An empirical guide to MLOps adoption: Framework, maturity model and taxonomy

Citation for the original published paper (version of record):

John, M., Olsson, H., Bosch, J. (2025). An empirical guide to MLOps adoption: Framework, maturity model and taxonomy. Information and Software Technology, 183.
http://dx.doi.org/10.1016/j.infsof.2025.107725

Information and Software Technology 183 (2025) 107725

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

An empirical guide to MLOps adoption: Framework, maturity model and taxonomy

Meenu Mary John [a,*], Helena Holmström Olsson [a], Jan Bosch [b]

[a] Department of Computer Science and Media Technology, Malmö University, Malmö, 211 19, Sweden
[b] Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, 417 56, Sweden

ARTICLE INFO

Keywords:
MLOps
Framework
Maturity model
Taxonomy
Multi-case study

ABSTRACT

Context: Machine Learning Operations (MLOps) has become a top priority for companies. However, its adoption has become challenging due to the need for proper guidance and awareness. Most of the MLOps solutions available in the market are designed to fit the specific platform, tools and culture of the providers.
Objective: The objective is to develop a structured approach to adopting, assessing and advancing MLOps adoption.
Methods: The study was conducted based on a multi-case study across fourteen companies.
Results: We provide a comprehensive analysis that highlights the similarities and differences in the adoption of MLOps practices among companies. We have also empirically validated the developed MLOps framework and MLOps maturity model. Furthermore, we carefully reviewed the feedback received from practitioners and revised the MLOps framework and maturity model to confirm its effectiveness. Additionally, we develop an MLOps taxonomy for classifying ML use cases based on their context and requirements into the desired stage of the MLOps framework and maturity model.
Conclusion: The findings provide companies with a structured approach to adopt, assess, and further advance the adoption of MLOps practices regardless of their current status.

1. Introduction

Companies are constantly interested in embracing innovations [1] to achieve phenomenal growth. One of the disruptive innovations with the potential to impact business is Machine Learning (ML) [1], a subfield of Artificial Intelligence (AI). ML allows for the identification of patterns, making predictions, and evolving based on new and unseen data [2]. It has applications in different fields. For example, healthcare, telecommunications, manufacturing, insurance, banking, financial services, automotive, and energy [1,3].

Many companies assign high priority and spend time, effort and resources in the development and deployment of ML but often struggle to deliver expected business value [4,5]. This is due to the unsuitability of the linear and sequential workflow of traditional Software Development Lifecycle (SDLC) for managing the iterative nature of ML workflows [6,7]. The continuous exploration of data, experimentation and refinement of the model in ML projects, as well as its cyclical and exploratory process, cannot be handled by traditional approaches [6]. Furthermore, manual management of most parts of ML workflow in projects leads to deployment and operations issues [8,9]. Practitioners often spend a significant amount of time on low-level manual tasks (for instance, data wrangling and hyperparameter tuning), which introduces biases into the ML pipeline rather than focusing on high-level activities [6] (for instance, developing production-ready ML models) [8].

To address the issues mentioned above, companies have started adopting MLOps [10] to unify the development (Dev) and operations (Ops) of ML systems [11]. Various tools and platforms introduced in the market can promote the adoption [9,12]. However, many companies are hesitant to perceive and adopt MLOps [13]. Consequently, they adopt similar approaches due to a lack of awareness of best practices in the field [13]. There is a significant disagreement between researchers and practitioners in operationalising ML models [14,15] despite the fact that ML activities are organised in various phases using different approaches. This demands a need for a better understanding of different perspectives on MLOps, including best practices, platforms, tools and roles [8]. The lack (or very few) of scientific studies focusing on MLOps highlights the need to develop structured approaches based on practical experiences from companies [12,16,17]. Therefore, this is an opportunity to develop an MLOps framework, maturity model, and an MLOps taxonomy to adopt and assess MLOps practices.

* Corresponding author.
E-mail address: meenu-mary.john@mau.se (M.M. John).

https://doi.org/10.1016/j.infsof.2025.107725
Received 1 October 2024; Received in revised form 12 March 2025; Accepted 13 March 2025
Available online 22 March 2025

In our previous study, we examined adoption of MLOps practices in seven companies. Based on our findings, we developed a five-dimensional MLOps framework, with each dimension representing five stages in our developed MLOps maturity model. Furthermore, we also mapped the companies we studied to specific stages within the MLOps maturity model. This work was published as *"Advancing MLOps from Ad hoc to Kaizen"* in the 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) [18] in 2023.

The current study extends beyond the initial findings presented in our above-mentioned previous work [18] in the following ways: (a) It broadens the scope of the study to fourteen companies to empirically validate the developed MLOps framework and MLOps maturity model and refine them based on feedback from practitioners, and (b) It introduces an MLOps taxonomy to enhance the utility of the MLOps framework and maturity model. This taxonomy classifies ML use cases to the desired stage of the MLOps framework and maturity model based on their context and requirements. The findings offer comprehensive MLOps guidelines for companies and enable them to scale their practices more effectively.

