Consultancy & Public University

Table 3: Expert Interview Participants

### 2.4.3 Research Contribution based on different expert roles:

The research has benefited from the diversity of the experts participating in the study. This subsection describes how various expert roles have contributed to the research with their knowledge.

**Product owners:** Product owners provide a broader perspective on how machine learning products and services are developed and operationalized in the organization. They inform the research on how the communication takes place with different stakeholders, the expected business requirements, and the technical and non-technical challenges during the process.

**Data Scientists:** The data scientist mainly focus on developing the ML model and inform the research on how ML models are developed and deployed to production systems from a technical and process-oriented perspective. Also provide insights on the way of working and activities related to managing the life-cycle of the model.

**Architects :** Provide insights on how the overall ML solution is designed, implemented and integrated within the organization. They also possess a strong understanding of the business domain, data architecture and technology infrastructure in the organization. They also act as a bridge between the business and the IT and can provide insights on how various stakeholders are involved and aligned while developing technical solutions to support the machine operationalization process.

**Engineers:** The engineers inform the research about the various practices related to software development, data engineering, MLOps and DevOps workflows. They also elaborate on the interaction with data scientists in order to develop, deploy and monitor fully integrated ML applications.

**Specialist Leads, AI and Data:** Provide insights on how on defining and developing end-to-end data products for various clients, especially in the banking sector. They possess a mix of in-depth technical and managerial knowledge. Technically they can provide insight on the impact of data quality, MLOps practices, and technology-related challenges. From a managerial perspective they can provide insights on how teams organize, collaborate and work to deliver a ML application in production.

**Enterprise Advisor:** Provides information regarding the firm's vision with respect to advanced analytics and how it aligns with the strategic objectives and operating capacities. Informs the research with several internal and external factors that influence the machine learning operationalization process.

**Risk Consultants:** The risk consultant provide information relating to the identification and mitigation of the number of risks during the operationalization of ML applications. They also provide their perspective on interaction with data scientist and engineers, and further elaborate the challenges involved in managing different risks.

## 2.5 Case Study Analysis

Case studies provide an in-depth analysis of the phenomenon of interest within real life context (Sekaran & Bougie, 2016). In the context of this research, the case study aims to understand how the identified factors apply to a real use case that involves the operationalization of an ML application. With this, not only does the case study provide practical relevance to the study, but also provides an opportunity to test the transferability of the findings from the literature study and expert interviews in an organizational setting. Thereby acting as a means of triangulation, which refers to the process of corroborating data from multiple perspectives to enhance the depth of understanding and to provide verification of the results.

### 2.5.1 Case Selection and Recruitment

The case study was selected in a non-probabilistic manner similar to the recruitment of the expert interviews. The following criteria were considered while selecting the case: 1) The organization or the team has sufficient experience in ML experimentation and production deployment 2) At least one ML application is currently running in production and is being consumed by the business 3) The ML application involves certain risks which are acknowledged by the organization 4) The operationalization process exhibits a cyclic nature, where the ML application requires redevelopment or retraining. The case study selected analyzes the ML operationalization of an anti-money laundering application in a large Dutch bank, the details of which are shown in Table 4.

Expert ID	Role / Function	Role Description	ML Application	Organization
EC1	Data Scientist, Team Lead.	The participant leads a teams of data scientists in the transaction monitoring domain. The participant is involved in ML development and leads the collaboration with a number of stakeholders involved in the process.	Anti-Money Laundering	Bank

Table 4: Case Study Interview

The banking sector, owing to its nature values high degrees of confidentiality, as a result of which finding participants for in-depth investigation has been a challenging task. Therefore a single case-study was conducted which involved interviewing a data science team lead in a large bank. The information-gathering was done in one elaborate interview, proceeded by two follow-ups to provide further clarity on the data collected.

### 2.5.2 Data collection

The Data collection for the case study was done using semi-structured interviews. The expert participants were asked open questions with respect to the challenges encountered in the ML operationalization process, drivers that enabled the process, and the successful outcome of ML operationalization. This was followed by more specific questions that aimed to validate the identified socio-technical factors. While the overall interview protocol was similar to Appendix C, extra attention was given to organizational specific elements which included the topology of the technical teams, communication with business users, and interaction with risk assessment teams. Since the research narrowed the focus to risk management, two follow-ups were conducted with the participant in order to get more specific details of different risk assessment processes and to understand the controls used during the development of the ML application. The data collected was unified into a single document containing the summary of the entire discussion.

## 2.6 Data Analysis

The methods used for data analysis apply to both the expert interviews and to the case-study interviews. The data collected from the interviews were stored in the form of audio recordings and text transcripts. The goal of the data analysis was to reduce the vast qualitative information into conceptual findings from which socio-technical factors influencing ML operationalization can be extracted. To perform the data analysis, Atlas.ti, a computer-assisted qualitative data analysis software was used.

In the first step, the audio recordings and transcripts were imported to Atlas TI, where the transcripts were cleaned by re-listening to the audio recordings. The cleaning process then involved the creation of interview summaries and the assignment of qualitative codes. As mentioned previously, abduction which combines a deductive and inductive approach has been used for data analysis in this research. A number of predefined

codes were developed based on the literature study, these predefined codes were then investigated for confirmation in the transcripts. While the predefined codes provided a starting point and structure for the data analysis, an open mind was kept for new codes and themes that could emerge from the data. This first phase of coding predominantly involved the assignment of open codes which corresponds to one or more socio-technical factors. This is followed by organizing codes into different categories and developing relationships between them. Once the socio-technical factors are identified, the next task is to establish relationships between them. Some of the relationships were derived from the codes using the network analysis tool of Atlas.ti. This process is depicted in Figure 3.

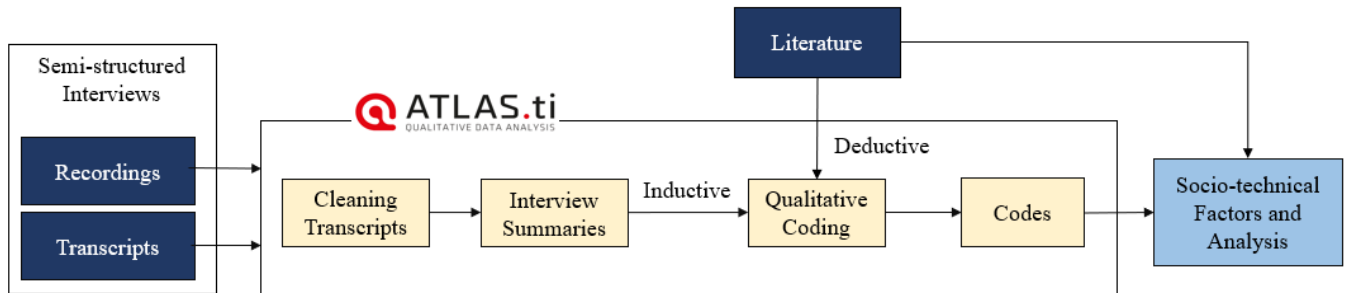


Figure 4: Data Analysis

To ensure that the data collection and analysis were ethical and compliant with regulations such as GDPR, the research was assessed and approved by the Human Research Ethics Committee (HREC) of TU Delft. Some of the steps followed to reduce the risk in the study included informed consent from the expert participants, anonymization of personal information of the participants, and the right to withdraw from the study at any point in time. The HREC process is explained in more detail in Appendix C.

# 3 Literature Review

Academic research in the field of machine learning has predominantly focused on computation and mathematical aspects to produce sophisticated algorithms and techniques. This has positively contributed to successful proof-of-concept that demonstrated how ML can enhance productivity and increase profitability for businesses and economies. However, some scholars argue that ML solutions developed in research often do not consider the practical implication of the real world (Domingos, 2012; Sarwate & Chaudhuri, 2013). Recently, there has been a growing interest in applied ML research, which recognizes the practical challenges and opportunities in operationalizing ML applications for real-world uses cases (Collins et al., 2021; Mikalef et al., 2019). Even though ML models are operationalized as software applications (in most cases), researchers and practitioners posit, ML techniques introduce distinct complex challenges which create a dichotomy between ML and traditional software operationalization methodologies (Treveil et al., 2020). There has also been increasing awareness in the scientific community that the process of operationalizing machine learning cannot be viewed purely from a technical and procedural lens, it involves several social factors relating both to its functioning within organizations and the ethical consequence of its results. Therefore, a socio-technical perspective is imperative to understand and mitigate the challenges of ML operationalization. These themes are further elucidated in the following subsections.

## 3.1 ML Process and Challenges

A number of efforts have been made to create a standards process for developing and deploying data mining and analytics applications. This includes the KDD: Knowledge Discovery in Database (Fayyad et al., 1996), SEMMA: Sample explore Modify, Model, Assess and CRISP-DM: CROss-Industry Standard Process for Data Mining (Wirth & Hipp, 2000). While these standards take into consideration important stages required to create intelligent data-driven applications, they do not explicitly state the steps specific to learning algorithms, which could otherwise affect the generalizability of the standards. More specific to machine learning use cases, the process of delivering a machine learning model to production involves the following steps: 1) Data extraction 2) Data analysis 3) Data preparation 4) Model training 5) Model evaluation 6) Model validation 7) Model serving 8) Model monitoring (Salama & Kazmierczak, 2021). Intuitively these steps seem sequential, but in reality, they are highly cyclic and iterative, owing to which the process is often viewed as a continuous life-cycle, referred to as Machine learning life-cycle (Garcia et al., 2018).

### 3.1.1 Data Dependencies

The initial phases of the model development involve the extraction, analysis, and preparation of data. Depending on the availability and quality of data, these initial steps can be excruciatingly long and challenging. Furthermore, both the performance and behavior of models are tied to the data on which the model is trained (Ashmore et al., 2021; Polyzotis et al., 2018). Therefore, if the data is imbalanced or skewed, it can result in a biased ML model, which has led to unfair and discriminatory decisions in many real-world applications (Mehrabi et al., 2021). Data in large financial institutions exhibits complex behaviours: ‘non-stationarity, non-linear interaction, heteroscedasticity, and biases’ (Creamer et al., 2021). From an organizational point of view, data is collected and processed through a chain of activities involving multiple actors and systems, thereby resulting in a data chain. This implies, that the quality of decisions made by machine learning models is not only influenced by data quality (accuracy, timeliness, completeness, consistency, and relevance) but also by how the overall data chain is managed (Janssen et al., 2017). Finally, the nature of real-world concepts (customer preferences, financial transactions, etc) are not stable and can change with time, as a result, the underlying data distribution might change as well. This can make ML models trained on old data inconsistent with the new data. This problem is referred to as concept drift or data drift which complicates the task of retraining ML models (Tsymbol, 2004).

### 3.1.2 Hidden Technical Debt

Technical Debt is a term used to describe the cost (time, money and resources) incurred by expediting the delivery of software. Technical debt may be paid by refactoring code, better testing, reducing dependencies, and improving documentation (Fowler, 2018). Sculley et al. (2015) elaborate that ML systems have a special capacity for incurring technical debt which exists at the system level, this not only makes it difficult to detect but also difficult to address by the typical methods of paying down technical debt. In software systems, strong abstraction boundaries based on encapsulation and modular design simplify the process of maintaining code and isolating changes. However, creating such abstraction boundaries in ML systems is more challenging since, ML systems are sensitive to changes in other systems, external data, and feedback loops. Further, as different data scientists and engineers work on multiple experiments to produce a working model, tracking experiments and reproducing them for refactoring can often be a challenge.

### 3.1.3 Governance of ML models

There are several complexities involved in managing and controlling ML models and life-cycle processes in an organization. ML experiments are often conducted in an ad-hoc style without a standardized way of storing and managing the resulting artifacts, making it difficult to track and reproduce them at a later point in time (Schelter et al., 2018; Zaharia et al., 2018). Furthermore, organizations develop and deploy multiple ML models to serve different business needs. These models may be built by different teams having their own custom tools and practices for addressing requirements of different stages of the machine learning life-cycle (Hummer et al., 2019). ML models may also be shared across business cases. In some cases a particular ML model may be reused to serve another business need (Hernández-Orallo et al., 2016) or different ML models may be combined together to solve another problem, referred to as ‘ensemble learning’ (Zhou, 2021). Lastly, since ML models in production are continuously updated due to retraining with new data and changes in code, they require proper model versioning.

Without proper governance, as the number of ML models increases, the amount of inefficiencies and disorganization in the ML processes also increases, hindering organizations from attaining the maximum possible return on their ML investment. Since multiple teams with different expertise work together to operationalize an ML model, such a chaotic entanglement of models can also result in a lack of shared understanding and accountability. The importance of ML governance is to ensure the desired outcomes and behavior of ML models, with mechanisms that can monitor, regulate and provide accountability. While a rich body of literature exists for IT and data governance mechanisms, for governance mechanisms specific to ML the literature is sparse (Schneider et al., 2020).

## 3.2 MLOps Paradigm

To tackle many of the mentioned challenges, a new paradigm termed as MLOps ( machine learning operations) has emerged. MLOps suggests practices for managing and monitoring models, model quality assurance, and the storage and discovery of artifacts (van den Heuvel & Tamburri, 2020). Scholarly articles analyze different parts of the ML workflow and introduce techniques for optimizing the efficiency, reliability, and reproducibility of the process (Hummer et al., 2019; Muthusamy et al., 2018; Ruf et al., 2021). On similar grounds, Van den Heuvel & Tamburri (2020) also introduce an Intelligent Enterprise Applications Architecture (iA2 ), consisting of the data layer, intelligence layer, and the data-driven application layer to create a logical separation for the MLOps process. However, given its recent introduction, there is limited research available on MLOps practices, and more empirical evidence is needed to verify the proposed advantages. Furthermore, little research has been done on how companies practice MLOps, with only two papers that considered collaboration between different stakeholders (John et al., 2021), and organizational roles (Ruf et al., 2021) in the MLOps process.

### 3.3 Socio-technical Systems Theory

A socio-technical perspective is a research outlook that considers the interdependence between sociological systems and technological systems. It emphasizes that the two systems must be jointly optimized to produce positive practical outcomes. The theory was introduced by Eric Trist, Ken Bamforth and Fred Emery, in their study of English coal mine during the world war II era. In recent years, socio-technical perspective has been a foundational viewpoint in Information System research and has helped in comprehending the relationship between information technology and social collectives (Sarker et al., 2019). In their seminal work, Bostrom & Heinen (1977a) describe how the absence of a socio-technical systems approach lead to an inadequate design of information systems, thereby leading to their failure. The technical system consist of the tasks, processes, and technology that transform inputs to outputs. Whereas the social system is concerned with the attributes of people, and their relationships. Their further posts that a socio-technical systems approach is key to developing a realistic view of an organization and making changes within it.

While still at a nascent stage, a number of researchers have approached the process of ML from a socio-technical perspective. Cabitza et al. (2020) highlights various socio-technical challenges relating to ‘hiatus of human trust’ which refers to lack trust among the users of ML and ‘hiatus of machine experience’ which refers to problem with data accuracy and availability. Further, the unrealistic expectations and ineffective communication in organizations operationalizing ML are also acknowledged (Baier et al., 2019). Significant work has also been done in developing best practices and frameworks to bolster the adoption, integration, and deployment of ML in organizations (Desouza et al., 2020; Fountaine et al., 2019; Makarius et al., 2020). While these works provide strong conceptual grounds based on theory, they do not critically analyze the interplay between technical and social aspects, or the possible trade-offs between conflicting goals. A predominantly technical approach by Robbins (2020), introduces an ‘envelopment’ around AI and ML applications by specifying constraints on different properties. This theory of envelopment is approached from a socio-technical perspective and applied to real use-cases in the public sector by Asatiani et al. (2021). The research elaborates how the envelopment technique can be used to manage the accuracy-explainability trade-off, in which the explainability of AI models decreases as the complexity and accuracy of the model increases (Adadi & Berrada, 2018). Other than this, no related work was found to acknowledge the different trade-offs in the operationalization process. Furthermore, the concept of AI and ML envelopment is built on the premise of tightly coupling the application to specific constraints, which scholars argue could hinder innovation in applications such as chatbots and self-driving cars that use machine learning for strong AI goals (Formosa & Ryan, 2021).

### 3.4 Risks and Trustworthiness of ML applications

The issues of data dependency, technical debt, and ML governance are closely intertwined with the organizational processes, and collaboration between stakeholders. However, as machine learning pervades into important applications in the public and private domain, it is becoming a central force in society. This introduces other socio-technical obstacles relating to the transparency, explicability, fairness, and accountability of ML systems that hinders its successful operationalization. AI systems, including machine learning, have an opaque nature, which makes it difficult to explain the outcomes of the system. This compromises the transparency and accountability of ML systems, making them difficult to trust (Enni & Herrie, 2021; Rossi, 2018). Experiments conducted to analyze decision making using business rules and ML, showed that even though both business rules and ML could make mistakes, mistakes made by black-box ML algorithms were harder to detect compared to white box business rules (Janssen et al., 2020). As a result, a large number of enterprises that have adopted machine learning technologies fear liability issues (Rossi, 2018). Another important concern is data privacy which has sparked several controversies worldwide. In order to achieve high accuracies, ML models must be trained on large relevant datasets from different sources, which often raises privacy concerns (Mohassel & Zhang, 2017). Organizations must therefore carefully assess the risks and benefits of using private data. The General Data Protection Regulation (GDPR) introduced in the European Union, puts forward principles to ensure data is protected and the decisions made by ML algorithms

can be explained and accounted for. Therefore creating an additional impetus for organizations to build auditable ML systems and processes. In response to the concerns on the trust and black-box nature of ML systems, the field of explainable artificial intelligence (XAI) has recently gained widespread academic attention. XAI employs ML techniques such as Taylor decomposition (Montavon et al., 2017), layer-wise relevance propagation (Bach et al., 2015), counterfactual algorithms (Chou et al., 2022) to produce more explainable ML algorithms without compromising performance. However, most of the academic research in the field is based on mathematical and computational constructs. Also, explainable algorithms do not always translate to explainable processes. De Bruijn et al (2021) argue that questions like which actors will be involved, what roles will the involved actors have, what main issues will be discussed, still remain unanswered, and therefore there is a need to move from explainable algorithms to explainable processes. such as training data, input, output, and functions to regulate the ML application.

### 3.5 Conclusion of Literature Review:

Focus Area	Literature Source	Key Findings
Engineering	(Sculley et al., 2015) , (Schelter et al., 2018), (Zaharia et al., 2018), (Lwakatare et al., 2020)	Elaborate on the technical challenges and opportunities in developing, deploying, and maintaining ML applications in production.
MLOps	(Ruf et al., 2021), (Zhou et al., 2020), (John et al., 2021), (Treveil et al., 2020), (van den Heuvel & Tamburri, 2020)	Use the MLOps paradigm to view the process of ML operationalization. Provide insights on the technical workflow, practices, tools, and the involved roles.
Data dependencies of ML	(Polyzotis et al., 2017), (Polyzotis et al., 2018) (Renggli et al., 2021) (Ashmore et al., 2021) (Janssen et al., 2017) (Janssen, Brous, et al., 2020)	Focus on challenges with respect to data and its impact on ML and ML operationalization.
Socio-technical theory	(Bostrom & Heinen, 1977a) (Bostrom & Heinen, 1977b) (Sarker et al., 2019)	Introduce socio-technical systems theory in the context of information systems. Further, highlight the application and importance of the same.
Socio-technical ML development	(Cabitza et al., 2020) (Robbins, 2020), (Asatiani et al., 2021) (Passi & Jackson, 2018)	Provide an overview of socio-technical challenges and practices for developing and implementing ML.
Risks and Trust-worthiness of ML	(Enni & Herrie, 2021)(Rossi, 2018) (de Bruijn et al., 2021)(Adadi & Berrada, 2018) (Mohassel & Zhang, 2017)	Focus on various risks in ML systems with respect to bias, explainability, privacy, and accountability.

Table 5: Summary - Literature Review

The main topics and key findings of the literature study have been organized in Table 5. The findings suggest that operational aspect of ML (development, deployment and maintenance ) have been gaining prominence in literature in the recent years. However, these studies predominantly focus on the technical aspect of ML operationalization. While a number of studies analyze socio-technical aspects of ML, they do not do so in the context of ML operationalization. It can therefore be conclude that, despite the growing interest in the ML operationalization, there remain a need to investigate the process from a socio-technical perspective. In addition to this, while risk in ML applications remains a widely discussed topic, there is little empirical evidence on how organization such as banks identify and manage risks during the operationalization of ML applications.

# 4 Socio-technical analysis

The goal of this chapter is to analyze the socio-technical system in which the ML operationalization process takes place. A stakeholder analysis is presented to identify the various stakeholders that influence the process. This is followed by the analysis of the ML operationalization workflow which integrates the understanding of how stakeholders interact with one another on different tasks to develop, deploy and maintain an ML application. The workflow, therefore, captures all the four variables ( people, structure, technology, and tasks) to highlight the socio-technical complexities of the ML operationalization process. The analysis of this chapter is developed through a combination of literature study and expert interviews. The expert interviewees are referenced through their expert id.