The contribution of the paper is threefold.

- Providing comprehensive analysis for MLOps adoption across different companies.
- Creating a blueprint to adopt, assess and advance MLOps practices.
- Developing a taxonomy to help companies in tailoring MLOps practices to fit their specific contexts.

The paper is structured as follows: Section 2 provides an overview of the background, and Section 3 presents related work. Section 4 outlines the utilised research methods. Section 5 summarises the empirical findings: Analysing the adoption of MLOps practices — Similarities and differences, Empirical validation of the MLOps framework and Maturity model, and MLOps taxonomy. Section 6 discusses our findings. Section 7 presents the potential threats to validity. Finally, Section 8 concludes the study.

2. Background

2.1. ML workflow

ML workflows have been extensively represented in different forms across various research and company settings [19–21]. For instance, Microsoft has proposed a nine-stage workflow that includes data-oriented stages, model-oriented stages, and multiple feedback loops [22]. The stages are: "(a) Model Requirements, (b) Data Collection, (c) Data Cleaning, (d) Data Labelling, (e) Feature Engineering, (f) Model Training, (g) Model evaluation, (h) Model deployment, and (i) Model monitoring". However, the process of experimenting, developing, deploying, monitoring, maintaining, and documenting ML models is considered challenging [23]. Some notable concerns include operational support, mixed team dynamics, access to experts, continuous delivery, ethical considerations, focus on technical solutions, country-level regulations and involvement of end-users [24].

2.2. MLOps

The challenge of transitioning developed ML models into production [25], as well as the impact of the paper "Hidden Technical Debt in Machine Learning Systems" [26], has greatly influenced the emergence and advancement of MLOps. As a result, researchers and practitioners are adopting MLOps as a practice similar to DevOps (aiming to deliver operational features in a faster and continuous manner [27]) to maximise the benefits. Technically, MLOps can be defined as "a development methodology aimed at bridging the gap between Development (Dev) and Operations (Ops), emphasising communication and collaboration, continuous integration, quality assurance and delivery with automated deployment utilising a set of development practices" [28]. It aims at standardising and streamlining the ML lifecycle, which is critical given the complexity of managing the ML lifecycle in companies [10]. The adoption of MLOps offers several benefits, including [29]: (a) Reduced cycles of development, (b) Enhanced collaboration between team members, (c) Making ML systems more reliable and scalable, (d) Optimised processes for operations and governance, and (e) Increased profit of ML projects.

3. Related work

3.1. MLOps methods, processes, tools and adoption challenges

As ML has become prevalent in software products, it is crucial to establish best practices and tools for deploying, managing, and monitoring ML models in real-world production [30]. We outline the methods, processes, tools and challenges associated with MLOps adoption in Table 1.

3.2. MLOps maturity models

This section discusses three well-known MLOps maturity models proposed by Microsoft, Google and Amazon [11,36,37]. The Microsoft MLOps maturity model consists of five levels of maturity and provides a step-by-step progression: (a) No MLOps, (b) DevOps but no MLOps, (c) Automated Training, (d) Automated model deployment, and (e) Full MLOps automated operations. On the other hand, the Google MLOps maturity model is simpler with three levels of maturity: (a) MLOps level 0: Manual process, (b) MLOps level 1: ML pipeline automation, and (c) MLOps level 2: CI/CD pipeline automation. Furthermore, the Amazon maturity model has four phases: (a) Initial phase, (b) Repeatable phase, (c) Reliable phase, and (d) Scalable phase. The MLOps maturity models from Microsoft, Google and Amazon can guide companies in transitioning from manual to fully automated ML pipelines. However, the shortcomings are: (a) It advises a one-size-fits-all maturity model which needs to be adjusted based on each company, (b) It is not applicable for companies that want to skip or prioritise a specific stage based on their maturity, (d) It assumes that adopting companies have the required size, infrastructure and budget to progress through the model, (e) It does not consider external constraints (For instance, regulations) that may affect the ability of companies to adopt MLOps.

4. Research methodology

Below, we outline the research questions (RQs), the research design chosen to answer these RQs, and the research methods for data collection and analysis.

4.1. Research questions

Our study addresses the following RQs to adopt, standardise, assess and advance the MLOps practices within companies.

- RQ1. How does the adoption of MLOps practices vary across companies?
- RQ2. How can an MLOps framework, maturity model, and taxonomy be designed to address evolving ML technologies and business needs in companies?
- RQ3. What variations or adaptations are necessary to ensure the applicability of the MLOps framework and maturity model in different company contexts?

4.2. Research design

To address the RQs mentioned above, we have chosen a multi-case study research. A case study involves an understanding of a specific

Table 1
Overview of the methods, processes, tools and challenges associated with MLOps.

MLOps elements	Description
Methods	- CI/CD automation [31] for Continuous integration, delivery and deployment [8][9] [32] - Workflow orchestration to manage tasks of an ML pipeline [8] - Ensuring reproducibility and versioning of data, model and code [8] [33] [32] - Collaborative and communicative work culture [8] - Continuous training and monitoring of ML system [8] [9] - Tracking and logging of metadata [8] - Establishing multiple feedback loops [8] - Apply SE principles to ML workflow [15] - Consider interaction between MLOps and existing practices in companies [15]
Process	- An End-to- end MLOps architecture [8] (a) MLOps product initiation steps (b) Feature engineering pipeline, (c) Experimentation, and (d) Automated ML workflow pipeline up to the model serving - Proposed an ML pipeline platform with [31]:(a) CI/CD pipeline, (b) Kubeflow pipeline, and (c) ML platform – Proposed MLOps workflow inspired from CRISP-DM with stages [9]: (a) Business problem understanding, (b) Data Acquisition, (c) ML methodology, (d) ML training and testing, (e) Continuous Integration, (f) Continuous Delivery, (g) Continuous Training, (h) Continuous monitoring, (i) Explainable AI, and (j) Sustainability
	– Proposed an end-to-end process for Continuous Delivery for Machine Learning (CD4ML) [32]
Tools	- TensorFlow Extended, Airflow, Kubeflow, MLflow, Databricks managed MLflow, Amazon CodePipeline, Amazon SageMaker, Azure DevOps Pipelines, Azure ML, GCP - Vertex AI, IBM Cloud Pak for Data, gitea, Drone, DotScience, Gitlab, Jenkins, Google AI platform, Polyaxon, Seldon Core, Valohai, DVC, Pachyderm, Torch Serve, Weight Biases [8] [9] [17] [31]
Adoption challenges	- Cultural shift from model-driven ML towards product-oriented discipline [8] - Establishment of multi-disciplinary team [8] - Educating for AI Operations [34] - More importance on technical perspectives and significantly overlook the organisational and social perspectives [35]

individual, group, organisation(s) or phenomenon [38]. It utilises various sources of information [39] and is widely recognised in multiple scientific fields [40] and among researchers [41]. For example, a case study could be conducted on empirical software engineering (SE) studies within a company [42]. The insights from the case study can be applied to similar phenomena [43] beyond the specific case being studied. The case study method helps in exploring "how" and "what" questions in our RQs [44], which require in-depth analysis or documentation of MLOps adoption in companies. Fig. 1 depicts the overall research process.

4.3. Data collection

4.3.1. RQ1 - Adoption of MLOps practices across companies

To understand how MLOps practices are being adopted across companies, we conducted a multi-case study involving a selection of companies in AI Sweden [45]. AI Sweden is an initiative with 120 partners from the public sector, private sector, and academia, with an intention to develop tools and resources to increase the application of AI in Sweden through projects, talent programs, courses and informative sessions [45]. To gather details on the adoption of MLOps practices, we conducted semi-structured interviews [46] in these selected companies. Interviews are a frequently used technique in empirical SE research to collect data about a phenomenon that is difficult to obtain using quantitative measures [47]. We designed an interview guide with questions related to the end-to-end ML workflow (from data and features to model development, deployment, and operations) based on existing scientific literature. The guide also includes questions about the organisation and the challenges they face when adopting MLOps. We asked a key person from AI Sweden to provide feedback on the interview guide to verify its completeness and then incorporated the feedback. The first author conducted trial interviews with three to four colleagues to ensure that the interview guide could be completed within an hour. The interview guide is available in Appendix A.

After finalising the interview guide, we began searching for companies relevant to our study with the help of responsible contacts in AI Sweden. We sent emails to key contacts in these selected companies by explaining the objective and purpose of the interview study. Once we received a final list of interviewees from each company, we scheduled one-hour online interviews between February and April 2023 via video conferencing using Microsoft Teams. Seven out of 120 companies in the statistics, banking, healthcare, pharmaceuticals, manufacturing, gaming, and mining domains have expressed their interest in participating in interviews. We selected these seven companies as they were considered representative of the broader population [48]. The second and third authors also participated in selected interviews. Before asking questions, we obtained consent from the interviewees to record the interviews for further analysis. During our interviews with practitioners, we requested an additional five to ten minutes, depending on their availability, to ensure complete coverage of the interview guide if we were short on time. Additionally, when faced with time constraints, we chose not to ask the supplementary questions included in the interview guide. The details of the interviewees (represented by P*) are shown in Table 2.

We studied different use cases in seven companies.

- Company A: A use case to predict the probability of individuals answering phone calls to schedule interviews.
- Company B: A use case that prioritises Environmental, Social and Governance factors in the execution of a bank.
- Company C: A use case that involves identifying and highlighting cancerous areas.
- Company D: A use case that analyses medical images to draw conclusions.
- Company E: A use case that developed a speech-to-text app that automatically tags vehicle maintenance issues.
- Company F: A use case in which ML bots automatically play new levels of game for assessment before release.
- Company G: A use case that monitors the health and sends alerts for the devices at various mining locations.