## 4.1 Stakeholder Analysis

The activities associated with machine learning are widely considered to be technology-oriented. However, a process such as ML operationalization comprises of heterogeneous interactions and collaboration between diverse stakeholders. With respect to a socio-technical perspective, the seminal work of Bostrom & Heinen (1977), posits that the social systems consist of the attributes of the people (skills, values, interests) and the structure of their relationships. A failure to understand the social system often results in the dysfunctional design of technology systems and processes. This section, therefore, delves into the analysis of the social system through stakeholder analysis. The findings of this section have been developed by combining both literature and expert interviews.

### 4.1.1 Identifying Relevant Stakeholders

The definition of a stakeholder in literature can be ambiguous. This research follows the definition of Schilling (2000) and Carroll et al. (1996) who consider stakeholders as individuals and groups who have a “stake” in a firm or process. This enables a more inclusive approach that takes into account the actors implementing the process, the parties influencing the process, and the parties being affected by the process. The stakeholders can be broadly classified as internal stakeholders that exist within the environment of the bank, and external stakeholders, that exist outside the environment of banks, there are identified in Table 6 and Table 7 respectively. This study acknowledges the external stakeholders, however, the focus of the analysis remains on the internal stakeholders. The internal stakeholders are grouped into an area of work that best represents their interests.

**Business:** The Business corresponds to a certain business function within the bank, for example, retail, banking, private banking, mortgages, etc. In an ideal case, a member of the business defines and proposes the business problem that must be solved using ML techniques, this member is referred to as the problem owner (Ruf et al., 2021). After an ML application is developed and deployed, the results of the ML application are consumed by a business user to make a certain decision or prediction. As an example, for an ML application that predicts if a particular client should be granted or denied a credit loan, the business user is a credit risk manager. The business users have an important stake in the ML operationalization process and can strongly influence the process outcome. If they are not satisfied with the ML application, do not trust the model, or cannot easily access the results they may not adopt the application, thereby rendering the ML application redundant (Passi & Jackson, 2018).

**Data Science:** Consists of data scientists and data science product owners. The data scientists conduct experiments using different ML algorithms, features, and hyper-parameters to develop and select an ML model. Data scientists are also involved in various phases of the operationalization process (Ruf et al., 2021). The data science teams must work in close collaboration with the business in order to understand their

requirements and benefit from the business domain knowledge. They must also collaborate with the engineering teams for multiple technical capabilities and are often dependent on the data providers to provide them with relevant data that is fit for the purpose of developing a model. In addition to this, they must work with the risk assessment teams to identify and mitigate the possible risks with the respect to operationalizing the ML model. A product owner having a data science background creates a product vision, and roadmap and coordinates the communication and collaboration of the data scientists with other stakeholders. Therefore

**IT engineering** represents individuals or groups that are responsible for providing technical and engineering capabilities for production-grade ML applications. This could include, ML engineers, Data Engineers, Infrastructure or Platform engineers, and technical architects (Treveil et al., 2020). The required engineering capabilities can vary from building software around the ML application to engineering data pipelines, or providing infrastructure and platform support. The IT engineering teams work in close collaboration with data scientists in order to develop an end-to-end ML application (Ruf et al., 2021). Further, the engineers may also take part in risk assessment activities that evaluate the risk of ML applications or platforms. The data scientists and engineers represent a technical group that is mainly responsible for the implementation of the ML operationalization process.

**Data providers:** are either individuals or teams who have ownership of the data. In banks, data providers usually reside within a particular domain and take responsibility and ownership of the data within that domain (E1, E6). The two common data provider roles include data owners and data stewards. The data owners are responsible for defining policies that ensure the quality, security, and compliance of data (Brous et al., 2020). Whereas the data stewards concern themselves with the supervision and implementation of the policies by monitoring data quality and data usage (Janssen, Brous, et al., 2020). While they are not directly involved in the ML development and deployment, they decide how the data will be accessed and are responsible for communicating any changes that could take place with respect to the data under their ownership.

**Risk:** The risk assessment, consists of stakeholders responsible for evaluating the risks in the development and deployment of ML models. They are primarily concerned with identifying the likelihood and impact of undesirable events that could occur as a consequence of operationalizing an ML application (Alhawari et al., 2012; Karimi & Yahyazade, 2021). The possible stakeholder roles include compliance officers, ethics and privacy officers (E4, E9, E14), model risk validators (Treveil et al., 2020) (E1, E5, E12, E15), and experts from the chief information security office (CISO) (E3, E4, E12). These stakeholders must be involved early in the process starting from the problem formulation and experimentation. The involvement of the risk assessment team is essential to ensure the ML development and deployment are lawful, ethical, and reliable.

**Regulators:** The regulators work to ensure the enforcement of laws and regulations across various banks. In the Netherlands, the two primary financial regulators are The Dutch Authority of Financial Markets (AFM) and the central bank of the Netherlands (DNB). In the context, of ML operationalization they indirectly influence the process through laws, regulations, and supervision. In certain use cases, where ML is used for regulatory models (eg: detecting financial crime, credit risk management, etc.), the regulators through their regulatory guidelines, can impact the underlying assumption of the ML model and its implementation (E1, E12).

**End Customers:** The end customer refers to the clients and customers that avail of the service of the bank. They have no direct influence on the ML operationalization process but can be impacted by the decision made by an ML application. The end customers, while are not involved in the ML operationalization process, they rightfully expect fair decisions and explanations in the case the decision is unfavorable (de Bruijn et al., 2021; Hacker et al., 2020). Banks must operate in the interest of the end customer in order to remain lawful and trustworthy.

Stakeholder Group	Stakeholder role	Skills / Expertise	Involvement	Key Interest	Preferences
Business	Problem Owner	Business domain knowledge	Define a business problem to be solved with ML.	Improvement of the business function.	<ul style="list-style-type: none"><li>• Accurate results from the model.</li><li>• Ease of use &amp; Adequate technical support.</li><li>• Interpretable results.</li><li>• Intuitive results.</li></ul>
	Business User	Business domain knowledge	Consume the results of the ML model to make certain business decisions.		
Data Science	Data scientists	Data science, ML	Solve a business problem using ML techniques.	Performance of the ML model.	<ul style="list-style-type: none"><li>• Access to high-quality data.</li><li>• Ability to use complex ML modeling techniques when required.</li><li>• Work on new and innovative projects</li><li>• Optimization of model performance metrics</li></ul>
	Product Owner	Data Science and Product management.	Sets product vision, aligns with other stakeholders, sets priorities.	Innovation and New challenges.	
Data Providers	Data Owners	Data governance, data management, data quality.	Manage data access, and set policies for data governance.	Data privacy and governance.	<ul style="list-style-type: none"><li>• Restricted access to data and controlled usage.</li></ul>
	Data Steward	Data governance, data management, data quality.	Supervise the management and governance of data.		
IT Engineering	ML Engineer	Software engineering, Machine Learning, MLOps	Software engineering for ML models, continuous integration & deployment.	Engineering performance	<ul style="list-style-type: none"><li>• Access to structured and high-quality ML model code.</li><li>• Ability to use complex ML engineering techniques when needed.</li><li>• Optimization of engineering metrics such (testability, scalability, and availability).</li></ul>
	Data Engineer	Data ingestion, transformation, streaming, etc.	Transforming data, and engineering data pipelines for ML models.	Innovation and new challenges.	
	Technical Architect	IT System design and integration	Making design choices for compute, storage, and networks.		
	Ops / Platform Engineers	DevOps, IT infrastructure.	Platform provisioning, and operational maintenance		
Risk	Compliance officer	Laws and Regulations	Asses the ML process for regulatory and compliance risk.	Minimization of risks	<ul style="list-style-type: none"><li>• Restricted access to data and controlled usage.</li><li>• ML model is sound and reliable</li><li>• All the applicable laws and regulations are adhered to.</li><li>• The process is transparent and explainable.</li></ul>
	Model Risk Validators	ML, Data Science, Business Domain	Validate the ML model for soundness		
	Legal and Ethics officers	Ethical principles, laws and regulations	Assess the model for fairness and explainability		
	CISO	Information security and risk.	Asses the ML process and application for IT risks.		

Table 6: Internal Stakeholders in ML operationalization

Stakeholder	Involvement	Expectation and Concerns
<b>DNB</b>	Supervise and provide guidelines for the develop ML, to ensure the stability and soundness of banks.	<ul style="list-style-type: none"> <li>Compliance with the applicable financial and non-financial regulations.</li> <li>ML development and deployment are sound, accountable, fair, and transparent.</li> </ul>
<b>AFM</b>	Supervise and regulate the financial markets. Provide guidelines for	<ul style="list-style-type: none"> <li>ML solutions developed by banks promote fairness and transparency in the financial markets.</li> </ul>
<b>European Commission</b>	Formulate EU wide regulations on data and AI like the GDPR, and upcoming AI act.	<ul style="list-style-type: none"> <li>Compliance with EU wide regulations.</li> </ul>
<b>End Customer</b>	The end-users are impacted by the operationalized ML model. They influence the profitability and reputation of the bank.	<ul style="list-style-type: none"> <li>Fair decisions are made.</li> <li>Expect an explanation in the event of an undesirable decision.</li> </ul>

Table 7: External Stakeholders in ML operationalization

#### 4.1.2 Structure and Topologies of technical teams

To further understand the interaction between stakeholders this section discusses how technical teams are organized for the end-to-end operationalization of ML projects. The development and deployment activities with respect to an ML model are initiated and pursued by data science teams. The data science teams are often set up as cross-functional teams with members who have different expertise (Raina & Krishnamurthy, 2022) . A typical structure of a data science team consists of a product owner, technical architects, data scientists, ML engineers, and data engineers and ops engineers. In addition to this, data science teams consist of a member from the business to provide expertise in the business domain (Treveil et al., 2020).

Despite the cross-functional nature of data science teams, all the technical capabilities may not be centered within a single data science team. Most of the experts participating in the study mentioned that cross-functional data science teams work with other engineering and operations teams throughout the operationalization process. As a result, engineering and operations activities can be distributed among the different teams in different ways. Based on this, four team topologies were identified which are depicted in Figure 5.

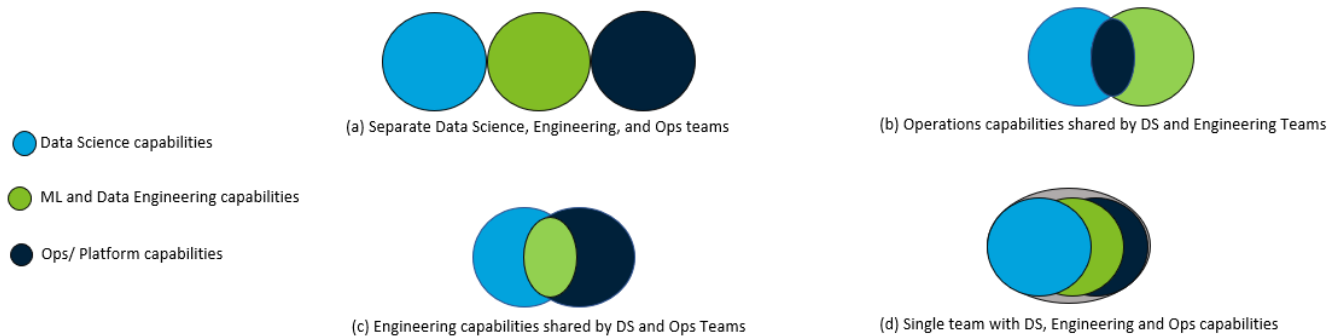


Figure 5: Topology of technical teams

Figure 5(a) represents a clear distinction between ML development, engineering and operations, which usually exist in new teams that are adopting or trying to move towards an MLOps or DevOps paradigm. In Figure 5(b) the operations capabilities are shared between data science and engineering teams, which could be applicable for an organization where the operations and infrastructure capabilities lie within the engineering teams, or the

infrastructure operations are covered by a self-service model. In Figure 5(c) the engineering capabilities are shared between the data science and operations teams. This configuration is well suited for organizations that have data science teams with mature engineering capabilities, or operations teams that have wider engineering capabilities to support data science teams. Lastly, Figure 5(d) shows a fully integrated data science team that takes responsibility for all the capabilities covering model development, application engineering, and operations.

## 4.2 ML operationalization Workflow

The ML operationalization workflow explicates a series of tasks and how different stakeholders interact one another during the execution of these tasks. The workflow is aimed at the development, assessment, deployment and maintenance of the ML applications. A predominantly technical overview of these tasks is provided by the Salama & Kazmierczak (2021) and Ruf et al (2021). By combining the knowledge from these works with the information garnered from expert interviews, a comprehensive workflow representation is created which takes into consideration the decision making processes and the collaboration between different stakeholders, in addition to the technical tasks. As shown in Figure 6, the workflow represents a sequence of six tasks, of which, tasks corresponding to ‘experimentation’ and ‘deploy and monitor’ are compound tasks consisting of a collection of sub-tasks. These tasks are expanded and elaborated in Figure 7 and Figure 8 respectively. It is important to note that the ML operationalization is highly interactive and cyclic process that does not end once the ML application is deployed in production.

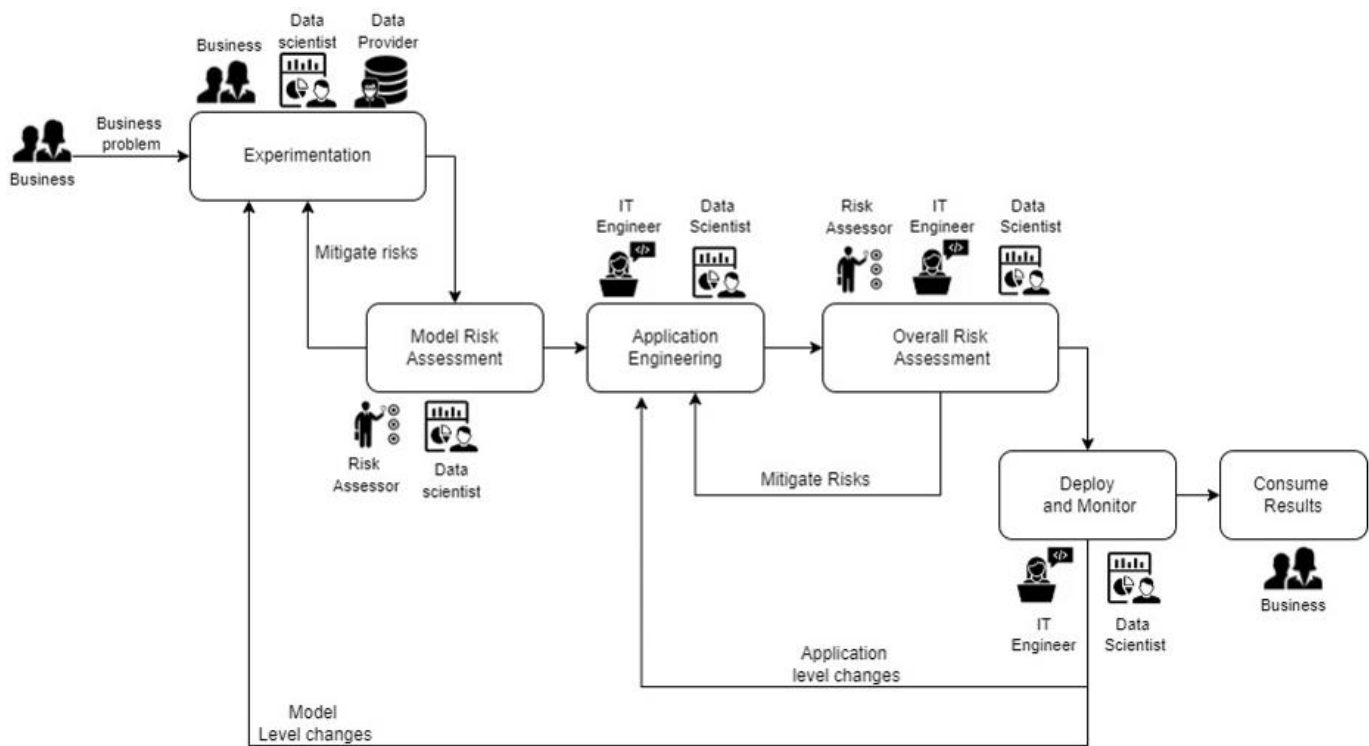


Figure 6: ML operationalization workflow

### 4.2.1 Experimentation

Unlike traditional software engineering, ML and data science require experimentation based on business domain understanding, data exploration, and developing and testing one or more causal inferences. The experimentation phase is initiated when a business problem is proposed by a particular business function to the data science team. The phase begins with understanding of the business problem, which often requires

close collaboration with the business teams and data science team ( E1, E3, E7) . This is followed with data collection, where data is extracted from the relevant sources and integrate for analysis. In order to collect data, the data scientist are dependent on the data providers. Especially the data owner who grants the permission to access data who have concerns. The data scientist must be able to convey the value of the business case they are working on and provide assurance on controlled usage and adherence to governance policies set on the data. Expert interview highlight that getting access to high quality and fit purpose data remains a key challenge in the banking industry owing to concerns with respect to the security and privacy of the data. Having collected the data, the data scientists explore the schema and distribution of data through exploratory data analysis. Data is then prepared for the ML task by performing data cleaning and feature engineering. In the next, step different algorithms and hyperparameters are used to train an ML model. The trained model is evaluated on a set of metrics to gauge its quality and performance. Depending on the results of the evaluation, data scientists and may either select the ML model or choose to reiterate with another experiment for better model performance (Salama & Kazmierczak, 2021).

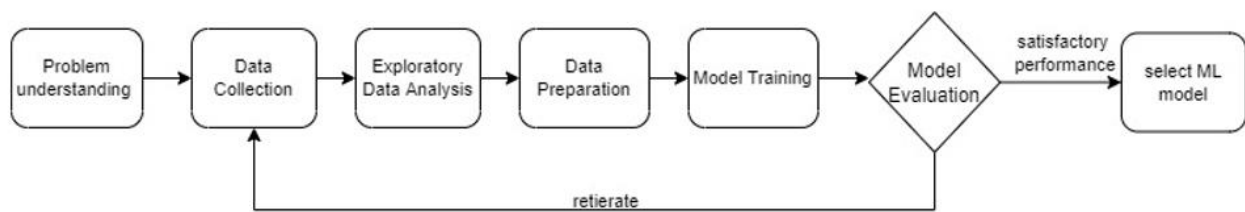


Figure 7: Experimentation

#### 4.2.2 Model risk assessment

Once a ML model is selected by the data scientist, it must go through a model risk assessment. Model risk assessment is a long standing practice in the financial service industry, and applies to all kinds quantitative models. The goal of this task is to ensure that any potential risks, due to errors in the development, implementation and use of internal models is identified and mitigated. In this phase the ML model is critically evaluated in terms of its design by members of risk assessment team. However, the tradition model risk assessment process needs to be revised in order to effectively evaluate ML models (Baquero et al., 2020; Haakman et al., 2021).

This can primarily attributed to the fact that ML models are inherently dynamic and need to updated regularly, and the lack of ML-specific expertise in the traditional model risk assessment teams. Based on the expert interviews, model risk assessment is performed by model validators who belong to the risk department and have strong ML expertise to evaluate the technical and functional components. The assessors review the underlying software code, and performance metrics along with the documentation describing the underlying assumption and design choices made by the data scientists during the development of the model ( E1, E13). In order to proceed with the further steps the data scientist must get an approval from the model risk validation team. It is important to note that the need and intensity of the model assessment may vary depending on the application use case. For example, model risk assessment may be very critical for applications such as fraud detection and credit risk management, but might be less intense or even absent in low-risk applications such as ML-based targeted marketing for retail banking.

#### 4.2.3 Application Engineering

During the experimentation phase, the data scientists develop a ML model by writing code in interactive notebooks. However, a production deployment requires that the model is packaged and integrated with other software components to create a ML application or service (Amershi et al., 2019; Zaharia et al., 2018) . Depending on the complexity and requirements of the ML service, this stage requires the data scientists

collaborate with various IT engineers for the integration of different capabilities. These roles and capabilities include:

- a. ML engineer: Provides software development capabilities for the ML models, example: developing a ML training pipeline or integrating ML model into an accessible microservice ( E4, E5, E7).
- b. Data engineer: Develops pipelines for ingesting and pre-processing vast amounts of data required by ML model (John et al., 2021; Ruf et al., 2021).
- c. Operations engineer: Provides capabilities for the ML platform, continuous integration, continuous deployment and monitoring of the ML application (Ruf et al., 2021).
- d. Technical Architect : Makes key design choices regarding the compute, network and storage for an integrated and end-to-end ML deployment (Treveil et al., 2020).