4.3.2. RQ2 and RQ3 - Development and validation of MLOps framework, maturity model, and (development of) taxonomy

We have developed an MLOps framework, maturity model, and taxonomy based on the empirical findings from different companies in AI Sweden. The MLOps framework and maturity model help companies assess their current level of MLOps adoption and ways to advance for future growth. The MLOps taxonomy classifies the ML use cases into the desired stage of the MLOps framework and maturity model. To validate the developed MLOps framework and maturity model, we organised a validation MLOps workshop involving six companies in the Software Center [50]. The Software Center collaborates with sixteen companies and five universities, with a focus on improving the

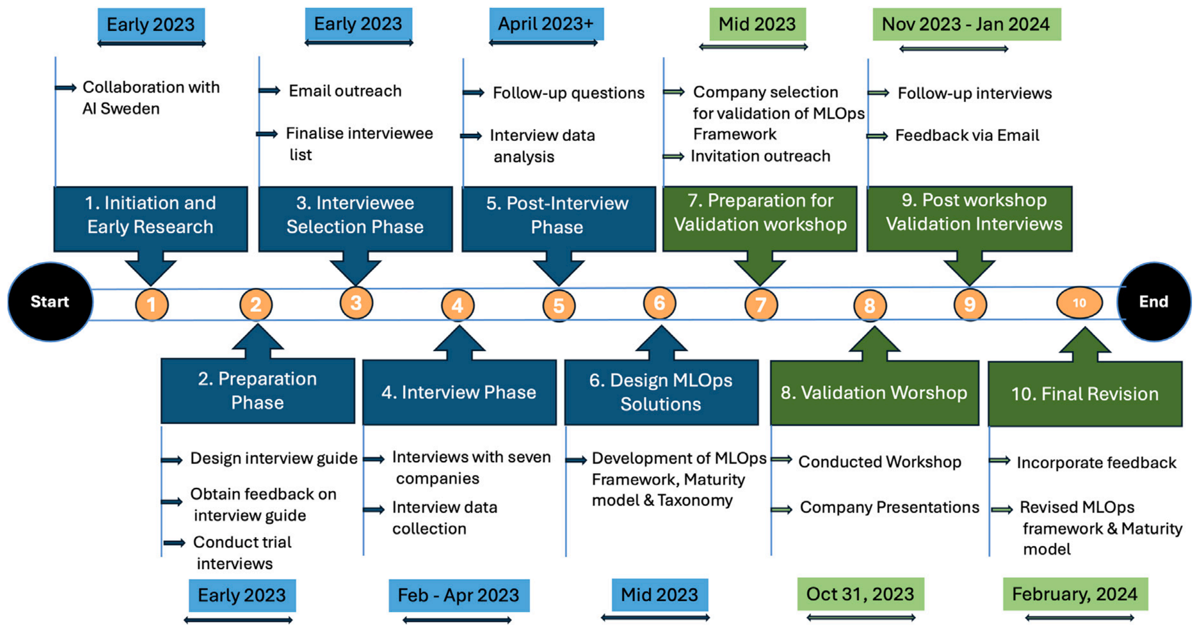

Fig. 1. Overall research timeline.

Table 2
Description of interviewees involved in Phase I [18] [49].

Company	Description	Practitioners		Date
		ID	*Roles*	
A	Statistics	P1 P2	Business developer Statistical expert	2023/2/21 2023/3/3
B	Banking	P3	Chief data scientist	2023/2/21
C	Healthcare	P4 P5	Head of AI competence centre Product manager	2023/2/22 2023/2/22
D	Pharmaceutical	P6	AI team lead	2023/3/2
E	Vehicle manufacturing	P7	Head of data science chapter	2023/2/23
F	Gaming	P8	Lead of AI Center Excellence	2023/4/5
G	Mining	P9	Data scientist	2023/4/13

digitalisation capabilities of companies within Europe [50]. One of the main themes in Software Center is AI Engineering. The second and third authors worked together to identify suitable companies for the workshop and provided the first author with the key contacts of each company. This helped in sending invitations to the key contacts and asking them to forward the invitation to suitable practitioners. We conducted a two-hour MLOps workshop on October 31, 2023, where each company presented its current status of MLOps adoption, providing an opportunity to share insights and ask questions. Twelve practitioners from six different companies attended the MLOps workshop. The details of the MLOps workshop participants are summarised in Table 3.

To gather more details, the first author requested the opportunity to conduct separate validation interviews with companies during the validation workshop. Later, the first author sent invitations to practitioners in companies via email. Four out of six companies from the MLOps workshop agreed to this invitation. Additionally, one company that could not attend the MLOps workshop agreed to an interview. Below, we present the ML/DL use cases studied in five companies.