A data scientist's preference of tools, programming languages and platforms may differ from the preference of the engineering teams. Therefore, the integration of data science capabilities with engineering capabilities also requires that the data scientists and engineers make negotiated design choices to ensure the compatibility of the complete solution (E1- E6, E9, E12).

#### 4.2.4 Overall risk assessment

Once an integrated ML application is developed it must be assessed for new risks before it is introduced into a particular business function. These risks correspond to process risks, IT risks, regulatory risks, and ethical risks (Baquero et al., 2020; Wirtz et al., 2022). The assessments corresponding to these risks are elaborated as follows:

- a. Change risk assessment: This assessment investigates the process risks that may arise when the ML application is introduced in a particular business function. Operational risk may arise when the business users cannot effectively use the ML application for a number of reasons such as lack of explainability, lack of user knowledge and experience, incompatibility with the existing business rules (Wirtz et al., 2022).
- b. IT risk assessment: Primarily concern the risks and vulnerabilities relating to the IT systems on which ML applications are hosted. This may include aspect such as security, availability, scalability, etc. (E2, E4, E6, E12).
- c. Compliance assessment: The ML application is assessed for regulatory risks to ensure it complies with all the applicable laws and regulations. This may include a data protection impact assessment to ensure compliance with GDPR, along with additional regulations that may apply, for example a ML model used for anti-money laundering must comply with Wwft. Additionally, the processes of how the ML application is developed and deployed must ensure high levels of transparency to ensure it is auditability (Dwivedi et al., 2021; Wirtz et al., 2022) . Nearly all expert participants acknowledged the importance and presence of compliance assessment in the ML operationalization process.
- d. Ethical risk assessment: The ML application is also assessed for various ethical risks that could arise from its implementations. The expert participants highlighted a strong focus on fairness during the ethical risk assessment (E4, E13, E14) .

Experts highlight that the overall risk assessment of the ML models is one of the most time-consuming tasks during the operationalization process. The task requires that data scientists and engineers collaborate with different stakeholders from the risk department who in many cases do not have strong expertise in ML, as

result of which it often takes time to converge to a decision of how all the relevant risks can be identified, mitigated or accepted. The risk assessors focus on minimizing risks whereas the data scientists are focused on innovation and performance optimization. As a result the stakeholders are often met with conflicting viewpoints and preferences which they must balance to make a joint decision. The risk assessment teams act as gate-keepers and their approval is crucial for the data science and engineering teams to deploy the ML application to production.

#### 4.2.5 Deploy and Monitor

ML deployment is a critical stage in the ML operationalization process, where in the ML application must be deployed in a production environment to produce results for its intended purpose. Recently emerging MLOps paradigm introduced a set of practice in order to apply DevOps practices from software engineering into ML based applications. This essentially involves the adoption of continuous integration (CI) and continuous deployment (CD) pipelines (John et al., 2021; Ruf et al., 2021). As shown in Figure 8, this consist of a number of orchestrated tasks, starting with a CI pipeline which integrates, build and tests the code from the source code repository. CI plays an important role, in ensuring new code modules can be effectively integrated when new features are added or changes are made. When the CI pipeline is successful, it initiates a CD pipeline which deploys the ML pipeline. Mature production deployments require that the model is recreated through an automated ML pipeline to ensure consistency, and to facilitate retraining of the model. The ML pipeline serves as an end-to-end construct that orchestrates data extraction, validation, preparation, model training, and model evaluation. Given the evaluation of the model yields expected results, the model is registered. Registering the model ensures that a versioned catalog of ML models in maintained in production. The model is then deployed served or deployed in one of the following formats: 1) A microservice with a REST API to serve online predictions. 2) A model that executes on a batch dataset. The performance of the model is continuously monitored, based on which a new iteration in the ML workflow can be initiated to retrain or re-experimentation the model.

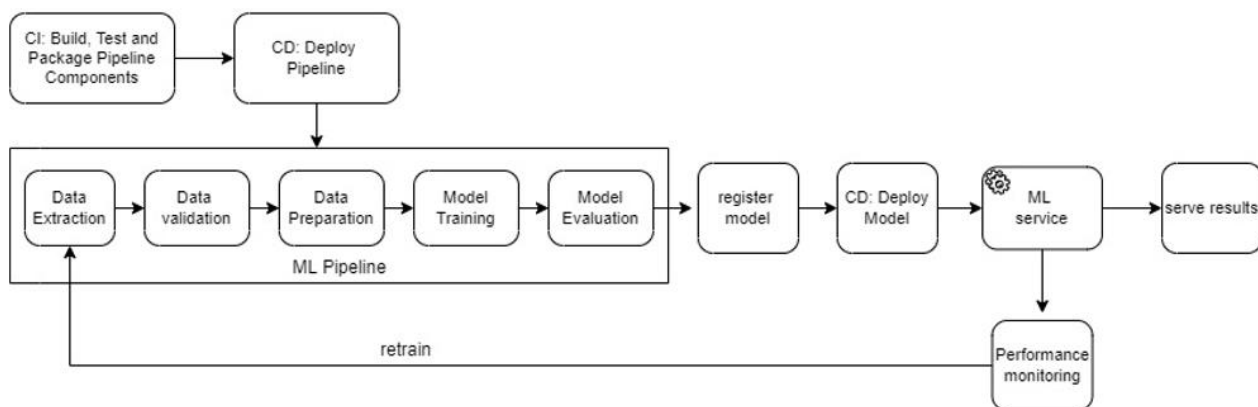


Figure 8: Deploy and Monitor

Providing operational support for deployed ML application can often be challenging. Post its deployment, the ML application may produce errors and unexpected results due to system dependencies (Schelter et al., 2018; Sculley et al., 2015). It may also require changes based on the feedback of the business user, or need to be retrained with new data, due to changing real-world data distribution. Further, unlike tradition software applications, ML applications are an amalgamation of the ML model, software code and data. This can often make it challenging to identify the source of the various problems that occur in production (E2, E4, E5). Additionally, organizations often face the dilemma of who should provide production support and how can the right person be involved to make a change in the deployed ML application (E5, E12). Figure 8 is relatively

mature orchestration workflow. The participants of interview showed varying level of maturity, with most still implementing or improving CI/CD capabilities in their ML deployment.

#### 4.2.6 Consumption of ML Results

For the ML application to deliver the desired business value, the results produced by the ML application should also be consumed by the intended business user for the intended purpose (E1, E4, E9). This often requires the results to be integrated with a user interface or a system that is accessible and familiar to business users. Further, the results produced by the ML application should not only be accurate but also explainable and easy to interpret for the user. ML as technology is new for the majority of the business functions in banks, as a result of which business users can show apprehensions and sometimes even resistance towards the adoption of ML applications (E2, E12, E13). To build trust, the data scientists need to work closely with the business users to educate them on how to use the ML application and interpret the results. Further, the business users can have new requirements and feedback which must be incorporated into the ML application after its initial deployment. Therefore, the data science teams need to provide continuous support to the business user. A number of experts highlighted, that providing post-deployment support is a major challenge since data scientists are innovation-driven stakeholders who prefer taking up new challenges and moving to new projects (E1, E5, E9, E10, E15).

### 4.3 Conclusion of the Socio-technical Analysis

This chapter aims to answer the first sub-research question: *What are the socio-technical complexities in the ML operationalization process?*

This chapter shows that ML experimentation is only of the many stages during the ML operationalization process. The process requires participation from diverse stakeholders having different interests, knowledge and preferences. At every stage different stakeholders that must collaborate with one another and make certain joint decisions. The workflow shows that the data scientists are the drivers of the process and are involved throughout the process. They are also the stakeholders that primarily specialize in ML, therefore it is crucial for them to translate and communicate the ML techniques in a way that is understood by other stakeholders who do not specialize in ML ( for example business user, risk experts, data providers). This can often be challenging owing to the novel and opaque nature of ML. Further, the stakeholders are often met with conflicting preferences which can lead to stalemates thereby delaying the process.

Specifically, with respect to technology, ML applications are inherently different from traditional software applications. They require a more inductive and experimental approach for model development. Further, the ML application is dependent on a number of components which includes training data set, ML configurations and, and application software. The inability to track of these artefacts and the possible dependencies and changes associated with them can add to the technical debt and complexity of the process.

Further, ML operationalization process is a cyclic process which does not end after the deployment of the ML application. Once deployed the ML application may require retraining and redevelopment owing to changes in real-world data distribution, technical bugs, or due to changes in business needs. This not only requires dynamically updating the ML application and its components, but also renegotiating certain decisions, if required.

# 5 Socio-technical Factors

This chapter identifies the socio-technical factors that influence the ML operationalization process. First, the socio-technical factors and sub-factors are identified along with the empirical evidence relating to both literature and expert interviews. This is followed by a network relationship diagram based on the expert interviews to illustrate how the factors are related to one another based.

## 5.1 Socio-technical factors influencing ML operationalization

The analysis of academic literature and the expert interviews, lead to a comprehensive list of socio-technical factors. The socio-technical factors represent socio-technical subsystems, within the larger socio-technical system in which the ML operationalization process takes place. This implies that each factor in some way is associated with ML and the tasks during the ML operationalization process (technical system), and is dependent on or requires the involvement of certain stakeholders (social system).

Each of the identified factors consist of granular subfactors. This structure of factors and sub-factors can be attributed to two primary reasons. First, much of the literature study provided broader concepts and problem areas, which needed to be further refined and made specific to ML operationalization and the banking industry. Secondly, the qualitative analysis of expert interviews required frequent abstraction and splitting of concepts in order to develop a comprehensive and coherent encapsulation of the ML operationalization process. The factors are categorized into three categories which include the drivers, challenges, and the successful outcomes which are organized in Table 8, Table 9, and Table 10 respectively. A detailed description of all the factors can be found in Appendix D.

### 5.1.1 Successful Outcomes

The expert participants of the study were asked how they perceive the success of the ML operationalization process. Based on this three factors were identified through qualitative analysis. The significance of the factors was further understood by exploring these factors in literature. While a number of literature sources do not explicitly regard the factors as successful outcomes, they aimed to achieve or optimize these factors. *Unlike the other categories i.e. challenges and drivers, the successful outcomes do not consist of sub-factors. This is because, the success outcomes were derived through a funneling approach during the expert interviews. Therefore, they led to more specific factors and not broader concepts.*

Successful Outcomes		
Factor	Literature Source	Expert
1. Business User Adoption	(Chowdhury et al., 2022) (Cabitza et al., 2020) (Guo et al., 2022) (Brock & von Wangenheim, 2019)	E1, E3, E4, E5, E9, E11, E12, E13, E14.
2. ML Decision Quality	(Chowdhury et al., 2022) (Lwakatare et al., 2020) (Ruf et al., 2021)(Granlund et al., 2021) (Schelter et al., 2018) (Sculley et al., 2015)	E1, E2, E3, E4, E5, E7, E9, E11, E12, E13, E14.
3. Compliance and Auditability	(Muthusamy et al., 2018) (Karimi & Yahyazade, 2021) (Haakman et al., 2021) (Wirtz et al., 2022) (de Bruijn et al., 2021)	E1, E2, E3, E4, E5, E6, E7, E9, E10, E11, E12, E13, E14.

Table 8: Socio-technical factors - Successful outcomes

### 5.1.2 Challenges

This category represents factors that are essential to the ML operationalization process, but are challenging for organization to realize or implement. These factors indicate the areas in which organization need improve in order to ensure the success of the ML operationalization process.

<b>Challenges</b>		
<b>Factor</b>	<b>Literature Source</b>	<b>Experts</b>
<b>4. Data Quality</b>		
Data accuracy	Renggli et al., 2021, Kaddoumi & Tambo, 2022, Lwakatare et al., 2020), Cabitza et al., 2020, Janssen et al., 2017, Keller & Staelin, 1987)	E1, E12, E5, E6, E4, E13
Data availability and accessibility	(Sebastian-Coleman, 2013) (Ruf et al., 2021) (Renggli et al., 2021) (Kaddoumi & Tambo, 2022) (Lwakatare et al., 2020)	E2, E5, E6, E9, E12, E10, E11, E13
Data Consistency	(Renggli et al., 2021) (Kaddoumi & Tambo, 2022) (Lwakatare et al., 2020) (Passi & Jackson, 2018)	E1, E2, E4, E5, E9, E13,
Data Completeness	(Renggli et al., 2021) (Janssen, Brous, et al., 2020; Ruf et al., 2021) (Miceli et al., 2022) (Passi & Jackson, 2018)	E3, E4, E13
<b>5. Adaptability</b>		
Adapting to system dependencies	(Lwakatare et al., 2020)(Sculley et al., 2015), (Schelter et al., 2018) (Yokoyama, 2019)	E1, E9, E13, E7, E15
Adapting to business needs	(Guo et al., 2022), (Q. Z. Chen et al., 2022), (Omidvar et al., 2022), (Borg, 2022)	E1, E2, E9,
Adapting to concept drift	(Tsymbal, 2004) (Žliobaitė et al., 2016) (Bourgais & Ibnouhsein, 2021) (Renggli et al., 2021)	E1, E4, E6, E9, E15
<b>6. Enterprise Integration</b>		
Integration with existing systems	(Kaddoumi & Tambo, 2022)(Snoeck et al., 2020) (Kuguoglu et al., 2021) (Brock & von Wangenheim, 2019) (Schlögl et al., 2019)	E1, E4, E9, E10, E13,
Integration with existing process	(S. M. Miller, 2018), (Wamba-Taguimdje et al., 2020), (Olan et al., 2022), (Kerzel, 2021)	E1, E4, E9, E10, E14
<b>7. Risk Management</b>		
Managing Operational Risk ( model risk, process risk, IT, regulatory and compliance)	(Muthusamy et al., 2018) (Karimi & Yahyazade, 2021) (Haakman et al., 2021) (Wirtz et al., 2022)	E1, E2, E3, E4, E5, E13, E15
Managing Ethical and Reputational risks	(Haakman et al., 2021) (Wirtz et al., 2022) (de Bruijn et al., 2021)	E2, E3, E5, E12, E15

Table 9: Socio-technical factors- Challenges

### 5.1.3 Drivers

The factors in this category, can be understood as enabler of the ML operationalization process that essentially bolster or control the process. The drivers often deliver value indirectly, however they have been increasingly be stressed on by multiple expert participants.

Drivers		
<b>8. Shared Knowledge</b>		
Cross-functional knowledge	(Brock & von Wangenheim, 2019)(Ruf et al., 2021) (de Bruijn et al., 2021) (John et al., 2021) (Karamitsos et al., 2020),	E1, E3, E4, E5, E9, E7, E12, E15
Negotiated Knowledge	(Passi & Jackson, 2018), (de Bruijn & ten Heuvelhof, 2018) (de Bruijn et al., 2021)	E2, E3, E4, E5, E6, E12, E14, E15
Collaboration	(Sambasivan et al., 2021) (Schelter et al., 2018) (Spjuth et al., 2021) (Sculley et al., 2015) (Hummer et al., 2019) (Zaharia et al., 2018)	E2, E3, E4, E5, E9, E12, E13
Communication and Translation	(Cabitza et al., 2020) (Passi & Jackson, 2018) (John et al., 2021) (Hummer et al., 2019) (Lebcir et al., 2021) (Baier et al., 2019) (van den Heuvel & Tamburri, 2020) (Fountaine et al., 2019)	E1, E2, E3, E4, E5, E6, E7, E9, E10, E13, E14, E15
<b>9. Controls</b>		
Monitoring	(Haakman et al., 2021) (Lwakatare et al., 2020) (Ruf et al., 2021)(Granlund et al., 2021) (Schelter et al., 2018) (Sculley et al., 2015) (John et al., 2021) (Zaharia et al., 2018) (Hummer et al., 2019) (Wirtz et al., 2022)	E1, E6, E7, E9, E12
Traceability and Lineage of Artefacts	(Sculley et al., 2015) (Schelter et al., 2018) (Zaharia et al., 2018) (Hummer et al., 2019) (John et al., 2021) (Ruf et al., 2021)	E2, E3, E4, E6, E9, E12, E15
Testing and Quality Assurance	(Garcia et al., 2018), (Sculley et al., 2015) (Schelter et al., 2018) (Zaharia et al., 2018), (Hummer et al., 2019), (John et al., 2021) (Chowdhury et al., 2022)	E1, E2, E6, E9, E15
Model Explainability and Fairness	(Chou et al., 2022), (Bach et al., 2015), (Adadi & Berrada, 2018), (Mehrabi et al., 2021) (Enni & Herrie, 2021) (Turner Lee, 2018)	E1, E3, E9, E12, E13, E15
Roles and Responsibilities	(de Bruijn et al., 2021), (Hummer et al., 2019) (van den Heuvel & Tamburri, 2020) (Janssen et al., 2020) (Toreini et al., 2020) (Asatiani et al., 2021) (Robbins, 2020) (Wirtz et al., 2022)(Preece et al., 2018)	E2, E5, E6, E9, E12

Table 10: Socio-technical factors - Drivers

## 5.2 Inter-relationships between socio-technical factors

Based on the expert interviews, the relationship between the socio-technical factors are identified. These relationship are represented in Figure 9, which is developed in ATLAS.ti using the graphical network manager. The figure can be interpreted as a network of nodes and links. Each node represents a qualitative code corresponding to a socio-technical factor. Nodes are connected to one another with relational links that represent the relationship between the factors.

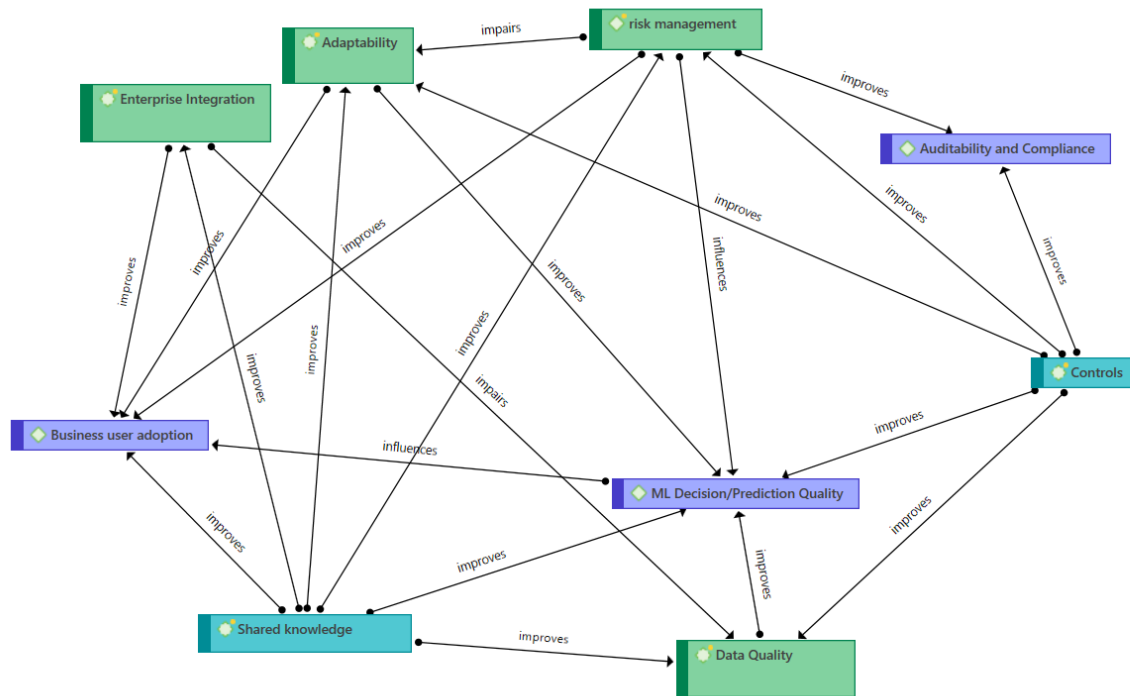


Figure 9: Network representing the relationship between socio-technical factors

The socio-technical factors in the category of drivers, challenges and successful outcomes are shown in turquoise, green and purple respectively. To best represent the relationship, four types of relational links are used: ‘improve(s)’, ‘impair(s)’, and ‘influence(s)’. The relations ‘improve’ and ‘impair’ are simple to comprehend, and can be directly interpreted as positive and negative relationships respectively. However, relations represented by ‘influence’ are relatively complex. A ‘influence’ relation has a dual nature, and indicates that a particular factor can both improve and impair another factor. The impair relationship was found in two cases: between risk management and ML decision quality, and between ML decision quality and business user adoption.

In the first case, risk management was found to both improve decision quality or impair it depending on the type of risk management and how it was applied to the ML operationalization process. For example, model risk validation can ensure soundness and robustness of the model, thereby improving the ML decision quality. However, strict risk measures that constraint the usage of data and restrict complex modeling and engineering techniques can impair the ML decision quality. In the second case, decision quality was found to improve business user adoption in many cases. Accurate results increase the confidence in the business user thereby aiding them to perform their business function in a better way. However, in some cases, accurate decisions may not coincide with the intuitive beliefs of the business user. In such cases, even if the model is highly accurate, but too complex for the business user to effectively explain and interpret the results, the business users may resist the adopt the ML application.

### 5.3 Conclusion of the Socio-technical factors

This chapter aims to answer the second sub-research question: *What socio-technical factors influence the ML operationalization process?*

Based on literature and expert interviews this chapter identifies nine socio-technical factors that influence the ML operationalization process. These factors can be summarized as: business user adoption, machine learning decision quality, and compliance and audibility which are categorized as successful outcomes; data quality, adaptability, enterprise integration, and risk management which are categorized as challenges; shared knowledge, and controls which are categorized as drivers. To develop a concrete understanding, each factor is composed of more granular subfactors. While a number of these factors are acknowledged and investigated in the literature, the knowledge is vastly fragmented. One of the key contributions of this chapter is to integrate this fragmented knowledge to develop a comprehensive understanding of the ML operationalization process.

Depending on how the factors influence the ML operationalization process, the factors are organized into three categories which include drivers, challenges, and successful outcomes. Based on the expert interviews, the relationship between the socio-technical factors is presented through a network relationship diagram which diagram shows a dense inter-relationships between the factors. Based on this certain factors can improve or impair the other. More complex relationships also exist where a factor may both impair and improve another factor under certain conditions. While inter-relationships identified from the interviews provides an extensive understanding of the factors, they also makes it challenging to develop a parsimonious model directly from the findings.

# 6 Case Study Analysis: Anti Money Laundering

The case study analyzes a concrete ML operationalization business case in the banking industry. The purpose is to develop a pragmatic understanding of how the identified socio-technical factors play out in practice. Further, it also allows for triangulation and validation of the factors. Based on an interview with an industry expert, the case study examines a prominent use case where a large Dutch bank has operationalized an ML application to detect fraudulent transactions as a part of their anti-money laundering initiative.

## 6.1 Case Background

Money laundering refers to the illegal process of concealing money generated from criminal activities such as drug trafficking or terrorist funding in a way such that, it appears to have been generated from a legitimate source. Since money laundering imposes a significant threat to the global economy and its security, governments and central banks have introduced a number of regulations that enforce banks to monitor and report suspicious transactions of their customers. In the past, failure to comply with such regulations has cost banks hefty fines ranging from \$350million to \$1.2bn in Europe and the Americas (Z. Chen et al., 2018). Specifically in the Netherlands, all Dutch banks must comply with the *witwassen en financieren van terrorisme* (Wwft) or the money laundering and terrorist financing prevention act. In doing so, banks must monitor the transactions of their clients and notify the Financial Intelligence Unit – of the Netherlands (FIU) of the unusual transactions.

A large Dutch bank is using ML techniques for detecting anomalies in transactions and flagging suspicious behaviour of customers. The bank traditionally used a rule-based system to detect suspicious transactions. However, rule-based systems have a relatively static nature and cannot detect new patterns in transactions. This gives criminals the opportunity to adapt their activities and bypass the existing rules. Additionally, with high volumes of the transaction, rule-based systems generate a large number of false positives that lead to over-reporting of suspicious behaviour, which can be expensive and time-consuming to filter through (Han et al., 2020; Jullum et al., 2020). The organization believes that the use of ML techniques can detect new and emerging patterns in financial transactions and augment the understanding of financial crime analysts on how criminals are laundering money.

The data scientists have operationalized an ML application that has been running in production for two years. While the team has succeeded on many fronts, there still remain a number of challenges that need to be overcome. The team's prior experience in operationalizing the ML application and the existing challenges provide the research with pragmatic insights regarding the ML operationalization process. In order to conduct the case study, a data scientist team lead working in the transaction monitoring domain is interviewed. The insight provided by the interviewee delves into both the social and technical aspects of the ML operationalization process. *All the quotations made in this chapter correspond only to the data scientist interviewed for the case study.* The results of the case study are organized into four sections which elaborate on the factors relating to the successful outcomes, challenges, and drivers.

## 6.2 The Successful Outcomes: Decision Quality, Business User Adoption, and Compliance.

### ML Decision Quality:

The quality of the decision made by the ML model is extremely important for a critical use case like anti-money laundering. The ML model must accurately flag the anomalies in a transaction, that can be suspicious. Not being able to detect fraudulent activities (false negatives) could not only result in hefty fines for the bank but also threaten the financial stability and the overall social well-being at the national and international levels. Whereas inaccurately detecting unsuspicious activities as suspicious (false positives) is expensive, time-consuming, and detrimental to the reputation of the bank.

### Business User Adoption:

The most accurate ML models would deliver no business value if not used by the financial crime analysts who are responsible for examining suspicious transactions and reporting them to governmental institutions. It is also worth highlighting that the bank does not intend to fully automate the detection of suspicious transactions. Instead, ML techniques are used to assist the analyst in improving the quality and efficiency of the reporting system. Therefore, adoption by the business users i.e. the financial crime analysts in this case is imperative for the success of operationalizing the ML application.

### Compliance:

Anti-money laundering is a mandatory preventive measure that financial institutions must take in order to comply with the laws and regulations. The outcomes of the preventive measures are closely monitored by supervisory bodies such as the DNB and AFM. The application must comply with two regulations namely wwft and GDPR. Wwft provides a set of rules and obligations corresponding, to customer due diligence, transaction monitoring, and reporting in order to prevent financial crime. Like any other data-driven application the ML application and operationalization process must comply with GDPR and ensure that the personal data and privacy rights of the customers are protected. Compliance serves as a hard requirement for the success of the ML operationalization process.

## 6.3 The Challenges: Data Quality, Adaptability, Enterprise Integration, and Risk Management

### Data Quality:

Data Quality is highlighted as a prominent factor in the operationalization of ML models. Data quality issues continue to be a challenge for the teams working on operationalizing the ML model. However, the teams have come a long way in solving these challenges and improving the quality of the data used in the operationalization process.

The data on which the ML model is dependent is stored in the legacy systems of the bank. During the experimentation phase, the data scientists encountered several data quality issues relating to inaccurate values, incomplete fields, and incomplete data sets. While some of the issues were detected and resolved during the development of the model, some issues were undetected until the model was deployed in production.

*Data Scientist ( Team Lead): “Over these two years, we frequently had the analysts ( business users) come to us with numerous examples showing that they got alerts that they didn't understand. So they requested us to investigate, those alerts further. It turned out that it was the data quality issues that were leading to these false alerts.”*

In the initial stages, when the model was operationalized in production, the data quality issues lead to a number of false alerts being generated. The business users reported the unexpected alert which lead to the detection of certain additional data quality issues. Additionally, this also got the model validation team from the risk department involved. The model validation team along with the data science have been currently working on

a data validation framework to address the data quality issues. The data validation framework serves as a control measure that ensures that the data quality issues are detected, reported, and prevented before they impact the ML model. The control measures and the synergy between business, risk, and data science teams also validate the importance of communication and collaboration and shared knowledge to reduce data quality issues.

### **Enterprise Integration:**

As previously indicated the ML application required a technical integration with the legacy systems for the data sources, which also introduced a number of data quality issues. However, integration with other systems also contributed positively to business user adoption. ML application had to be integrated with the existing user interface which the analyst used for viewing results. This required the ML results to be sent to the user interface with the right fields of data and in the right format.

*Data Scientist (Team Lead): “ We had to send our alerts to a UI tool the analysts were using, but that the required of course, like a technical integration, and making sure that we send the information within the right format with the right required fields such that it is compatible with the UI tool”.*

In addition to the technical integration, the ML application had to be integrated into the existing business process. The results generated by the ML model are not used in isolation, they must be combined with the result of the rule-based systems based on particular business logic. ML-generated results were also new and relatively opaque compared to traditional rule-based systems for business users. The data science team had to therefore guide the business users on how to interpret the ML results, and provide them with information on why the model would consider a particular transaction suspicious. This involved conducting workshops and regular review meetings to address the questions and concerns of the analysts.

*Data Scientist (Team Lead): “We generate multiple signals, which need to be combined based on particular business logic. Since machine learning is new for the analyst we also had to prepare the analysts by creating working instructions. Explainability was an important topic because these, machine learning models are mostly a black box. So you need to help the analysts and provide them with the right information so they know why did the model consider a transaction suspicious? and How do I need to investigate this?”*

### **Adaptability:**

After the initial deployment, the data science teams have had to adapt to and accommodate different changes in their ML solution. The development of the ML model was initiated before the covid-19 crisis and therefore the model was trained on pre-pandemic historical data. This led to a scenario of concept drift, where the covid-19 crisis lead to a change in the transaction behavior of the customers, as a result of which the ML model was no longer relevant with the current trends. Further, the application also had to adapt to changing business requirements. The ML model is built on certain risk indicators specified by the financial action task force (FATF), based on which a transaction is considered to be suspicious. However, the risk indicators themselves can change. For example one of the risk indicators involves a list of high-risk countries, which is subject to change. Further, the data being used by the model was also processed by other source systems, as a result of this, errors in the processing of the source system also affected the model.

*Data Scientist (Team Lead): “The system used by the analyst at one point in time changed the way it processed data, and that affected how we were retrieving the labels. For example, they started identifying certain types of transactions ‘cash’, which were earlier considered ‘wire’ transactions. So that also had an impact on the model and we had to adapt to it.”*

In order to deal with the changing transaction patterns, controls such as model monitoring have played an instrumental role. The data science team closely monitors the results of the model and compares them with the historical trends of the results. In addition to this, the team maintains versions and lineage to detect, what data

was a version of the model trained. This allows them to detect changes in feature distribution. Based on this, the data scientists take a call if the model needs to be re-trained or redeveloped in some way. However, the decision is a joint decision made by both the data science team and stakeholders from the business, since it requires both ML and business domain expertise. When the changes are made to the model, data scientists create documentation of the changes and submit it to the model risk validators for review.

### **Risk Management:**

To manage the risk the bank has three lines of defense. The first line of defense consists of the developers of the ML application. The second line of defense consists of independent risk assessment teams that validate the ML application through a number of risk assessments, and the third line of defense consists of internal audits, which is a broader organizational process that audits all the processes and systems within the bank. The risk assessment processes form an important component of the ML operationalization process. The expert interviewee highlighted the following risk assessments: model risk validation, change risk assessment, and data protection impact assessment. This model risk validation assesses the risks related to the soundness and robustness of the ML model. The model risk validation is done by an independent model validation team from the risk department which has a strong technical background, and expertise in ML and data science. The change risk assessment assesses the process risks which include the risk of ML application being introduced into a business process. This can entail the risk of analysts not being able to effectively use the results generated by the model for reasons such as explainability and interpretability, thereby resulting in an inefficient and ineffective process.

The data protection impact assessment was conducted to cover aspects relating to regulatory risks. The operationalized ML application must comply with both GDPR and wwft. Compliance with GDPR requires that the use of sensitive personal data points is restricted to train the ML model. However, the expert interviewee mentions that this makes the development of the ML model with high accuracy extremely challenging which can risks compliance with wwft.

*Data Scientist (Team Lead): “So in principle, we have two regulations to comply to. One is the wwft, which obliges us to report suspicious transactions. And on the other hand, we have like the GDPR that requires us to use as little personal data as possible. And these two, these two legislations are basically conflicting. So you need to find the balance that that is good from both perspectives and that is done in the data protection assessment.”*

Finally, the ML application can also impose ethical risks associated with the bias and fairness of the decision made by the model. This also relates to the regulatory risks, and the restricted use of personal and sensitive data. However, bias may still creep into the ML models indirectly, for example, innocuous data points such as postal code, employment, etc. may serve as a proxy for restricted fields such as ethnicity. Therefore, dealing with ethical risk requires extra caution. In the initial deployments, the ethical risks were managed during the data protection impact assessment. Recently, the bank has introduced ethical checks during the experimentation phase, to ensure any ethical issues are flagged early in the operationalization process. The risk assessments require the data scientist to collaborate and negotiate on a number of key decisions. The key challenge faced by the data scientists during the risk assessment was the lack of ML and data science knowledge among the risk teams ( except model validation teams) which resulted in a long and cumbersome assessment process. The risk assessment and data science teams often end up in repeated conversations before they converged on a shared understanding.

*Data Scientist (Team Lead): “We find ourselves often going back to the basics of machine learning. Because those teams, of course, don't have a technical background. And then I guess they're not supposed to remember all of this, but. Yeah, takes a lot of time for us to go through these processes because we need to do a lot of explanations.”*

## 6.4 The Drivers: Controls and Shared knowledge

### Controls

The case study involves a highly regulated business application and therefore exhibits a relatively high maturity in implemented controls. Control measures such as monitoring have enabled data science teams to continuously evaluate the performance of the ML model and adapt to changing data distributions. Additionally, these also serve as mitigation measures to overcome model and process risks. In order to keep the track of the deployments, the data science team is maintaining a version of the deployed model. This enables the team to maintain a history of the deployments and track the performance of the model over multiple versions.

*Data Scientist (Team Lead): “We monitor the performance of the models very closely. Every month we have performance reports that we generate. So in this specific use case, we check how many alerts we generated and how do they compare to the historical trend. We also track the data, and measure the changes in the feature distribution as an additional check. So there are various angles by which we monitor what is happening.”*

The case study interview highlights the importance of versioning and creating a lineage of the artifacts in the ML operationalization process. The data science team versions both the deployed ML models, and the data set on which the model is trained on. In addition to this, all the experiments done by the data scientist are tracked. This allows the data scientists to keep a list of ML algorithms, hyper-parameters, and the corresponding performance in the experiment phase. This plays an important role in facilitating an end-to-end lineage of artifacts and addressing problems in production. When a particular problem is detected in the ML application, the data scientists can trace which model version is deployed in production, on which data set is the model trained, and the experiment that it corresponds to. This aids the team to reproduce the problems in their development environment thereby enabling the data scientists to understand the root cause of the problem and respond to production issues in a timely and systematic manner. Further, versioning and lineage also serve as measures to create audit trails, which can enhance the transparency and auditability of the ML operationalization process. Another control measure is the data validation framework which is being developed through the joined collaboration between the data science team and model validation team. This can be a form of testing and quality assurance that helps to minimize the impact of data quality issues. While not in place yet, data science is further trying to develop automated tests in CI/CD.

### Shared Knowledge

The expert participant has repeatedly highlighted the importance of shared knowledge between diverse stakeholders throughout the ML operationalization process. This primarily includes data scientists, engineering teams ( ML engineers and Ops Engineers ), risk teams, and business teams. To develop a shared understanding and strengthen communication, the data science teams have conducted a number of workshops and knowledge-sharing sessions with the business users and with the non-technical risk assessment teams. This is complemented by regular engagement with business users to support them in the adoption of the ML application.

A number of instances highlighted previously confirm the value of these efforts. For example, shared knowledge between the data science and business teams has helped in detecting data quality issues and changing data distribution. Whereas shared knowledge between the data science and risk assessment teams has been important in the joined development of the data validation framework which can potentially reduce the data quality issues. This corroborates the importance of shared knowledge which is not merely sharing facts and information, it is built on the concept of developing shared values and synergy between diverse stakeholders. The expert participants elaborated that the technical teams are organized as separate data science, engineering, and operational teams. In its current effort, the data science team is aiming to strengthen its collaboration with the ML engineers in order to develop standard CI/CD pipelines that test, build and deploy ML applications in production throughout its lifecycle.

However, at the same time, data scientists still face challenges in establishing shared knowledge among risk teams, other than the model validation team. The effect of this can be seen in the repetitive and arduous nature of the risk assessment.

## 6.5 Conclusion of the Case Study

The main aim of the case study is to validate the identified socio-technical factors and relationships and to develop an understanding of how the factors apply to ML operationalization in practice. With this, the case study adds rigor and relevance to the second sub-research question.

The case study also helps to understand each factor as a socio-technical subsystem, that consist of technology, tasks, and interaction between stakeholders. In addition to this, certain other aspects, which haven't been captured in the previous chapter are highlighted in the case study. Shared knowledge is an important factor during ML operationalization. But developing shared knowledge during the ML operationalization process remains a complex task. Despite, the data science team's attempts to conduct workshops with the non-technical risk teams, the risk assessment process remains long and cumbersome, which indicates that the shared knowledge between the data science team and the non-technical risk assessment needs more maturity. The use case also shows how the stakeholders are required to innovate in a highly regulated environment. During the risk management process, the stakeholders are confronted with opposing relationships or trade-offs relating to data privacy and ML performance, both of which are required to comply with the regulatory obligations.

Further, the case study also highlights some hidden relationships which were not explicitly captured during the qualitative data analysis of expert interviews. Integration with existing processes can introduce process-related risks, thereby impairing risk management. Secondly, integration with existing systems can introduce system dependencies which can impair adaptability.

## 6.6 Intermezzo: Narrowing the focus

The findings from the literature, expert interviews, provide a comprehensive list of factors and their relationships. At the same time, such a comprehensive analysis makes it challenging to conduct an in-depth investigation into all the factors within the constraints of the time. Further, keeping the number of factors, sub-factors and relationships in mind, it is especially challenging to develop a conceptual model that is parsimonious and useful for researchers to conduct future research. Therefore, a choice was made to center the focus on a single factor in the next steps of the research. The factor selected was risk management. The choice can be attributed to a number of interesting findings:

1. Given the highly regulated environment and the high stakes involved, risk management assumes high priority and strategic importance in the financial service industry. ML brings new and unfamiliar risks which can be difficult to identify and mitigate.
2. Risk management has been a re-occurring in the qualitative analysis of the expert interviews. It has also been prominently highlighted during the case-study. The experts highlighted risk assessment as a challenging phases which is excessively cumbersome and time-consuming.
3. Both the expert interviews and literature study indicate a predominantly algorithm-centric approach to risk management. This indicates broader system and process-related risks of operationalizing ML have often not been investigated in detail.

Based on this it was concluded that investigating risk management and addressing its challenges can support the ML operationalization process, thereby contributing to the main research question.

# 7 Risk Management In ML Operationalization

One of the primary concerns with respect to the use of ML in banks is the possible financial and reputational risk it entails. Some of the high-profile cases which exemplify the consequences of risks include Knight capitals algorithm trading glitch which led to a loss of ~450 million dollars putting it on the verge of bankruptcy in 2012 (Koshiyama et al., 2021), and the recent news of Budapest bank being fined 250 million euros for a GDPR violation with respect to one of its ML applications.

Risk management refers to the structured process of identifying, classifying, and quantifying, risks and then mitigating and controlling them (Srinivas, 2019). While risk management is a long-standing practice in the banking sector, the ML operationalization process can involve new and varying risks that are not easy to manage by traditional risk management functions. As shown in Chapter 5, the risk management factors consist of managing operational risks and managing ethical and reputational risks. Operational risk consists of managing model risks, regulatory risks, process risks, and IT and Data risks. The explanation of these risks can be found in Appendix D.5. Most banks use three lines of defense model, where the risks should be managed by the developers and owners of the ML application (first line), risk assessments conducted by risk experts (second line), and through internal audit (third line).

This chapter starts by discussing technological aspects of ML that make risk management a challenge during the operationalization process. In addition to this, interactions, translations, and negotiations between the stakeholders are explained to understand the social challenges of risk management. Four strategic guidelines are proposed to help banks adapt their risk functions to address the socio-technical challenges of ML operationalization. The chapter concludes with a conceptual model that captures the essence of the risk management process during ML operationalization.

## 7.1 Technical challenges of Risk Management

### 7.1.1 Model Opaqueness and Biases

Traditionally the risks associated with algorithmic decision-making have focused on accuracy and computational costs, however, the use of ML techniques introduces new sources of risks (Koshiyama et al., 2021). To understand the risks associated with the model, it is important to understand the internal functioning of ML algorithms. Widely discussed in the literature is the opaque or black-box nature of complex ML algorithms, which makes it difficult to clearly explain how ML algorithms make a particular decision or prediction (Burrell, 2016; Janssen et al., 2020). The tight coupling between large-scale data and complex ML algorithms can make it difficult even for data scientists and engineers to clearly understand the working of ML applications (Passi & Jackson, 2018). Thereby making it even more challenging to explain the working during risk assessments to stakeholders who lack specialized knowledge in ML and data science. Another aspect of the technological challenge is the bias that can get embedded in ML algorithms when trained on skewed training data set (Wirtz et al., 2022). Not only can biased models lead to inaccurate decisions, they can also affect the social well-being of the society and damage the reputation of the bank.

### 7.1.2 Risks in the Operationalization process

When viewed from the perspective of the ML operationalization process, the risks may lie beyond the ML algorithm. An ML application is composed of a number of components which include datasets, ML configurations, and application software (Zhou et al., 2020). These components are tightly coupled with one

another and rarely static. Depending on (re)experimentation and re-training activities, components can change frequently and asynchronously during the life cycle of the ML application. This makes it crucial to have a transparent overview of the different components, and their inter-linkages to stay in control of the operationalization process. Referring back to the case of Knight capital, the glitch did not precisely lie in the algorithm, but in the wrong version of software being deployed to production. Further, when post deployments issues occur in production, it is often challenging to trace the components of ML application to reproduce the results (Zaharia et al., 2018) and the errors. Therefore, the risk may not only reside in the ML algorithm but also in how the components of the ML application and their dependencies are managed during the operationalization of ML.