- Company H: A use case in which an internal MLOps platform is utilised to streamline and automate the end-to-end ML lifecycle.
- Company I: A use case that trains a simulation model to be deployed in a test rig.
- Company J: A use case involving the use of microcontrollers inside embedded products to obtain training data.
- Company K: A use case that provides embedded solutions.
- Company M: A use case that involves a service solution that can be used to store operational data and enable customers to make informed decisions.

We conducted a semi-structured interview study to validate the MLOps framework and MLOps maturity model empirically. The validation interviews were conducted between November 2023 and January 2024, each lasting 50 to 60 min. We used an interview guide for the validation study, which explained the objective of the validation interview, the developed MLOps framework, and the MLOps maturity model. We then asked for feedback, which was supported by some follow-up questions. This is made available in Appendix B. One out of five companies sent their validation feedback via email. Based on the validation feedback, we revised the MLOps maturity model and framework. The details of the validation interviewees are shown in Table 4.

4.4. Data analysis

During the interviews, we transcribed all interview recordings using Microsoft Teams. We then analysed the transcript using elements of open coding [51] and triangulation [52]. The first author discussed the interesting and relevant findings with the other authors and reached a consensus. We used the interview findings as a reference for developing and validating the MLOps framework, maturity model and taxonomy.

Table 3
Details of MLOps workshop participants involved in Phase II.

Company	Description	Practitioners		Date
		ID	*Roles*	
H	Telecommunications	VP1	Expert in Network Architecture evolution (AI/ML)	2023/11/10
		VP2	Main Technical Coordinator	2023/10/31
		VP3	Specialist in ML and Data Science	2023/10/31
J	Home appliance manufacturing	VP6	Software Engineer	2023/10/31
K	Aviation	VP7	Head of Signal and Data processing Applications	2023/10/31
		VP8	Technical specialist	2023/10/31
L	Autonomous driving	VP11	Team Lead	2023/10/31
		VP12	Technical Expert	2023/10/31
		VP13	Data Scientist	2023/10/31
M	Energy	VP14	Project Manager	2023/10/31
		VP15	Junior Software Architect	2023/10/31
N	Pumps	VP16	Lead data scientist	2023/10/31

Table 4
Description of practitioners involved in validation study.

Company	Description	Practitioners		Date
		ID	*Roles*	
H	Telecommunications	VP1	Expert in Network Architecture evolution (AI/ML)	2023/11/10
		VP2	Main Technical Coordinator	2023/11/10
		VP3	Specialist in ML and Data Science	2023/11/10
I	Vehicle manufacturing	VP4	Platform Architect for ML	2023/11/15
J	Home appliance manufacturing	VP5	Line Manager	2023/11/20
		VP6	Software Engineer	2023/11/20
K	Aviation	VP7	Head of Signal and Data processing Applications	2024/01/11
		VP8	Technical specialist	2024/01/11
		VP9	Engineer	2024/01/11
		VP10	System Architect	2024/01/11
M	Energy	VP14	Project Manager	2024/01/11
		VP15	Junior Software Architect	2024/01/11

5. Findings

Below, we detail the empirical findings from our multi-case study.

5.1. Analysing similarities and differences in adoption of MLOps practices

We have identified similarities and differences when adopting MLOps based on the multi-case study. To enhance the readability of our findings, we have organised them into five dimensions based on an extensive literature review from our prior study [53]. These dimensions cover critical components of the ML lifecycle, i.e., (a) Data, (b) Model, (c) Deployment, (d) Operations & Infrastructure, and (e) Organisation. Table 5 outlines the assigned set of codes for MLOps practices in each dimension. Table 6 highlights the similarities and differences observed in MLOps practices within each company. In Table 6, we use checkmarks (✓) or cross mark (X) to indicate the presence or absence of specific MLOps practices.

Table 6 shows that all companies (A - G) follow standard processes for data management and adhere to data quality checks or validation and mechanisms for handling sensitive data. Each company has different approaches to dealing with sensitive data. For instance, company A ensures data confidentiality through automated methods, company C employs federated ML, synthetic data generation and homographic encryption, and Company D supports GxP (good practice).

Most companies (A, B, D, E and G) version their data, while only a few utilise feature stores (except companies C, D and E). The majority of companies (A, B, E and F) have standard processes for model development, except for companies C, D and G. All companies, except company A, support code versioning, and companies except A and F support model metadata management. Only companies E and F have provisions for reproducible experimentation setup.

Table 5
Assigned set of codes for MLOps practices in each dimension.

Codes	Description
D1	Standard process for data management
D2	Use of data versioning
D3	Data quality checks or data validation
D4	Mechanisms for handling sensitive data
D5	Data governance (including policies and security)
D6	Use of feature store
M1	Standard process for model development
M2	Use of code versioning
M3	Reproducible experimentation set up
M4	Model metadata management
DY1	Automated deployments
DY2	Set up for CI/CD pipelines
DY3	Flexible deployment options
OI1	Automated Monitoring and retraining
OI2	Performance alerting mechanisms
OI3	Adequate infrastructure or tools for hosting and maintenance of models
OG1	Defined roles and collaboration within teams
OG2	Regular communication or meetings with stakeholders

Interestingly, only company B has implemented automated deployments, while the rest still rely on ad hoc or manual processes. Additionally, companies have varying frequencies of deployments, with Companies B, D, F, and G having a CI/CD pipeline set up. Moreover, all companies except A, C, and G have flexible deployment options.

Table 6
Similarities and differences in the adoption of MLOps practices among Companies.

Dimension	Codes	Companies						
		A	B	C	D	E	F	G
Data	D1	✓	✓	✓	✓	✓	✓	✓
	D2	✓	✓	X	✓	✓	X	✓
	D3	✓	✓	✓	✓	✓	✓	✓
	D4	✓	✓	✓	✓	✓	✓	✓
	D5	✓	✓	✓	✓	✓	✓	✓
	D6	✓	✓	X	X	X	✓	X
Model	M1	✓	✓	X	X	✓	✓	X
	M2	X	✓	✓	✓	✓	✓	✓
	M3	X	X	X	X	✓	✓	X
	M4	X	✓	X	✓	✓	✓	✓
Deployment	DY1	X	✓	X	X	X	X	X
	DY2	X	✓	X	✓	X	✓	✓
	DY3	X	✓	X	✓	✓	✓	X
Operations and Infrastructure	OI1	X	✓	X	X	X	X	X
	OI2	X	✓	X	X	✓	✓	X
	OI3	✓	✓	✓	X	✓	✓	X
Organisation	OG1	X	✓	X	✓	✓	✓	X
	OG2	X	✓	✓	✓	✓	✓	✓

For instance, some companies utilise cloud platforms, and others rely on in-house tools.

From Table 6, it is evident that only company B has automated processes for monitoring and retraining. Furthermore, a few companies (B, E and F) have performance alerting mechanisms. Companies other than D and G have the necessary infrastructure or tools for hosting and maintaining models. Additionally, companies A, C and G need more well-defined roles, and company A needs better communication with stakeholders and the team.

Based on observations from the multi-case study, the underlying MLOps processes and workflow remain the same even though the tools vary from company to company. This has led to the development of a standardised (a) MLOps framework, and (b) MLOps maturity model. The MLOps practices outlined in the MLOps framework cover various practices observed and extracted from different studied companies. The MLOps framework helps companies standardise their MLOps practices, whereas the MLOps maturity model allows companies to assess and advance their maturity. The MLOps framework is presented in Table 7, and the MLOps maturity model in Fig. 2. The MLOps maturity model consists of five stages: (a) Ad hoc, (b) DataOps, (c) Manual MLOps, (d) Automated MLOps, and (e) Kaizen MLOps. These stages span five dimensions, i.e., Data, Model, Deployment, Operations & Infrastructure, and Organisation. In the Ad hoc stage, companies have ad hoc processes for data, model development, deployment and operations. They also have limited infrastructure and tools, undefined roles, and irregular communication and collaboration with stakeholders and teams. In the DataOps stage, companies shift towards standardised and automated processes for data, while the processes related to the model (i.e., development, deployment and operations) remain ad hoc. The Manual MLOps stage represents the transition of companies towards standardisation in model development, deployment and operations processes. Furthermore, they have regular communication and collaboration with stakeholders and teams. As companies progress to the Automated MLOps stage, they begin to introduce automation into all processes related to models and data. Finally, in the Kaizen MLOps stage, companies aim for a continuous improvement mindset across all dimensions. During this stage, all processes related to data, models, deployment, operations, infrastructure, and roles within companies are refined and optimised. Fig. 2 illustrates the mapping of studied companies to their respective MLOps stages based on their maturity.

5.2. Empirical validation of MLOps framework and MLOps maturity model

We conducted an empirical validation study in five companies in the Software Center to verify the effectiveness of the MLOps framework and maturity model. Through this empirical validation, we aim to gather valuable feedback from companies to identify strengths, weaknesses, and areas for improvement within the MLOps framework and maturity model. Based on the feedback we received from the validation study, we revised and enhanced the MLOps framework and maturity model. Below, we detail the received validation feedback.

5.2.1. Examining feedback received from validation study

To structure and present the feedback from the validation study, we have categorised them into three: (a) Confirmatory (CF), (b) Constructive (ConF), and (c) Optimisation (OF). Confirmatory feedback indicates that specific elements of the MLOps framework and maturity model either meet or exceed the expectations of practitioners involved in the validation study. Constructive feedback points out that specific elements of the MLOps framework and maturity model fall short of the expectations of practitioners or need improvement. Optimisation feedback suggests that existing specific elements of the MLOps framework and maturity model need fine-tuning or optimisation.