*Product Owner (ML Platform): “We made assumptions about the data that turned out to be incorrect and they were also relatively difficult to check. And therefore the whole application wasn't behaving the way that we were supposed to think. It was also difficult to find out. Having a strange behavior in the data can cause problems things throughout the application. And you don't really know if it's an application bug or a data problem.”*

Unlike a number of high-tech organizations, most banks currently have a small number of ML applications in production. As banks aim to scale their AI initiative throughout the enterprise, a large number of ML applications running in production can further add to the complexity of managing the process-related risks. In addition to the existing inter-linkages, the technical stakeholders may also reuse (transfer learning) or combine ML models (ensemble learning) to create new ML applications. The output produced by some ML applications may also be consumed by other ML applications. Practitioners argue that in such situations ML systems can exhibit complex entanglements, making it hard to create strict abstraction boundaries between them (Sculley et al., 2015). As a result of which changes and problems in one application or its component can have a cascading effect on other applications.

## 7.2 Risk Management in a network of Stakeholders

ML risk management takes place in a network of diverse stakeholders. In such a network every stakeholder (1) has different goals and interests, (2) provides different resources, (3) and depends on other stakeholders for the realization of their goals (de Bruijn & ten Heuvelhof, 2018). The stakeholders must also negotiate with one another to make joint decisions and key trade-offs. This network is activated during the risk assessment of the ML application and consists of three primary stakeholder groups: Data scientists, IT engineers, and Risk teams. Analysis of the stakeholder network provides deep insights into the social nuances and affective relationships which sets the logical context for understanding the social challenges.

### 7.2.1 Diversity of Stakeholders

Based on their knowledge, interests, and goals stakeholders develop a certain cognitive state which guides their behavior and actions (Grover & Lyytinen, 2015) during the risk assessment process. In this subsection, the different characteristics of stakeholders and their effect on risk management are analyzed.

The stakeholders in the risk group include model risk validators who specialize in evaluating all kinds of quantitative and financial models that are deployed in the bank. The stakeholders in the risk group also include privacy, compliance, and ethics officers who specialize in the understanding of different laws, regulations, and ethical standards. Change risk managers are also part of the group, and specialize in change management and analyzing process-related risks. Another important stakeholder is members of the information security department who have a specialized understanding of different information risks. The key interest of this stakeholder group is to ensure that all risks associated with the ML application are minimized and serve the primary goal of safeguarding the bank against financial and reputational losses. As a result of this stakeholder,

the group has a cognitive state which is predominantly risk-averse. Based on this, the risk assessment has a preference for approaches and design choices that have the lowest possible risks.




Stakeholder group	Risk 	Data science 	IT engineering 
Knowledge/ Expertise	Laws, regulations, Information security, Risk identification & analysis	Data Science, Machine Learning	Data and ML engineering
Primary interests	Minimization of risks	Timely deployment and model performance	Timely deployment and Engineering performance
Primary Goal	Prevention of financial and reputational loss	Approval for production deployment	Approval for production deployment
Cognitive State	Risk averse	Innovation driven	Innovation driven

Figure 10: Diversity of Stakeholders based on knowledge, interests, goals and cognitive states

The data scientists and IT engineers on the other hand have deep knowledge of the technology. The data scientist possesses an in-depth understanding of data preparation, feature engineering, and ML modeling techniques. Whereas the IT engineers have strong knowledge about the data pipelines, software components, CI/CD pipelines, and IT platforms on which the ML model is operationalized (Ruf et al., 2021).

The primary interest of data scientists and IT engineers lies in a lean risk assessment and the timely deployment of the ML. Further, the data scientists may be more interested in the performance metrics of the ML model such as accuracy, receiver operating characteristic (ROC) curve, precision, and recall. and the engineers may have a greater interest in the performance metrics of the application such as testability, scalability, and maintainability. The primary goal of the Data Science and IT engineering teams lies in the timely deployment of the ML application to production, which is conditioned to the approval of the risk assessment teams. These stakeholder groups possess a cognitive state that is predominantly innovation-driven. This implies that they are more likely to favor approaches that are efficient in terms of time and provide better performance.

## 7.2.2 Resources and inter-dependencies

Differences among stakeholders are integral to organizational diversity and productive friction (Passi & Jackson, 2018). When assessing risks, the different viewpoints of the stakeholders can add rigor to the process, thereby resulting in an ML application that is sound, robust, compliant, and responsible.

When diverse stakeholders participate in risk management during ML operationalization, they bring with them certain resources which determine their influence in the process. The resources also create inter-dependencies between stakeholders and provide incentives for them to collaborate with one another (Khan et al., 2022). The resources, in this case, are knowledge and authority. In order to conduct the risk each stakeholder depends on the other for their unique specialized knowledge. The data scientists inform the risk assessment with features used to develop the ML model, the ML algorithm along with its working, and the assumptions underlying the ML model development. The engineers provide information with respect to data processing and storage, platforms security, code quality and vulnerabilities. Whereas the risk teams inform the risk assessment with various techniques to analyze, quantify and mitigate risks. If the stakeholders can clearly articulate and share knowledge with each other, the risk assessment can be highly to beneficial to all the parties.

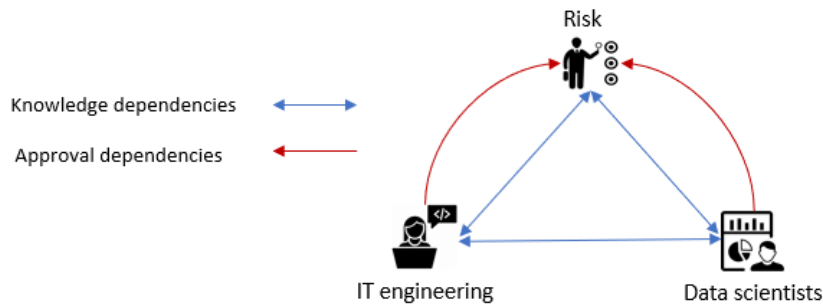


Figure 11: Inter-dependencies between stakeholders

In addition to this, during the process of risk assessments, the risk teams act as gate keepers who can approve or disapprove the production deployments. As a result of this, the risk teams also possess the additional resource of authority, which gives them a relatively higher power position in the network. These inter-dependencies are shown in Figure 11.

### 7.2.3 Trade-offs and Negotiations

The trade-offs are conflicting relationships, which can be observed when achieving or increasing one attribute makes the achievement of another attribute more difficult (Matthewson & Weisberg, 2009). Such conflicting relationships can often show up during risk management. When stakeholders have opposing preferences for the conflicting attributes they must be able to cooperate with one another to reach negotiated decision, instead of ending up in a deadlock that will stall or delay the process. This section introduces two common trade-offs that were identified through expert interviews and literature.

#### 1. Privacy vs ML performance

In order to preserve the privacy of the data and ensure compliance with GDPR, the access to data in banks is limited. In many cases, limited access to data can compromise the performance, especially the accuracy of the ML model (Goldsteen et al., 2021; Raynal et al., 2020).

Data scientists often require access to large and diverse datasets in order to conduct exploratory data analysis and build ML models. Further, in many cases, the use of personally identifiable information (PII) and sensitive data features can improve the accuracy of the model. Risk-averse stakeholders on the other hand are more concerned about the privacy and security of the data. They want to ensure that access to large data sets is limited and the use of PII data and sensitive fields are either prohibited or tightly controlled to ensure regulatory compliance and ethical standards. Further, data scientists are commonly restricted from accessing production data for model development. This often makes it difficult for data scientists to ensure consistency and reproducibility between experimentation and production.

*Product Owner ( Data Science): “There are several privacy and security considerations that need to be made. Do you have permission to move the data to the cloud? Does the data need to be anonymized? Sometimes you have only a fraction of the data to develop your model, and that can be a problem because now the performance of the model in production might decay.”*

*Solution Architect ( Data & AI ): “Data scientists, especially in highly regulated industries like banks struggle to procure data that is good and fit for purpose. With ML you need more than just a small sample set. To make the data available it can be anonymized but the same distribution, dispersion, variance, etc. must be present in the anonymized data set. Sometimes creating this anonymization can also be a challenge. Since the goal is to ensure that a model trained on anonymized data will not behave differently in production.”*

This trade-off does not always result in active resistance since both parties acknowledged the importance of privacy and model performance. However, it is challenging to find a negotiated solution that balances both aspects. This may require dropping certain features to train the ML model or applying certain privacy preservation techniques to obfuscate the data points. However, a ‘latent conflict’ may arise when these negotiations are done too late in the process. This may especially happen if the data scientists have spent a lot of time and effort developing an ML model based on certain data points that they later realize cannot be approved for production, thereby leading to excessive re-work and significantly delaying the production deployment.

*Team Lead (Advanced analytics): “You cannot use certain sensitive features else the model will be incompliant and can discriminate on unethical grounds. One of the smart things we did, was to communicate with risk guys right from the start and ensure that sensitive fields are not used in the model development.”*

A number of literature resources indicate the use of novel techniques to preserve privacy. This includes the use of ensemble learning (Rezaei et al., 2021), differential privacy (Xu et al., 2018), and federated learning (Liu et al., 2019). However, it is important to keep in mind certain limitations of these techniques. Firstly, most of the techniques are in a nascent stage and well suited for a distributed device architecture, it remains uncertain how these techniques can be applied to centralized architecture (which is the case for many banking applications). Secondly, the techniques do not completely overcome the trade-off, a compromise in accuracy (relatively less) must still be made to preserve privacy. Thirdly, introducing new techniques further adds to the complexity of the solution thereby making it even harder to translate and explain the overall ML application to the non-technical risk assessment teams.

## **2. Explainability vs Model Performance:**

It is argued in theory that the explainability decreases as the accuracy of the ML model increase with the use of more complex ML algorithms. The primary interest of data scientists and engineers is to achieve high accuracy, whereas the risk teams are primarily interested in explainability, both to understand the functioning of the ML application and to avoid to risk relating to reliability, compliance and ethics. The expert interviews highlighted key negotiations based on the explainability and complexity of the ML model. These negotiations usually took place with the model validation team and were focused on analyzing the benefits of using the more complex model over the simple model.

*Team lead (Advanced Analytics): “People are reluctant to some extent because sometimes they do not completely understand or at least they feel they don’t fully understand. These are all challenges You have to address. The challenge is to be able to explain the value or what really is the additional benefit of using maybe more complicated models versus maybe simple models.”*

*Data Scientist: “One criterion for model validation is to check what’s the performance increment you get when using a complex model compared to a simple model. There are a lot of machine learning explainability tools like SHAP and LIME and those sort of things that you can use to explain the model output. So, then you can have at least some sort of an understanding of what are you doing you also do model debugging with residual analysis or your sensitivity analysis to see which features are contributing to most of my residuals”*

While complex algorithms do exhibit high accuracy and low degree of explainability this representation is only partially correct and situational. A linear model can produce highly accurate results through massive data pre-processing performed (example: non-linear features). In such cases, the lack of explainability can be attributed to the data pre-processing rather than the algorithm (Koshiyama et al., 2021). To corroborate this further, Herm et al. (2022) argue that this trade-off is situational, and may not only depend on the algorithm but also on other factors such as data complexity. To establish this, they conduct experiments to plot the

accuracy against the explainability perceived by users for two data sets, one representing a high complexity, and another representing low complexity. The results are aggregated to create a plot as shown in Figure 12. *Therefore, the trade-off between explainability and model performance is not solely dependent on the complexity of the ML algorithm, but also on the complexity of data and its processing.*

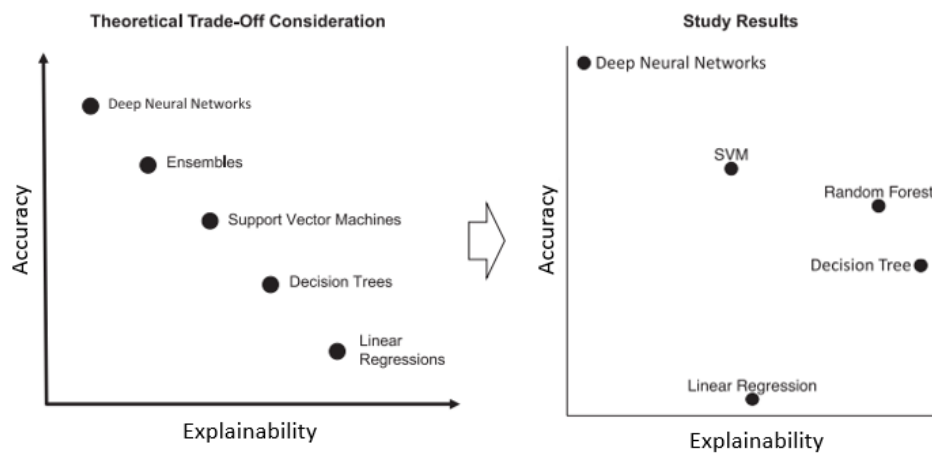


Figure 12: Variations in accuracy vs explainability trade-off based on experiments of Herm et al, (2022)

The data from the expert interviews suggests that negotiation while making this trade-off primarily involves the choice algorithms. However, the risk teams, data scientists, and engineers must additionally consider the dependency of accuracy and explainability on data and its processing, which can introduce more options to reach a negotiated solution.

## 7.3 Social challenges in risk management: Stakeholder Conflicts

### 7.3.1 Knowledge Asymmetries

While diversity among stakeholders is important, findings from the expert interviews and the case study suggest that the current knowledge gap between the stakeholders is very wide and is a major challenge during risk management.

**Data Scientists and Engineers vs Risk:** The expert interviews and case study suggest that the model risk validation team was the only team to possess a strong understanding of ML and data science. Whereas other risk teams that conduct a change risk assessment, data and IT risk assessments, and regulatory assessments often do not have specialized knowledge in ML and data science.

*Enterprise Advisor:* “It can at times be difficult to explain how the model works to the risk assessment teams. Sometimes you have to explain that you are using statistical features to make a prediction, and you are dealing with entropy, volatility, or statistical features which do not have a lot of relevance for the change risk assessment teams.”

*Solution Architect (Data & AI):* “Interaction with different assessment teams can be a challenge. ML solutions are integrated solutions. They are both data-driven and have a software angle to them. Some people view it as pure software, some people view it as a pure data solution. ML is somewhere in between, it has both angles. To converge to this understanding is a challenge.”

Lack of shared knowledge between stakeholders makes it difficult for them to effectively anticipate the likelihood and consequences of the risks (Alhawari et al., 2012). This can also lead to long and repetitive discussions, which can cause conflicts between the stakeholders. Filling this knowledge gaps is especially challenging owing to the technical complexity and opaqueness of ML techniques.

**Data Scientists vs Engineers:** Engineers were found to have better knowledge and understanding of coding standards, software vulnerabilities, and testing. Based on this they prefer following standard practices that can add to the robustness of the ML application and reduce risks. However, if the data scientists are not aware of these, or if these practices are conveyed late in the process, they can be perceive them as overheads that are delaying the ML deployment.

*Solution Engineer ( ML platform): “Data scientists and ML engineers can have very different ways of seeing software and optimizing it. Data Scientists want to optimize model performance, ROC curves, confusion matrix, etc. Whereas ML engineers are more concerned about the coding standards being followed, packages being used, software vulnerabilities, etc. So the two roles have different skills and goals, which can often be challenging.”*

Knowledge asymmetries can also lead to conflicts data scientists and engineers post the deployment of the ML application. Once deployed in production, risks can arise due to ML application failure or performance degradation. If there is a lack of synergy and shared knowledge between data scientists and engineers, it can lead to a lack of accountability and delayed mitigation measures.

*Enterprise Advisor: “When something goes wrong with deployed ML application, you have the IT team saying that we don’t know much about the ML model, and the data science teams saying we don’t know much about the IT application in which the model is embedded. This can often cause a lot of miscommunication and is very difficult to resolve.”*

### 7.3.2 Power-Interest Asymmetries

**Power-Interest in ML Operationalization:** Despite a high power position, the risk assessment teams have relatively fewer incentives to participate in the ML operationalization. While it is a part of their duty to protect the bank against threats, they do not specifically have an intrinsic motivation to involve themselves in the complexities of ML and data science. On the contrary, dealing with the complexities of ML solutions can be taxing for risk teams since they are already tasked with assessing numerous business functions and technologies within the bank (Baquero et al., 2020). This power interest relationship can be show in Figure 13.

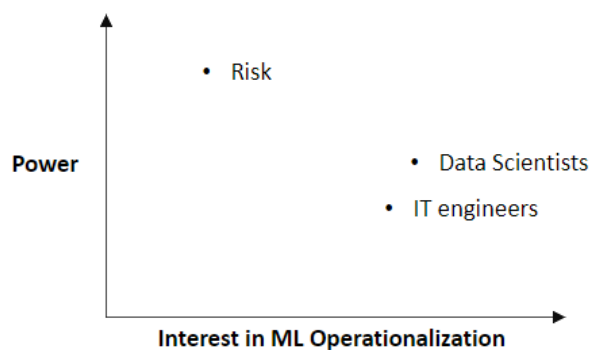


Figure 13: Power-Interest in ML operationalization

The high power and low interest position of risk teams in the ML operationalization can introduce certain difficulties during the assessments. This may in some cases manifest itself in form of reluctance towards ML and AI. Further, risk teams may introduce strict and general standards for approvals and be less open to

negotiations. This might be especially challenging since risk assessment requires negotiating some complex trade-offs and different ML applications can exhibit varying levels of risk.

*Product Owner ( Data Platform): “Discussion with assessment teams can be difficult at times. People are not used to ML and AI. So if you are in a risk-averse function like a chief data officer or something, and someone wants to do AI with their data. The first answer is no. It can take some effort, to educate and convince the people before they can start trusting it.”*

Due the lack of interest, risk teams will be less motivated to learn and retain information with respect to ML. As a result of which workshops and knowledge sharing sessions conducted by data science teams can be less effective. This ineffectiveness of knowledge sharing sessions and workshops was highlighted in the case-study and also by expert participants.

**Power-Interest in Risk Assessment:** On a similar note, Data scientist and IT engineers have a few incentives with respect to the risk assessment process. They acknowledge the importance of the risk assessment, however, given their innovation-driven temperament, they do not have an intrinsic motivation to follow several standards and undergo various risk-related procedures. Resulting in a power interest relationship as in Figure 14.

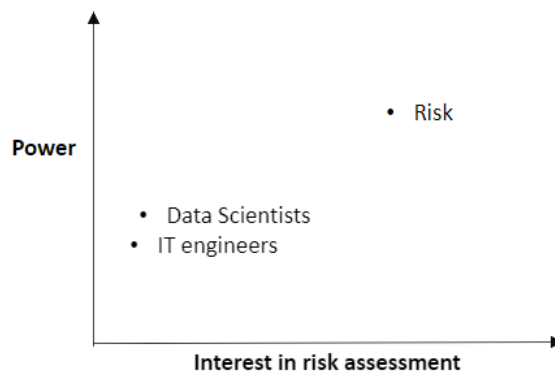


Figure 14: Power vs Interest in risk assessment

*Risk consultant ( Model Risk Validation): “They ( data scientists ) don’t like following a lot of standards. For example, writing long extensive documentation with every detail, and taking measures for reproducibility. They also don’t like doing all maintenance work, once the application is deployed. ”*

While the diversity of stakeholders is crucial to the ML operationalization process, when the knowledge base and interests of the stakeholders become too distant from one another, it can often lead to inter-group conflicts and a lack of cooperation (Nelson & Coopride, 1996). As a result of this the risk management during the ML operationalization can be excessively cumbersome and ineffective

## 7.4 Strategic Guidelines for Managing Risks in ML operationalization

Based on the identified challenges of risk management, banks should strategically aim to introduce controls to manage the intractable nature of the ML operationalization process, and also improve shared knowledge and interests between the stakeholders involved in risk management. This section provides four strategic guidelines which provide an explanation of how the challenges of risk management can be addressed.

### 1. Moving from explainable ML Algorithms to explainable ML operationalization:

Many banks have already started adopting various tools and techniques to explain ML algorithms. In addition, to this, they should also be able to explain various aspects of the operationalization process, which include:

*Which models are in production? How are they performing?*

Organizations must have an overview of all the models and their versions running in production. Further organization must also be able to monitor the performance of the ML model in production.

*What are the artifacts corresponding to a particular model?*

There should be a lineage between the ML model deployed in production, the data set it is trained on, application software code and configurations, and the experiment it corresponds to (including the feature engineering, algorithm, and hyper-parameters). When the artifacts are developed and updated, they must be tracked and versioned.

*What are the dependencies that impact the ML application?*

There should be a comprehensive understanding of the systems and data dependencies that can impact the ML models running in production.

*Who will fix issues and take accountability when things go wrong?*

There should be a clear set of roles and responsibilities to ensure accountability when production issues occur. Further, important decisions made by different stakeholders should be documented and kept track of (de Bruijn et al., 2021).

ML controls, which include monitoring, versioning, lineage, model explainability, roles, and responsibilities are needed to make the process of ML operationalization transparent and auditable. Developing these controls can serve as a strategic investment for banks as they aim to scale their AI initiatives.

### 2. De-risking ML by design

Risk management should not be an afterthought and must be embedded in the design of the ML application and the ML operationalization process. To achieve this:

i) Coordination between a Data scientist and IT engineers must be strengthened, in order to incorporate certain standards and controls during the experimentation and application engineering phase. There must also be sufficient communication and synergy between data scientist and engineers in order to take shared responsibility post the deployment of the ML application.

ii) Complementary skills can be introduced into the technical teams. This includes developing awareness of various ML-related risks, and the ability to translate and communicate complex data and ML techniques in simple and effective terms (Passi & Jackson, 2018).

iii) Stakeholders from the risk department must be involved right from the beginning to reduce potential conflicts and prevent rework. Basic ML and data science competencies can be developed in risk teams to effectively analyze risks in ML development and deployment.

Implementing risk measures in the design and development of ML solutions may not show immediate value. However, it is the key to reducing costly delays and strengthening the first line of defense.

### **3. Integrating risk management into agile workflows**

Most data science and engineering teams follow an agile way of working which is iterative and favors rapid development and deployment to production. Further, the risk assessment should not be limited to the first deployment to production and should take into account the changes made to the ML application throughout its life cycle. The traditional risk assessments and approvals must therefore be integrated with agile workflow of ML life-cycle (Baquero et al., 2020).

This does not necessarily require the risk teams to work in agile sprints. However, the risk teams have to think beyond one-time approvals and consider the cyclic nature ML operationalization. Further, responsibility of managing risk cannot be the sole responsibility of risk teams. The data scientist, engineers, and risk experts must make some important and joint decisions which include: how often should ML application be reviewed? What should be the intensity of reviews? What reviews and checks can be automated through testing? What reviews require manual interventions?

### **4. Red team and blue team Gamification**

A common approach in information security risk management is to conduct an intra-organizational exercise where the security teams are divided into the red team, which attacks the infrastructure of the organization, and the blue team which defends the infrastructure of the organization against the attacks. The goal of the exercise is to detect security vulnerabilities and strengthen the infrastructure of the organization.

This approach can be extended to ML applications, by conducting an exercise where the developers of the ML application (blue team) are challenged by a team of technical and non-technical risk experts (red team). Depending on the maturity of the organization and the complexity of the ML application, the attacks can range from simple manual scenarios to advanced adversarial attacks to test the robustness of the ML application. The exercise can therefore motivate development of new and relevant control mechanisms to minimize the risks in the ML operationalization process. Another advantage of this approach is that it facilitates implicit knowledge sharing, by making the process stimulating and interesting for the different stakeholders. At the same time, preserving the diversity of knowledge and interest between the different stakeholders,

## **7.5 Conclusion of Risk Management: Developing a Conceptual Model**

The aim of this chapter was to answer two sub-research questions. The third sub-research question: *What are the socio-technical challenges specific to managing risks during the ML operationalization process?* And the fourth sub-research question: *How can banks strategically address the challenges of risk management?*

The socio-technical challenges of managing risks in the ML operationalization process lie in the intractability of the process, and the knowledge and the power interest asymmetries between stakeholders. This often makes it difficult for stakeholders to translate, communicate and negotiate with one another. With these findings the chapter answers the third sub-research question. To answer the fourth sub-research question the chapter presents four strategic guidelines that provide actionable insights to address the identified challenges of risk management. The strategies proposed are based on the underlying concepts of shared knowledge and controls, and eventually contribute to the successful outcomes of the ML operationalization process. To capture this integrated understanding a conceptual model is proposed as shown in Figure 15.

The conceptual model posits that the drivers: shared knowledge and controls can be instrumental in reducing these challenges and achieving the successful outcomes identified in this research. The concept of shared knowledge goes beyond the idea of just sharing facts and information. Effective shared knowledge is represented by a synergy between groups and is achieved through mutual trust and influence (Nelson & Coopride, 1996). In the conceptual model, shared knowledge refers to effective collaboration between diverse stakeholders involved in the risk management based on shared interests and goals. The stakeholders must also possess some cross-functional knowledge of data science, engineering, and risk capabilities. Further, the stakeholders must also be able effectively to translate different forms of specialized knowledge to communicate with one another. During the ML operationalization process, the stakeholders will be faced with conflicting interests, preferences, and trade-offs. In such situations, the information is not right or wrong, rather the stakeholders need to make a joint compromise to reach a mutually agreeable decision. Such an established knowledge is referred to as negotiated knowledge (de Bruijn & ten Heuvelhof, 2018).

Controls provide explicit mechanisms to deal with the intractable nature of the ML operationalization process by identifying and mitigating risks. Controls such as monitoring, testing, model explainability and fairness, and roles and responsibilities have been individually discussed in the existing literature. While, versioning and lineage, have been emerging concepts in MLOps literature, it is mostly been viewed as a mechanism to complement the efficiency and adaptability of the ML Operationalization process. This research posits that versioning and lineage between various artifacts (model, experiments, training data) can serve as an important control that can enhance the transparency of the process, thereby aiding the identification and mitigation of various risks

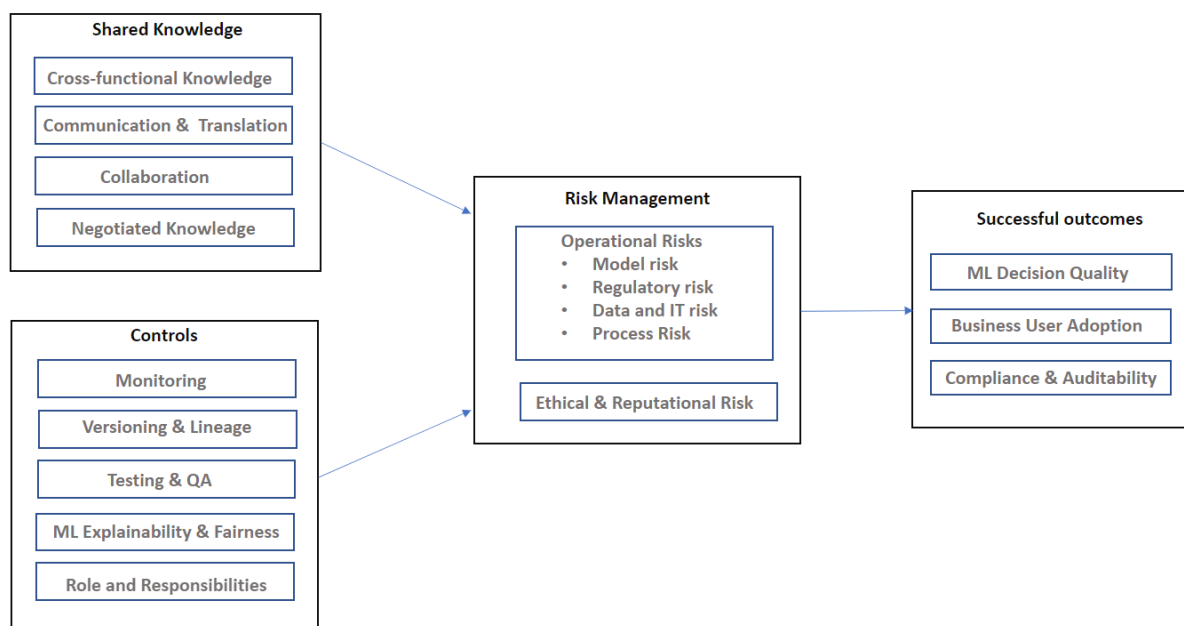


Figure 15: A Conceptual model for managing risks during ML operationalization

The conceptual model further supports the view effective risk management contributes to factors representing successful outcomes of ML operationalization. The discussion of this chapter makes it evident that risk management can contribute towards ML decision quality (through robust model risk management), and compliance and auditability (through managing regulatory and ethical risks). Additionally, the expert interviews and network inter-relationship diagram in Figure 9 indicate that risk management can increase the confidence in business users, thereby aiding business user adoption of the ML application. The conceptual therefore shows how a socio-technical approach towards risk management can support the ML operationalization in banks. While ML risks are widely discussed in literature, this research presents the first comprehensive conceptual model to manage risks during the ML operationalization process.

# 8 Conclusions

As banks embark on their AI transformation journey, ML operationalization serves as an important means to deliver business impact through sustainable production deployments. The aim of this study has been to develop a socio-technical framework to strengthen both the understanding and implementation of the ML operationalization process. This chapter synthesizes the main findings of the research and explains its practical and scientific relevance. It further analyzes the limitations of the study and provides direction for future research.

## 8.1 Main Findings

This research highlights the socio-technical complexities in the ML operationalization process. The ML operationalization workflow consists of a number of activities involving ML experimentation, application engineering, risk assessment, application deployment, and maintenance. The execution of these activities requires collaboration between a diverse group of stakeholders having different expertise, interests, and preferences. Despite the differences, the stakeholders must be able to translate with one another and make negotiated decisions. From a technology standpoint, the research highlights that, unlike traditional software engineering, the integrated nature of ML applications results in multiple dependencies that exist in application software, data, and ML configurations. Not only does this add to the complexity of the development, deployment, and maintenance activities, but also makes it challenging for technical and non-technical stakeholders to effectively communicate and translate with one another to develop a shared understanding of the process.

Based on the 15 expert interviews and literature study, the research identified nine socio-technical factors that influence the ML operationalization process. These factors include: business user adoption, machine learning decision quality, compliance & audibility, data quality, adaptability, enterprise integration, risk management, shared knowledge, and controls. The identified factors serve as theoretical concepts that shape the phenomenon of the ML operationalization process. Further analysis indicates that the socio-technical factors are not isolated, but show dense inter-relationships. A case study on operationalizing an ML application for anti-money laundering validates the factors and shows how the factors apply in a real world context.

While the analysis of all the factors and their relationships makes the research comprehensive, it can compromise the depth and parsimony of the research. Therefore, a decision was made to conduct an in-depth investigation into risk management which is one of the most crucial yet challenging factors in the process. The investigation of risk management highlights that technical risks may not only arise from opaqueness and bias in ML algorithms, but also from opaqueness and intractability of the process of ML operationalization. Further, the analysis of social systems highlights significant knowledge and power interest asymmetries between data science, engineering, and risk teams that lead to cumbersome and ineffective risk management. To overcome these challenges the research proposes four strategic guidelines: 1) Moving from explainable ML algorithms to explainable ML operationalization, 2) De-risking ML by design, 3) Integrating Risk management into the agile workflow, 4) Red Team and Blue Team gamification. Further a conceptual model is developed to capture the integrated understanding of risk management during ML operationalization. The model shows how factors such as shared knowledge and controls reduce challenges in risk management and contribute toward successful outcomes such as business user adoption, ML decision quality, and compliance, and auditability. With this, the research proposes the first conceptual model for managing risks during the ML operationalization process.

## 8.2 Practical Relevance

The socio-technical framework developed in this research can be used by industry practitioners both at the start of the implementation and during the execution of the ML operationalization process. To further elaborate on the practical relevance the practitioners are categorized into managers, developers, and assessors. In addition to this, the practical relevance for consultancy firms such as Deloitte is also elaborated.

Managers are responsible for setting strategic targets and making important decisions with respect to ML initiatives. This may include members from the senior management and middle management. The socio-technical complexities highlighted in the research can help managers set realistic expectations and achievable goals with respect to ML. Using the list of socio-technical factors, managers can develop a comprehensive view of the ML operationalization process and further investigate or prioritize the business functions that require attention in order to support the process. The managers can also use the strategic guidelines proposed in this research to bolster the operationalization activities by fostering a culture of mutual cooperation between the technical and risk team.

The developers refer to product owners, data scientists, engineers, or any entity directly involved in the development, deployment, and maintenance of the ML application. The socio-technical analysis factors help the developers understand the broader implications of the ML operationalization. The developers may especially benefit from understanding the social complexities of the process, which can help them work towards better translation and negotiation with non-technical stakeholders. By referring to the strategic guidelines, the developers can build control mechanisms and take preemptive measures to ensure robust yet efficient risk management.

The assessors refer to stakeholders who are tasked with scrutinizing the ML application for various risks. The research informs the assessors of risks embedded in the ML operationalization which are less discussed in literature and practice. It further provides strategic guidelines to develop ML and data science competencies in risk teams and to adapt risk assessments to complement the ML operationalization process.

Lastly, the socio-technical framework can be leveraged by consultancy firms like Deloitte, that deal with a number of clients in the financial service industry. Consultants at Deloitte can use the socio-technical framework as a reference to understand the current condition of the ML operationalization process of their clients. Based on the existing challenges and the required drivers, Deloitte can create a client roadmap to achieve the successful outcomes of the ML operationalization process. Further, both consultants in the AI and data teams and in Risk advisory can benefit from the strategic guidelines for risk management to support the ML operationalization process of their clients. Since the framework identifies a comprehensive list of socio-technical factors, corresponding to a very relevant problem, Deloitte may also use this framework for their research initiatives such as international CIO surveys which identifies technology trends across industries and geographies.

## 8.3 Scientific Relevance

This research makes two main contributions to scientific literature. First, it illustrates the socio-technical complexities of the ML operationalization process in banks and presents a comprehensive list of socio-technical factors. Second, it provides an in-depth analysis of risk management during ML operationalization and integrates the findings into a conceptual model.

One of the knowledge gaps identified in the literature study relates to the predominantly technical overview of the ML operationalization process. This knowledge gap is addressed by the socio-technical analysis and the socio-technical factors identified in this research. The analysis augments the technical workflow by capturing the collaboration and interaction between various stakeholders to elaborate on the socio-technical complexities of the ML operationalization process. Further, a comprehensive list of socio-technical factors is developed

which integrates a list of theoretical concepts influencing ML operationalization. To support the factors with empirical evidence, each factor is associated with its corresponding literature source and expert interview participant. While these findings are useful for scholars interested in socio-technical systems and organization science, they can also benefit scholars researching in the area of computer science and engineering, by informing them about the broader pragmatic considerations of the ML operationalization process.

While risk with respect to ML is a widely discussed topic in scientific literature, there has been little discussion with respect to managing risks during the ML operationalization process. To address this knowledge gap the research presents an analysis of the social and technical challenges in managing risk during the ML operationalization process. The findings provide an elaborated understanding of the interaction between data scientists, engineers, and risk teams which is rarely highlighted in the existing literature. Further, the analysis also aims to motivate scientific debates and discussions that not only focus on the risks of ML algorithms but also on the risks embedded in the operationalization of ML applications. Lastly, a first such conceptual model is presented which captures the understanding of risk management within the context of the ML operationalization process. The conceptual model can serve as tentative theory (Maxwell, 2005), and can be used by other scholars to develop and test hypothesis in future research.

## **8.4 Limitations**

This section addresses the limitations of the research in order to promote transparent and accurate interpretation of results, and at the same time provide an impetus for future work.

Firstly, there are limitations with respect to the generalizability of this research. While this study provides a comprehensive list of socio-technical factors and sub-factors, it does not claim to be an all-encompassing list. Therefore, leaving room for unknown and un-discovered factors. The study was conducted within the scope of the banking sector in the geographic delineation of the Netherlands. While a number of socio-technical factors and relationships may apply to other industries and geographies, the users of the framework must be cautious in doing so and should be open to possible changes and nuances. Further, of the experts interviewed, only two experts belonged to academia, with one of them also having a risk-averse function in a consultancy. In addition to this, the study uses only a single use case to validate the study. ML use cases can vary across industries but also in the banking industry itself. Different use cases in the banking domain may highlight variations in socio-technical factors and relationships. For example, the socio-technical factors that apply to a highly regulated use case such as anti-money laundering may differ from the socio-technical factors that apply to a more commercial use case in retail banking.

Secondly, the study only conducts an in-depth analysis of one of the factors: risk management. A number of other factors that represent such as data quality, enterprise integration, adaptability, and business user adoption, remain unexplored. These factors exhibit a strong socio-technical character and can yield valuable insights through a similar in-depth socio-technical analysis. Lastly, while a conceptual model is presented in this research, it remains to be empirically tested.

## **8.5 Future Research Directions**

To build on the current work, two future research directions are provided. The first direction proposes to improve the internal and external validity of the research. The research can further benefit from increasing the diversity of expert participants, especially by including more experts belonging to risk-averse roles and experts from academia. The inclusion of experts with these backgrounds can introduce new perspectives with respect to risk management, controls, compliance, and audibility which can augment or change the existing set of factors and their relationships. In addition to this, more diverse use cases can be analyzed in the context of the ML operationalization process. This can add more rigor to the framework, by increasing the external validity and reliability of the socio-technical factors and relationships. To further augment the generalizability of the

study, researchers may expand the findings to other industries and geographies. This approach can facilitate the identification of various patterns and differences with respect to ML operationalization.

The second research direction aims to add depth to the study. The analysis of risk management can be considered as an analysis of a socio-technical subsystem that exists within the broader socio-technical system of the ML operationalization process. Researchers can identify other such subsystems which are represented by factors such as data quality, adaptability, business user adoption, and enterprise integration. Thereby leading to a more in-depth analysis of other factors. Further, the conceptual model proposed for risk management can be used by researchers to develop research problems and designs that can validate the cause-effect relationship between the factors in the conceptual model, thereby leading to an established theory that is empirically tested.

## **8.6 Relevance to the MOT program**

This thesis is a part of the MSc program in Management of Technology, which aims to develop intellectual competencies in students, that allow them “to analyze technology and its commercial impact, and implement these technologies in the organizational context of the firm”. A number of courses studied in this program have been instrumental in developing the framework and guiding the direction of this research. This includes courses such as technology dynamics which explicate the multiple dimensions of socio-technical change and emerging and breakthrough technology which helps to understand how innovation progresses throughout out a timeline. The course in business process management contributed significantly to analyzing the workflow of the ML operationalization process. Followed by Inter and Intra organization decision making which elaborates the importance of managing stakeholders in networks and developing negotiated knowledge. The specialization in ICT design and management further augments the understanding of how ML is positioned in the organization as a technology capability and its dependencies on other systems and processes. Further elective courses such as technology battles have helped to structure various factors influencing a process or a technology. The overall thesis process is also guided by the course on research methods which helps to build an understanding of qualitative data analysis and framework development. Much of the research in this study has focused on inter-relating the social systems and technical systems in the organization in order to develop a comprehensive view of the ML operationalization process. The research bridges the gap between the technology-related and business-related aspects of ML and deeply analyzes the intuitional complexities involved in converting ML into a valuable corporate resource.

## **8.7 Personal Reflection**

In this section, I would like to take the opportunity to reflect and share some of the learnings of my thesis journey which could possibly benefit other researchers.

At the start of the thesis journey, it was difficult to completely comprehend the breadth of the topic given the nascent stage of ML operationalization. It turned out that the number of factors and relationships were far broader than I had expected. This made it extremely difficult to inter-relate each and every factor. This can often be the case when one is conducting exploratory research about a new phenomenon. Based on astute advice from my supervisors, I decided to narrow down the research and dive deep into a specific factor. This decision has been extremely conducive to the research. Not only did it add depth to the study, but on a personal level, it made the process of analyzing findings and writing my report a lot more organized and enjoyable. Thesis research is an interactive and iterative process, over the course of which a researcher may have to make some changes in order to refine the topic. One of the takeaways from this experience was to think about scoping the research topic in terms of constructs and variables. This also made me realize that conducting research is not about covering each and every aspect. It is more about developing a deep understanding of the phenomenon being studied, in a way that is useful to the readers.

I have learned that one needs to balance new experiments with other practical considerations. The essence of research lies in discovering the unknown, it is therefore only natural that one would like to conduct some experiments or use certain approaches that have not been done before. However, it is important to plan these experiments within the constraints of time ( or other resources ). During this project, I attempted some visualization techniques on qualitative data, which, however, did not produce results that were easy to interpret. While it was very tempting to continue the experiment, I believe made the right decision to move ahead and focus on other experiments and findings that were more relevant to the research.

I would like to emphasize that report writing can often be an overlooked component of the thesis. For me, this thesis report has been the most elaborate document I have written so far in my life. Over the course of the study, one develops a number of ideas, and can sometimes be challenging to integrate all the ideas in a coherent and compelling manner while documenting the report. What helped me was to write frequently and reflect on the output. Soon you develop a mental framework that guides you to put all the pieces together.

Finally, research has been a highly iterative and interactive process. I have particularly benefitted from discussing my work and receiving constant feedback from my supervisors and other Ph.D. scholars and professionals. Often others can see what you cannot, and it is, therefore, it is important to keep an open mind and help others help you.