In the transcripts of validation interviews, we search for (a) Responses that express satisfaction to identify confirmatory feedback, (b) Responses offering critiques, suggestions, or recommendations, and (c) Responses suggesting adjustments, improvements, or clarifications to existing specific elements of the MLOps framework and maturity model to identify optimisation feedback. The first author assigns each statement from the transcript to one of these feedback categories. Then, these are presented to the second and third authors for review.

Below, we detail the categorisation of validation feedback and include only a few examples of feedback we receive from practitioners involved in the validation study.

1. Confirmatory feedback: The validation feedback on the MLOps framework and maturity model is highly positive and is well-received for its applicability in real-world settings. The endorsement from different validation practitioners (i.e., practitioners involved in the validation study) adds credibility to the presented MLOps framework and maturity model. The acknowledgement from company H indicates that the MLOps framework and maturity model are considered comprehensive and beneficial for companies adopting MLOps practices. They explicitly mentioned this by saying *"Thank you for the initiative to propose a maturity model that can be used by several; I think it can help companies set a direction for their work with MLOps and also know what to strive for"*. They express that the MLOps framework and maturity model enhance automation across different product areas within their company. VP6 at Company J highlight that the presented MLOps framework and maturity model are more valuable to companies, especially top management,

Table 7
MLOps Framework (Five dimensions used to define stages in the maturity model) [18].

Dimensions	Stages				
	Ad hoc	DataOps	Manual MLOps	Automated MLOps	Kaizen MLOps
Data	Ad hoc processes for data management	Standardised and automated processes for data management	Standardised and automated processes for data management	Standardised and automated processes for data management	Data management is continuously improved through an iterative process
	Ad hoc processes for data governance	Standardised and automated processes for data governance	Standardised and automated processes for data governance	Standardised and automated processes for data governance	Data governance is continuously improved through an iterative process
	Ad hoc data versioning	Standardised and automated data versioning	Standardised and automated data versioning	Standardised and automated data versioning	Continuously improved and optimised data versioning
	Ad hoc feature store	Standardised and automated feature store	Standardised and automated feature store	Standardised and automated feature store	Continuously improved and optimised feature store
Model	Ad hoc processes for model development	Ad hoc processes for model development	Standardised and manual processes for model development	Standardised and automated processes for model development	Model development is continuously improved through an iterative processes
	Ad hoc code versioning and code review	Basic code versioning and code review	Highly structured and manual code versioning and code review	Highly structured and automated code versioning and code review	Continuously improved and optimised code versioning and review
	Ad hoc model metadata management	Ad hoc model metadata management	Standardised and manual model metadata management	Standardised and automated model metadata management	Continuously improved and optimised model metadata management
	No reproducible experimentation setup	Limited reproducible experimentation setup	Manual reproducible experimentation setup	Automated reproducible experimentation setup	Reproducible experimentation setup is continuously refined and improved
Deployment	Adhoc processes for model deployment	Ad hoc processes for model deployment	Standardised and manual processes for model deployment	Standardised and automated processes for model deployment	Model deployment is continuously improved through an iterative processes
	Infrequent model deployments	Less frequent model deployments	Frequent model deployments	Highly frequent model deployments	Continuous and iterative model deployments
	No/basic CI/CD pipeline	Well-defined CI/CD pipeline	Manual CI/CD pipeline	Highly automated CI/CD pipeline	CI/CD pipeline is continuously improved and optimised
Operations and Infrastructure	Ad hoc processes for model monitoring	Ad hoc processes for model monitoring	Standardised and manual processes for model monitoring	Standardised and automated processes for model monitoring	Model monitoring is continuously improved through an iterative processes
	Ad hoc processes for model retraining	Ad hoc processes for model retraining	Standardised and manual processes for model retraining	Standardised and automated processes for model retraining	Model retraining is continuously improved through an iterative processes
	Infrequent model retraining	Less frequent model retraining	Frequent model retraining	Highly frequent model retraining	Continuous and iterative model retraining
	Limited infrastructure and tooling	Infrastructure and tooling focus on data management and governance	Infrastructure and tooling focus on manual ways to manage ML lifecycle	Infrastructure and tooling focus on automated ways to manage ML lifecycle	Infrastructure and tooling support continuous improvement
Organisation	No well-defined roles	Clearly defined roles	Specialised roles	Roles that focus on automation	Roles with focus on continuous improvement and optimisation
	Informal communication with stakeholders when required	Regular communication with stakeholders to inform data-related activities	Regular communication with stakeholders to inform model performance updates	Regular communication with stakeholders to inform model performance updates with automated reports/dashboards	Regular communication with stakeholders to focus on continuous improvement
	Informal communication with team when required	Regular communication with team to focus on data management and governance	Regular communication with team to focus on manual deployment and management of models	Regular communication with team to focus on automated deployment and management of models	Regular communication with team to focus on continuous improvement

compared to other existing maturity models. They affirm the statement when saying *"Because we had a look into the MLOps level model for Azure just prior to this meeting and top management will not really get much out of that because they don't know the specifics of tools that they're what they suggest to be as an improvement to actually get to the next step for example...I agree the top management have more use for this framework than the tool specific ones"*. In addition, VP6 at Company J emphasises that our MLOps framework and maturity model can be used to visualise the current status of an ML project to external stakeholders and identify improvement areas. They express this by saying *"This framework that you just showed us would be awesome to have as a indication of where you're lacking...I think this one really gives a good insight on different domains...So we need to focus on this one and then, after that, having a look at, if we are lacking in this dimension, what are the next steps to go... This*