# REFERENCES

- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- Alhawari, S., Karadsheh, L., Nehari Talet, A., & Mansour, E. (2012). Knowledge-Based Risk Management framework for Information Technology project. *International Journal of Information Management*, 32(1), 50–65. <https://doi.org/10.1016/j.ijinfomgt.2011.07.002>
- Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. <https://doi.org/10.1109/ICSE-SEIP.2019.00042>
- Asatiani, A., Malo, P., Nagbøl, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems. *Journal of the Association for Information Systems*, 22, 325–352. <https://doi.org/10.17705/1jais.00664>
- Ashmore, R., Calinescu, R., & Paterson, C. (2021). Assuring the Machine Learning Lifecycle. *ACM Computing Surveys*, 54(5), 1–39. <https://doi.org/10.1145/3453444>
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE*, 10(7), e0130140. <https://doi.org/10.1371/journal.pone.0130140>
- Baier, L., Jöhren, F., & Seebacher, S. (2019, January). *CHALLENGES IN THE DEPLOYMENT AND OPERATION OF MACHINE LEARNING IN PRACTICE*.
- Baquero, J. A., Burkhardt, R., Govindarajan, A., & Wallace, T. (2020). *Derisking AI by design: How to build risk management into AI development*.
- Basias, N., & Pollalis, Y. (2018). Quantitative and qualitative research in business & technology: Justifying a suitable research methodology. *Review of Integrative Business and Economics Research*, 7, 91–105.
- Borg, M. (2022). *Agility in Software 2.0 – Notebook Interfaces and MLOps with Buttresses and Rebars* (pp. 3–16). [https://doi.org/10.1007/978-3-030-94238-0\\_1](https://doi.org/10.1007/978-3-030-94238-0_1)
- Bostrom, R. P., & Heinen, J. S. (1977a). MIS Problems and Failures: A Socio-Technical Perspective. Part I: The Causes. *MIS Quarterly*, 1(3), 17. <https://doi.org/10.2307/248710>
- Bostrom, R. P., & Heinen, J. S. (1977b). MIS Problems and Failures: A Socio-Technical Perspective, Part II: The Application of Socio-Technical Theory. *MIS Quarterly*, 1(4), 11. <https://doi.org/10.2307/249019>
- Bourgais, A., & Ibnouhsein, I. (2021). Ethics-by-design: the next frontier of industrialization. *AI and Ethics*. <https://doi.org/10.1007/s43681-021-00057-0>
- Brock, J. K. U., & von Wangenheim, F. (2019). Demystifying Ai: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review*, 61(4), 110–134. <https://doi.org/10.1177/1536504219865226>
- Brous, P., Janssen, M., & Krans, R. (2020). *Data Governance as Success Factor for Data Science* (pp. 431–442). [https://doi.org/10.1007/978-3-030-44999-5\\_36](https://doi.org/10.1007/978-3-030-44999-5_36)
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 205395171562251. <https://doi.org/10.1177/2053951715622512>
- Cabitza, F., Campagner, A., & Balsano, C. (2020). Bridging the “last mile” gap between AI implementation and operation: “data awareness” that matters. *Annals of Translational Medicine*, 8(7), 501–501. <https://doi.org/10.21037/atm.2020.03.63>
- Carroll, A. B., & Buchholtz, A. K. (1996). Ethics and stakeholder management. *Cincinnati: South-Western*.
- Chen, Q. Z., Schnabel, T., Nushi, B., & Amershi, S. (2022). HINT: Integration Testing for AI-based features with Humans in the Loop. *27th International Conference on Intelligent User Interfaces*, 549–565. <https://doi.org/10.1145/3490099.3511141>

- Chen, Z., van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems*, 57(2), 245–285. <https://doi.org/10.1007/s10115-017-1144-z>
- Chou, Y.-L., Moreira, C., Bruza, P., Ouyang, C., & Jorge, J. (2022). Counterfactuals and causability in explainable artificial intelligence: Theory, algorithms, and applications. *Information Fusion*, 81, 59–83. <https://doi.org/10.1016/j.inffus.2021.11.003>
- Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., & Abadie, A. (2022). AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research*, 144, 31–49. <https://doi.org/10.1016/j.jbusres.2022.01.069>
- de Bruijn, H., & ten Heuvelhof, E. (2018). *Management in Networks*. Routledge. <https://doi.org/10.4324/9781315453019>
- de Bruijn, H., Warnier, M., & Janssen, M. (2021). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, 101666. <https://doi.org/10.1016/j.giq.2021.101666>
- de Nederlandsche Bank. (2019). *General principles for the use of Artificial Intelligence in the financial sector*.
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., ... Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Enni, S. A., & Herrie, M. B. (2021). Turning biases into hypotheses through method: A logic of scientific discovery for machine learning. *Big Data & Society*, 8(1), 205395172110207. <https://doi.org/10.1177/2053951721102075>
- Flach, P. (2019). Performance Evaluation in Machine Learning: The Good, the Bad, the Ugly, and the Way Forward. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 9808–9814. <https://doi.org/10.1609/aaai.v33i01.33019808>
- Flick, U. (2017). *The Sage handbook of qualitative data collection*. Sage.
- Fountain, T., McCarthy, B., & Saleh, T. (2019). *Building the AI-Powered Organization*.
- Garcia, R., Sreekanti, V., Yadwadkar, N., Crankshaw, D., Gonzalez, J. E., & Hellerstein, J. M. (2018). Context: The missing piece in the machine learning lifecycle. *KDD CMI Workshop*, 114.
- Goldstein, A., Ezov, G., Shmelkin, R., Moffie, M., & Farkash, A. (2021). Data minimization for GDPR compliance in machine learning models. *AI and Ethics*. <https://doi.org/10.1007/s43681-021-00095-8>
- Granlund, T., Stirbu, V., & Mikkonen, T. (2021). Towards Regulatory-Compliant MLOps: Oravizio's Journey from a Machine Learning Experiment to a Deployed Certified Medical Product. *SN Computer Science*, 2(5), 342. <https://doi.org/10.1007/s42979-021-00726-1>
- Grover, V., & Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. *MIS Quarterly*, 39(2), 271–296. <https://doi.org/10.25300/MISQ/2015/39.2.01>
- Guo, L., Daly, E. M., Alkan, O., Mattetti, M., Cornec, O., & Knijnenburg, B. (2022). Building Trust in Interactive Machine Learning via User Contributed Interpretable Rules. *27th International Conference on Intelligent User Interfaces*, 537–548. <https://doi.org/10.1145/3490099.3511111>
- Haakman, M., Cruz, L., Huijgens, H., & van Deursen, A. (2021). AI lifecycle models need to be revised. *Empirical Software Engineering*, 26(5), 95. <https://doi.org/10.1007/s10664-021-09993-1>
- Hacker, P., Krestel, R., Grundmann, S., & Naumann, F. (2020). Explainable AI under contract and tort law: legal incentives and technical challenges. *Artificial Intelligence and Law*, 28(4), 415–439. <https://doi.org/10.1007/s10506-020-09260-6>
- Han, J., Huang, Y., Liu, S., & Towey, K. (2020). Artificial intelligence for anti-money laundering: a review and extension. *Digital Finance*, 2(3–4), 211–239. <https://doi.org/10.1007/s42521-020-00023-1>

- Herm, L.-V., Heinrich, K., Wanner, J., & Janiesch, C. (2022). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management*, 102538. <https://doi.org/10.1016/j.ijinfomgt.2022.102538>
- Hummer, W., Muthusamy, V., Rausch, T., Dube, P., el Maghraoui, K., Murthi, A., & Oum, P. (2019). ModelOps: Cloud-Based Lifecycle Management for Reliable and Trusted AI. *2019 IEEE International Conference on Cloud Engineering (IC2E)*, 113–120. <https://doi.org/10.1109/IC2E.2019.00025>
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 37(3). <https://doi.org/10.1016/j.giq.2020.101493>
- Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Kuk, G. (2020). Will Algorithms Blind People? The Effect of Explainable AI and Decision-Makers' Experience on AI-supported Decision-Making in Government. *Social Science Computer Review*, 089443932098011. <https://doi.org/10.1177/0894439320980118>
- Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research*, 70, 338–345. <https://doi.org/10.1016/j.jbusres.2016.08.007>
- John, M. M., Olsson, H. H., & Bosch, J. (2021). Towards MLOps: A Framework and Maturity Model. *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 1–8. <https://doi.org/10.1109/SEAA53835.2021.00050>
- Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., & Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control*, 23(1), 173–186. <https://doi.org/10.1108/JMLC-07-2019-0055>
- Kaddoumi, T., & Tambo, T. (2022). *Democratizing Enterprise AI Success Factors and Challenges: A Systematic Literature Review and a Proposed Framework* (pp. 640–652). [https://doi.org/10.1007/978-3-030-95947-0\\_45](https://doi.org/10.1007/978-3-030-95947-0_45)
- Karamitsos, I., Albarhami, S., & Apostolopoulos, C. (2020). Applying DevOps Practices of Continuous Automation for Machine Learning. *Information*, 11(7), 363. <https://doi.org/10.3390/info11070363>
- Karimi, T., & Yahyazade, Y. (2021). Developing a risk assessment model for banking software development projects based on rough-grey set theory. *Grey Systems: Theory and Application, ahead-of-print*(ahead-of-print). <https://doi.org/10.1108/GS-05-2021-0074>
- Keller, K. L., & Staelin, R. (1987). Effects of Quality and Quantity of Information on Decision Effectiveness. *Journal of Consumer Research*, 14(2), 200. <https://doi.org/10.1086/209106>
- Kerzel, U. (2021). Enterprise AI Canvas Integrating Artificial Intelligence into Business. *Applied Artificial Intelligence*, 35(1), 1–12. <https://doi.org/10.1080/08839514.2020.1826146>
- Khan, I. S., Kauppila, O., Fatima, N., & Majava, J. (2022). Stakeholder interdependencies in a collaborative innovation project. *Journal of Innovation and Entrepreneurship*, 11(1), 38. <https://doi.org/10.1186/s13731-022-00229-0>
- Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., & Lomas, E. (2021). Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3778998>
- Kuguoglu, B. K., van der Voort, H., & Janssen, M. (2021). The Giant Leap for Smart Cities: Scaling Up Smart City Artificial Intelligence of Things (AIoT) Initiatives. *Sustainability*, 13(21), 12295. <https://doi.org/10.3390/su132112295>
- Lebcir, R., Hill, T., Atun, R., & Cubric, M. (2021). Stakeholders' views on the organisational factors affecting application of artificial intelligence in healthcare: a scoping review protocol. *BMJ Open*, 11(3), e044074. <https://doi.org/10.1136/bmjopen-2020-044074>
- Liu, Z., Li, T., Smith, V., & Sekar, V. (2019). *Enhancing the Privacy of Federated Learning with Sketching*.
- Lwakatare, L. E., Raj, A., Crnkovic, I., Bosch, J., & Olsson, H. H. (2020). Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and Software Technology*, 127, 106368. <https://doi.org/10.1016/j.infsof.2020.106368>

- Matthewson, J., & Weisberg, M. (2009). The structure of tradeoffs in model building. *Synthese*, 170(1), 169–190. <https://doi.org/10.1007/s11229-008-9366-y>
- Maxwell, J. A. (2005). Conceptual framework: What do you think is going on. *Qualitative Research Design: An Interactive Approach*, 41, 33–63.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 1–35. <https://doi.org/10.1145/3457607>
- Miceli, M., Posada, J., & Yang, T. (2022). Studying Up Machine Learning Data. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1–14. <https://doi.org/10.1145/3492853>
- Miller, H. (1996). THE MULTIPLE DIMENSIONS OF INFORMATION QUALITY. *Information Systems Management*, 13(2), 79–82. <https://doi.org/10.1080/10580539608906992>
- Miller, S. M. (2018). *AI: Augmentation, more so than automation*.
- Mohassel, P., & Zhang, Y. (2017). SecureML: A System for Scalable Privacy-Preserving Machine Learning. *2017 IEEE Symposium on Security and Privacy (SP)*, 19–38. <https://doi.org/10.1109/SP.2017.12>
- Muthusamy, V., Slominski, A., & Ishakian, V. (2018). Towards Enterprise-Ready AI Deployments Minimizing the Risk of Consuming AI Models in Business Applications. *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, 108–109. <https://doi.org/10.1109/AI4I.2018.8665685>
- Nelson, K. M., & Coopridge, J. G. (1996). The Contribution of Shared Knowledge to IS Group Performance. *MIS Quarterly*, 20(4), 409. <https://doi.org/10.2307/249562>
- Olan, F., Ogiemwonyi Arakpogun, E., Suklan, J., Nakpodia, F., Damij, N., & Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. *Journal of Business Research*, 145, 605–615. <https://doi.org/10.1016/j.jbusres.2022.03.008>
- Omidvar, O., Safavi, M., & Glaser, V. L. (2022). Algorithmic routines and dynamic inertia: How organizations avoid adapting to changes in the environment. *Journal of Management Studies*. <https://doi.org/10.1111/joms.12819>
- Passi, S., & Jackson, S. J. (2018). Trust in Data Science. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–28. <https://doi.org/10.1145/3274405>
- Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2017). Data Management Challenges in Production Machine Learning. *Proceedings of the 2017 ACM International Conference on Management of Data*, 1723–1726. <https://doi.org/10.1145/3035918.3054782>
- Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2018). Data Lifecycle Challenges in Production Machine Learning. *ACM SIGMOD Record*, 47(2), 17–28. <https://doi.org/10.1145/3299887.3299891>
- Preece, A., Harborne, D., Braines, D., Tomsett, R., & Chakraborty, S. (2018). *Stakeholders in Explainable AI*. arXiv. <https://doi.org/10.48550/ARXIV.1810.00184>
- Raina, V., & Krishnamurthy, S. (2022). Building and Structuring the Team. In *Building an Effective Data Science Practice* (pp. 315–323). Apress. [https://doi.org/10.1007/978-1-4842-7419-4\\_22](https://doi.org/10.1007/978-1-4842-7419-4_22)
- Raynal, M., Achanta, R., & Humbert, M. (2020). *Image Obfuscation for Privacy-Preserving Machine Learning*.
- Renggli, C., Rimanic, L., Gürel, N. M., Karlaš, B., Wu, W., & Zhang, C. (2021). *A Data Quality-Driven View of MLOps*.
- Rezaei, S., Shafiq, Z., & Liu, X. (2021). *Accuracy-Privacy Trade-off in Deep Ensemble: A Membership Inference Perspective*.
- Robbins, S. (2020). AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. *AI & SOCIETY*, 35(2), 391–400. <https://doi.org/10.1007/s00146-019-00891-1>
- Rossi, F. (2018). Building trust in artificial intelligence. *Journal of International Affairs*, 72(1), 127–134.
- Ruf, P., Madan, M., Reich, C., & Ould-Abdeslam, D. (2021). Demystifying MLOps and Presenting a Recipe for the Selection of Open-Source Tools. *Applied Sciences*, 11(19), 8861. <https://doi.org/10.3390/app11198861>
- Salama, K., & Kazmierczak, J. (2021). *Practitioners guide to MLOps: A framework for continuous delivery and automation of machine learning*.

- Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). “Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3411764.3445518>
- Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, 43(3), 695–719. <https://doi.org/10.25300/MISQ/2019/13747>
- Schelter, S., Biessmann, F., Januschowski, T., Salinas, D., Seufert, S., & Szarvas, G. (2018). On Challenges in Machine Learning Model Management. *IEEE Data Eng. Bull.*, 41, 5–15.
- Schilling, M. A. (2000). Decades ahead of her time: advancing stakeholder theory through the ideas of Mary Parker Follett. *Journal of Management History*, 6(5), 224–242. <https://doi.org/10.1108/13552520010348371>
- Schlögl, S., Postulka, C., Bernsteiner, R., & Ploder, C. (2019). *Artificial Intelligence Tool Penetration in Business: Adoption, Challenges and Fears* (pp. 259–270). [https://doi.org/10.1007/978-3-030-21451-7\\_22](https://doi.org/10.1007/978-3-030-21451-7_22)
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 28, 2503–2511.
- Sebastian-Coleman, L. (2013). Data Quality and Measurement. In *Measuring Data Quality for Ongoing Improvement* (pp. 39–53). Elsevier. <https://doi.org/10.1016/B978-0-12-397033-6.00004-3>
- Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & sons.
- Snoeck, M., Stirna, J., Weigand, H., & Proper, H. A. (2020). Panel discussion: artificial intelligence meets enterprise modelling. *CEUR Workshop Proceedings*, 2586, 88–97.
- Spjuth, O., Frid, J., & Hellander, A. (2021). The machine learning life cycle and the cloud: implications for drug discovery. *Expert Opinion on Drug Discovery*, 16(9), 1071–1079. <https://doi.org/10.1080/17460441.2021.1932812>
- Srinivas, K. (2019). Process of Risk Management. In *Perspectives on Risk, Assessment and Management Paradigms*. IntechOpen. <https://doi.org/10.5772/intechopen.80804>
- Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 272–283. <https://doi.org/10.1145/3351095.3372834>
- Treveil, M., Omont, N., Stenac, C., Lefevre, K., Phan, D., Zentici, J., Lavoillotte, A., Miyazaki, M., & Heidmann, L. (2020). *Introducing MLOps*. “O’Reilly Media, Inc.”
- Tsymbal, A. (2004). The problem of concept drift: definitions and related work. *Computer Science Department, Trinity College Dublin*, 106(2), 58.
- Turner Lee, N. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society*, 16(3), 252–260. <https://doi.org/10.1108/JICES-06-2018-0056>
- van den Heuvel, W.-J., & Tamburri, D. A. (2020). *Model-Driven ML-Ops for Intelligent Enterprise Applications: Vision, Approaches and Challenges* (pp. 169–181). [https://doi.org/10.1007/978-3-030-52306-0\\_11](https://doi.org/10.1007/978-3-030-52306-0_11)
- Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., & Tchatchouang Wanko, C. E. (2020). Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. *Business Process Management Journal*, 26(7), 1893–1924. <https://doi.org/10.1108/BPMJ-10-2019-0411>
- Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, 101685. <https://doi.org/10.1016/j.giq.2022.101685>
- Wu, D., & Olson, D. L. (2010). Enterprise risk management: coping with model risk in a large bank. *Journal of the Operational Research Society*, 61(2), 179–190. <https://doi.org/10.1057/jors.2008.144>

- Xu, L., Jiang, C., Qian, Y., & Ren, Y. (2018). Privacy-Accuracy Trade-Off in Distributed Data Mining. In *Data Privacy Games* (pp. 151–177). Springer International Publishing. [https://doi.org/10.1007/978-3-319-77965-2\\_6](https://doi.org/10.1007/978-3-319-77965-2_6)
- Yokoyama, H. (2019). Machine Learning System Architectural Pattern for Improving Operational Stability. *2019 IEEE International Conference on Software Architecture Companion (ICSA-C)*, 267–274. <https://doi.org/10.1109/ICSA-C.2019.00055>
- Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., & others. (2018). Accelerating the machine learning lifecycle with MLflow. *IEEE Data Eng. Bull.*, 41(4), 39–45.
- Zhou, Y., Yu, Y., & Ding, B. (2020). Towards MLOps: A Case Study of ML Pipeline Platform. *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, 494–500. <https://doi.org/10.1109/ICAICE51518.2020.00102>
- Žliobaitė, I., Pechenizkiy, M., & Gama, J. (2016). *An Overview of Concept Drift Applications* (pp. 91–114). [https://doi.org/10.1007/978-3-319-26989-4\\_4](https://doi.org/10.1007/978-3-319-26989-4_4)

# APPENDIX

## A. CODEBOOK

### Code

- Adaptability
  - adapting changing business needs
  - adapting system changes
  - adapting to concept drift
- Auditability and Compliance
- Business user adoption
- Communication & Collaboration
  - inter-team collaboration
  - intra-team collaboration
- Controls
  - Model explainability
  - Monitoring
  - Reproducibility of results
  - Roles and Responsibilities
  - testing and QA
  - traceability and lineage of artefacts
- Data Quality
  - Data Accuracy
  - Data availability and accessibility
  - Data completeness
  - Data consistency
- Enterprise Integration
  - integration with existing processes
  - integration with existing systems
- ML Decision/Prediction Quality
- Problems with Norms and culture
- risk management
- Shared knowledge
  - Cross-functional teams and inter-disciplinary
  - Data literacy
  - negotiated knowledge
- stakeholder tensions
- Trade-off

## B. INTERVIEW PROTOCOL

### Introduction:

ML operationalization

- Transitioning from experimentation phase (ML POC/experiment) to production phase (production grade application).
- Sustainable management of the entire ML lifecycle after deployed to production.

### Interview Approach:

The interviewer starts with open / reflective question, the interviewer may then probe with more specific questions. Extent of probing depends on the details given by interviewee in open / reflective questions.

### Goal

- To understand the desired state of ML operationalization, and how experts view success in this context.
- To understand the current state of ML operationalization.
- To understand the stakeholders involved, their interest, influence, if there are any conflict of interest.
- To understand the challenges involved in the process
- To understand what enables the process
- To learn new aspects we have not encountered so far in literature ( abduction ).

### Questions:

Could you give an overview of your role and how does ML operationalization relate to you [open] ?

What are the attributes of the successful ML operationalization or when do you consider ML operationalization to be successful ? [reflect]

Who are the different stakeholders involved? How do they influence the process? [open]
--

- |   |
|---|
| <ul style="list-style-type: none"><li>a. Which tech teams are involved ? (infra, dev, ops, platform ) [probe]<br/>What is the responsibility of each team?<br/>- Who does the software engineering around ML model ?</li><li>b. Involvement Business Users? What is of importance to them? ( Some examples of the kind of application )</li><li>c. Which assessment teams/actors are involved ( CISO / risk/compliance)<br/>At what point are they involved ? What all must be explained to them? [probe]</li><li>d. Does the entire solution need to be audited? ( involvement of regulators/auditors) ?<br/>Who are the parties involved? at what point to the ?<br/>What all must be explained to them ? [probe]</li></ul> |
|---|

<p>What according to you are the major challenges in operationalizing ML application?:  [open / reflect]</p> <ul style="list-style-type: none"> <li>e. What challenges do you see specifically with respect to data? How does it affect operationalization? [probe]</li> <li>f. Do you think regulations impose additional challenges to the process ? What are they ? [probe]</li> <li>g. Challenges related to hand-overs/communication between teams? How are the teams structured ? Does it affect the process ? [probe]</li> <li>h. Do you find it challenging to explain the solution during the assessment? why ?</li> </ul>

<p>Once in production, do you still face challenges ? [ reflect]</p> <ul style="list-style-type: none"> <li>a. When new production issues occur, do you find it challenging to resolve the source of the problem ( is it data/code / model ?)</li> <li>b. When prod issues occur, is it easy to get the right person invovled ( DSE/ MLE/ platform engineer/ data engineer ) ?</li> <li>c. Can the business users, use the model outcome well ? if not why ?</li> <li>d. Is model reliability a challenge ( initially and over a period of time) ?</li> <li>e. Do you find it challenging to implement a feedback loop for retraining and redevelopment of the model? why ? [probe]</li> <li>f. Summarize the challenges. List Success factors. Could you identify how do the challenges affect the success factors ?</li> </ul>
<p>What enables ML operationalization ?</p>

( if time allows)

What are the best practices during development and deployment ? :

- a. Are the data science experiment tracked ? [probe]
- b. Is the data versioned ? [probe]
- c. Are the models versioned in repository ? [probe]

Is there any other question that you think I should have asked you ?

## C. RESEARCH ETHICS

Research that involve human subjects, requires by the Human Research Ethics Committee (HREC) of TU Delft. The process ensures that the human subjects are comfortable and that the data collection process follows GDPR standards. It proceeds for this research, as:

1. The HREC uses a checklist to check the level of risk involved in the resarch. The checklist asks questions are participants part of vulnerable groups, does research involve deceiving the participants, can collecting samples from participants cause discomfort/ stress.
2. Since the research involved interviews and collecting and storing audion/videos or other identifiable data of human subjects, Suitable informed consent forms are needed to be sent across to potential participants prior to their interviews. The informed consent forms comprise of two parts: information sheet and consent form options. The consent forms inform the participants about the purpose of the interview as well as their right to withdraw from the interview. They also inform the participants on how their data will be stored, used and protected.
3. Additionally, a data management plan (DMP) needs to be sent across after being vetted by the Data Supervisor of the concerned faculty. The DMP includes details on how the data will be collected, whether it is stored on secure servers or in a secure location. The data collected for this research was saved on the TU Delft Onedrive account of the author. This ensured the data was safe, not stored in an offline location and can only be accessed by the author of this thesis.
4. The checklist and consent forms have to be submitted to an online portal of the HREC. The committee meets once a week, debates on applications and either accepts an application or proposes changes. This research went through one revision which involved providing the exact details of the DMP. After the application was accepted, the interview process was started.

## **Information sheet for “Socio-technical framework for operationalizing machine learning in the banking sector” – 25/02/2022**

### **Purpose of the research:**

This research investigates the different socio-technical factors involved in converting machine learning (ML) pilots into production applications in the banking sector. The research aims to identify and analyze the opportunities and challenges banks face in the process of operationalizing ML.

### **Interviews:**

The participant will take part in a one-hour interview to provide insights into the ML operationalization process. These interviews will take place online or in-person depending on mutual agreement between the researcher and the participant.

### **Benefits and risks of participating:**

The participant can benefit from the interview since it gives them the opportunity to critically analyse, reflect and rethink their practices. As we are interviewing participants from different backgrounds, the diversity of the questions can help participants approach their practices from a holistic perspective. Risks of participating include mentioning information that is private for the participant's employer. While the participants are encouraged to share their knowledge and experience, they are requested to avoid sharing confidential details of the employer.

### **Procedures for withdrawal from the study:**

The participant can withdraw from the interview at any time. At any time, the participant can ask for their data to be destroyed by contacting the researcher.

### **Collection and use of personal information:**

All the personal information of the participant will be anonymized. The data collected will be used for understanding the ML operationalization process from different perspectives. The participant can, at any time, request access to or rectification or erasure of such personal data.

### **Research data:**

The audio recordings will be transcribed and anonymized into text transcripts, that will be only shared with the researcher working on the project. The audio recordings will be destroyed after transcription. In the planned scientific study, the researcher will draw insights from the interviews to identify and justify the factors that enable or hinder ML operationalization. The researcher may use anonymous quotes from the transcripts in their academic output.

### **Data retention period:**

The audio recording will be transcribed and destroyed within 1 month after the interview. The retention period for the anonymized transcripts will be of approximately 4 months i.e. until the thesis project is completed.

### **Contact details:**

Yash Singh, Y.S.Singh@student.tudelft.nl

*Figure 16: Consent Form Page 1*

## Consent Form for “Socio-technical framework for operationalizing machine learning in the banking sector”

### Taking part in the study

I have read and understood the study information sent to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.

I understand and agree that taking part in the study involves audio-recorded interviews, that will be transcribed as anonymized text later, and the recording will be destroyed.

### Use of the information in the study

I understand that information I provide will be used for writing an academic thesis.


I understand that personal information collected about me that can identify me, such as my name, employer name, and my work details, will not be shared beyond the interviewer.

I understand that the anonymized insights drawn from the interview will be used to inform the academic thesis research which will be published on the TU Delft educational repository.

### Signature of the Researcher

I have informed the potential participant to the best of my ability, ensuring that the participant understands to what they are freely consenting.

Yash Singh  
Researcher name

  
Signature

18/03/2022  
Date

Study contact details for further information: Yash Singh, +31644576818,  
Y.S.Singh@student.tudelft.nl.

Figure 17: Consent Form Page 2

## D. Description of Socio-technical factors

### D.1 Successful Outcomes

There is little explicit evidence in literature, of what the successful outcome of an ML operationalization process are. Without a fundamental understanding of this concept, organizations struggle to formulate clear goals and provide direction to their ML initiatives. Therefore, was derived from the expert interviews, where the experts were explicitly asked when do they consider the ML operationalization process to be successful, which contributed the following sub-factors:

*Senior Data Scientist: “In terms of ML operationalization, I think it is a success when the model is being used by the intended users for the intended purpose.”*

*Specialist Lead (MLOps): “Will your ML solution actually be used by your end user, and will they actually be able to realize their value. I think adoption is key because if your users don't adopt it, and are not going to make decisions based on the insights generated by the model, then it cannot be a success”*

- 1) **Business User Adoption:** refers to the ML application being used by the intended use for its intended purpose. For the operationalization process to be successful, the ML application must be adopted by the user, without which it has no business value.
- 2) **Quality of the ML decision or prediction:** refers to the correctness of the decision or prediction made by the ML application. A number of metrics are used to gauge the decision quality, these include accuracy, precision, recall, F1-score, ROC and many more (Flach, 2019). In essence, the decision made by the ML application must reflect reality and should be free from biases. Poor decision quality can lead to increased costs, have unethical consequences, and can also be detrimental to the reputation of the banks.
- 3) **Compliance:** ML applications should be operationalized in such a way that, they are compliant with all the applicable laws, regulations, and ethical standards. In highly regulated institutions such as banks, compliance is given high priority and requires that ML applications are compliant general regulations such as GDPR but also with regulations that apply to the business application case for which the ML is used.

### D.2 Data Quality

Data quality is a measure of how well the information is suited to fit its purpose. Data quality represents the quality of the information produced by the information system, on dimensions such as accuracy, timeliness, completeness, consistency, and relevance (Janssen et al., 2017; H. Miller, 1996). Data quality plays a critical role throughout the ML operationalization process and has a significant impact on the quality of the decisions made by ML. Based on data quality dimensions, the following sub-factors were identified specifically in the context of the ML operationalization process.

*Enterprise Advisor (Advanced Analytics): “It is important to make data available and ensure it is fit-for-purpose data . It is also quite a challenge. Do we have labeled data? Can I know beforehand know, that the data is statistically sound so that I don't invest months to experiment with several models and come back saying that the data is not fit for purpose? To ensure that the data is available and has good quality is quite a challenge. At this point, most organizations struggle with ensuring even the traditional data quality KPIs. It is relatively easier to procure historical data for model preparation, but getting the actual data for your*

*model to perform in production is challenging. It is often not available, not accessible, and not fit for purpose.”*

- 4) **Data accuracy:** refers to the extent to which the data is correct, reliable and representative of the underlying reality. The data on which the model is trained must not only be free of inaccurate values and errors but should have a distribution that reflects reality. Banks must therefore in order to ensure high accuracy of the ML model (Kaddoumi & Tambo, 2022; Renggli et al., 2021).
- 5) **Data Availability and accessibility:** Availability of labeled data sets is crucial to train ML models and can often be a challenge. Accurate data labeling in the financial sector often requires human experts and can be a cumbersome and time-consuming process (Renggli et al., 2021). Further, even if data is available in many cases it is not easily accessible due to constraints relating security and privacy, thereby making model development a challenging process (Lwakatare et al., 2020).
- 6) **Data consistency:** represents if the data stored in different systems in a consistent and uniform manner. With respect to ML operationalization the, the model is developed and deployed in different environments. Therefore, the data on which the model is trained, tested and must be consistent with the data in the production environment in order produce reproducible results (Ruf et al., 2021; Sebastian-Coleman, 2013).
- 7) **Data Completeness:** The data on which the ML model is trained must provide comprehensive information. Incompleteness may manifest itself in the form of missing value entries (Janssen et al., 2017; Ruf et al., 2021), or in the form skewed data sets that can lead to biased ML models (Miceli et al., 2022).

### D.3 Adaptability

The ML operationalization process must facilitate the adaptability of the ML application. Adaptability comes from the ability to iterate over the operationalization activities in order to respond to changes in a timely manner.

*Senior Data Scientist : “How you adapt to the change is quite a challenge. So for example, if certain models were developed before Covid that changes a lot of you’re the assumption you made during the development of the model. so that I would say is the changing external environment. The second things is changing internal environment, where the IT systems within the bank change. For example if you create a model with some data points based on a particular IT system. But data is now migrated to a new IT system, and now there is isn’t the same mapping of data points, then it lead to errors in your model. Therefore you need to take care of internal and external change.”*

- 8) **Adapting to system dependencies:** ML systems exhibit complex interactions at a system level, since they depend on other systems for data and input signals. Unlike traditional software applications, it is difficult to create strict encapsulations to isolate the ML application (Sculley et al., 2015). Therefore, in events when there are changes in the dependent system ( business rules, data format, connectivity, etc) the ML application must be able to detect and adapt to those changes.
- 9) **Adapting to business needs:** ML applications are rarely all-encompassing in the first deployment, and require continuous involvement and feedback from business users (Q. Z. Chen et al., 2022; Guo et al., 2022). Further business needs themselves are can change over time, as a result of which business user may request incorporation of new assumptions, or inclusion of new features to improve the performance or prevent the degradation of the model.
- 10) **Adapting to Concept drift:** The nature of real-world concepts (customer preferences, financial transactions, etc) are not stable and can change with time, as a result, the underlying data distribution might

change as well (Tsymbal, 2004) . Therefore ML models must adapt to concept drift either through retraining or re-experimentation in order to prevent the performance degradation of the ML model.

## D.4 Enterprise Integration

Introducing the ML applications in the core operations and processes of an organization requires a well-managed integration with the existing enterprise architecture (Kaddoumi & Tambo, 2022). In the case of the banking industry ML applications are commonly introduced in business functions which consist of pre-existing systems and processes.

*Specialist Lead (MLOps): “If your solution cannot be technically or operationally integrated into adjusting business process, then is it going to make the life for your business team easier or more complex? So when you're designing the proof of concept, and you think it has a lot of potential, let's operationalize it. You need to think about what are the technical constraints and the work environment in your model will be used, and you need to build towards that.”*

- 11) **Integration with existing system:** ML applications in their design and deployment must be integrated with other dependent systems. For example, developers of the application may need to consider the business rules and data format of the legacy systems from which the data must be read to train or score the ML model. Such integration can be instrumental in reducing data quality issues faced during the process. As another example, the ML application may also need to be integrated with the existing systems or interfaces on which the data must be displayed in order to make the result of the model easily accessible and support business user adoption (Brock & von Wangenheim, 2019).
- 12) **Integration with existing business processes:** Introduction of ML applications in a core business function it should bolster the workflow and not disrupt it. This may require designing the model to fit the existing business activities or reconfiguring the business activities to use the model. In many cases, the ML application results must be integrated with the existing business rules instead of completely replacing them. Additionally, developers of the ML model must also consider how the human workforce should be facilitated to will interact with the ML applications and use its results (S. M. Miller, 2018).

## D.5 Risk Management

The banking industry lays a strong emphasis on various risks associated with its business operations, products and services. It is necessary that all the relevant risks related to ML applications are identified and mitigate before the application is deployed to production. The finding suggest that, use of ML techniques often introduces unfamiliar risks, which is a challenges for the traditional risk management function in banks. To categorize different risks, each bank uses its own risk taxonomy. Based on the expert interview two main categories that apply to the ML operationalization process are operational risk and ethical and reputational risks.

- 13) **Operational risks:** The European banking authority defines operational risk as ‘the risk of losses stemming from inadequate or failed internal processes, people and systems or from external events. Operational risk constitutes model risk, process risks, regulatory risk, and IT risk. But it excludes reputational risk.
  - a. **Model Risk Management:** Refers to the management of risk arising from incorrectly implemented quantitative models. This can pertain to the use of questionable assumptions or assumptions that don’t hold any longer (Wu & Olson, 2010).

- b. **Process Risks Management:** Refers to the management of risks that arise from ineffective. In the context of ML operationalization, this refers to the risks arising from introducing ML applications into a business process. Expert participants mention that these risks may arise due to a lack of interpretability of the ML results, or inefficient integration with the business rules and existing systems.
- c. **Regulatory Risk Management:** Refers to the management of risks arising from incompliance of the ML application with the applicable laws and regulations (de Nederlandsche Bank, 2019). Inability to comply with laws and regulations can expose banks to legal penalties and forfeiture.
- d. **IT and Data Risk Management:** ML operationalization can face risks arising from system failure, data leakage and breach, and cyber-attacks (Wirtz et al., 2022). The management of these risks is referred to as IT and Data risk management.

14) **Ethical risks and reputational risks:** Refer to risks that could violate the ethical standards of the bank, lead to loss of trust among stakeholders, and potentially damage the reputation of the bank (de Nederlandsche Bank, 2019). Specifically, in the case of ML, these risks are often associated with principles such as fairness, accountability, and transparency. Given the rapid change in the societal expectations around the use of ML techniques, ethical and reputation risks have been of primary concern for many banks.

## D.6 Shared Knowledge

In information system research, shared knowledge is considered to be an important contributor to group performance. Shared knowledge refers to developing an appreciation and understanding of others environment rather than merely sharing information and translating technical and procedural terms (Nelson & Coopridge, 1996). Effective shared knowledge facilitates a common language and synergy between different social groups and plays a crucial role in determining the how diverse stakeholders interact with each other throughout the ML operationalization process.

*Product Owner (Data Science): “As a business teams, if you're not familiar with thinking in data terms or if you don't know what the model can do. Maybe you're overestimating, or maybe you're underestimating the possibilities of machine learning. And that that also both ways because if the data scientists don't have deep understanding of business and its problems, they think about a very sophisticated solution which is not practically possible. So both ways you need to get that understanding of each other and get an understanding for the possibilities, the impossibilities and the problems.”*

15) **Cross-functional knowledge:** Operationalizing ML applications require a multitude of expertise and cross-function knowledge ranging from data science, data engineering, software development, regulatory compliance and business domain expertise. Such knowledge must therefore be incorporated through cross-functional teams or through close collaboration with different stakeholders throughout the different stage of the ML operationalization process (Karamitsos et al., 2020).

16) **Negotiated Knowledge:** ML applications introduce uncertainties, and there is often a lack of uncontested and objective information. During the operationalization process, when different parties have different perspectives based on their relevant expertise, the parties establish the right information based on negotiation with each other to reach a compromise on various aspects such as what tools will be used, how will the responsibilities be distributed, and what will be the format of the assessment. Such an established knowledge is referred to as ‘negotiated knowledge’ (de Bruijn & ten Heuvelhof, 2018).

- 17) **Collaboration:** The stakeholders must collaborate based on shared interests and goals. Effective and continuous collaboration among technical teams primarily the IT engineers, data scientists and data providers is crucial for the development, deployment and maintenance of an integrated ML application (Ruf et al., 2021). Further, these technical teams also need to collaborate with business domain experts for requirements and feedback, and with the risk teams in order to jointly assess the ML application before it can be deployed to production.
- 18) **Communication and Translation:** Literature and expert interviews indicate that communication between technical stakeholders is crucial during the ML operationalization process. The ML operationalization process requires the technical teams to collaborate with non-technical stakeholders from business and risk (de Bruijn et al., 2021; Wirtz et al., 2022) . Therefore, the technical teams must be able to effectively translate of technical aspects of the ML to the non-technical stakeholders in the process.

## D.7 Controls

Controls refer to a set of mechanism or policies that are needed in order to ensure that the ML application is deployed and operated in reliable manner. Controls server as an important purpose of preventing, detecting and mitigating risks, but also contribute the operational efficiency of the process and the performance of the ML application.

*“When you are doing a POC, your machine learning is sort of this independent project. When it's in production you have a machine learning application that lives in your system. So it needs to be very clear where is it getting its data from, how much data is it getting, what's the frequency, what happens with the data in your pipeline. Everything has to be very transparent from end to end. ML solutions have many moving parts, it helps when you have explainability in the whole solutions so making an end-to-end pipeline that is really transparent helps in ensuring the root cause and problems in code, data and model can be detected and traced.”*

*-Product Owner, Data Science*

- 19) **Monitoring:** is essential to detect and track how the ML application is performing post-deployment. However, good performance means different things to different stakeholders (Treveil et al., 2020). The IT engineering teams are focused on monitoring the technical metrics such resource consumption and, response time. Whereas the data scientist are interested in the model metrics such as accuracy, drift detection, etc. The business on the other hand monitor broader implications such as business value and the operational cost of the application. Stakeholders from the risk are interested in monitoring the usage of ML applications (Wirtz et al., 2022).
- 20) **Versioning and Lineage of Artefacts:** Versioning of the different artifacts ( code, data, experiments, and models) generated throughout the operationalization process and maintaining a lineage of the same creates a systematic and transparent overview of the end-to-end ML solution (Ruf et al., 2021; Zaharia et al., 2018). This facilitates reproducibility of results, efficient troubleshooting of problems, and bolsters the auditability of the process.
- 21) **Testing and Quality Assurance:** is important to conduct necessary tests to prevent errors in the application and the undesirable behavior of the ML model in production. Given its integrated nature, ML operationalization requires the testing of software components, validating data sources and pipelines, and the performance of the model (Schelter et al., 2018; Sculley et al., 2015).
- 22) **Model Explainability and Fairness:** These refer to model specific controls, primarily to deal with the possible opaqueness and biases of ML algorithms. A number of techniques such as SHAP (SHapley Additive exPlanations) and LIME (local interpretable model-agnostic explanations) are available to

develop explainable and interpretable ML models. A number of practices and techniques have also been developed to aid the development of fair ML models (Enni & Herrie, 2021; Mehrabi et al., 2021).

- 23) **Roles and Responsibilities:** ML applications developed through collaboration between different actors for various components. A clear description of roles and responsibilities is required during the ML operationalization process in order to bolster the development efforts and guarantee a degree of ownership and responsibility post the deployment (Ruf et al., 2021; Treveil et al., 2020).