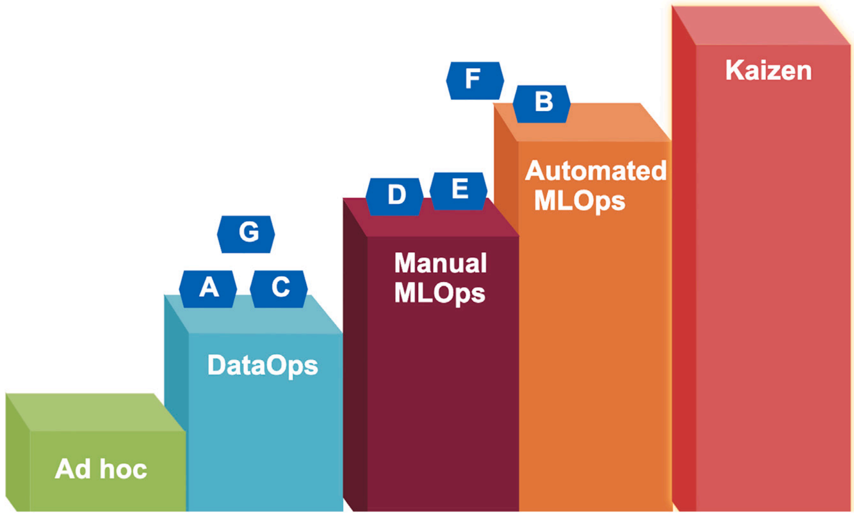

Fig. 2. MLOps maturity model [18].

framework would be really nice to give an overview for everyone external that doesn't necessarily know much about the topic but also give a view on the project and our status of where we at so we need to focus on this and then take it to the developers like say we're lacking this what do we need to do". Also, VP5 and VP6 from Company J highlight the versatility of the MLOps framework and maturity model. For instance, its application in (a) Internal projects, (b) Team discussions for improvement and (c) Providing insights for leadership and organisational management.

Company H appreciates the inclusion of the "organisation" dimension in the MLOps framework and maturity model. The emphasis on reproducibility solidifies the position of our MLOps framework and maturity model as a valuable tool in the MLOps landscape. For instance, *VP4* from Company I highlights the critical importance of reproducibility and indicates this by saying, *"We spend a lot of time to make sure that reproducibility is in place. Even with a rudimentary MLOps pipeline, we would still have a very good reproducibility mechanism; otherwise our development process becomes extremely difficult. We really prioritise that even if we don't have a highly automated CI/CD pipeline in the case of automated MLOps, we would absolutely have reproducibility because that's important for us".*

2. Constructive Feedback: Validation practitioners at Company H express concerns about finding the balance between the level of detail and simplicity in claiming adherence to a particular stage in the MLOps framework and maturity model. They indicate that a company or a product area within a company may fall into different stages when considering all elements of the MLOps framework and maturity model. This is evident when they say *"I think this may vary so that in some area, a company/product area is in "Manual" while in another in "Automated" (as an example). It is probably a balance, how much details to include to describe a Stage in such model while making it fairly simple to claim adherence to one stage (completely). This of course depends on the intent with the maturity model".* Validation practitioners at Company H find issues with including "Data" as a dimension in the MLOps framework and maturity model and suggest their preference for treating "Data" as a prerequisite rather than a maturity stage. They indicate this by saying *"Data is of course a foundation for any AI/ML operation, but we'd rather see it as a prerequisite when it comes to maturity in MLOps, at least when it comes to data management and data governance. As it is included now, putting is as a Maturity stage ("DataOps") where all components are "standardised and automated" including a Feature store even before you come to manual MLOps is one indication to this".* They recommend separating the "Data/DataOps", especially data management and governance elements, from the MLOps framework and maturity model. In addition, they suggest including the "feature store" as part of the "Model" dimension. In their opinion, the MLOps framework and maturity model lack quality and assurance in the "Model" dimension. Validation practitioners at Company H find that infrastructure and tooling seem non-congruent with the stages in the MLOps framework. They believe that utilising more tools helps in achieving greater maturity. This is obvious when saying *"We would see that you can include more tools to achieve more maturity. As example, for Adhoc and Manual MLOps I could see as simply using Notebooks/IDE throughout data pre-processing, feature engineering, training etc, while when you want to automate the process more one can start to orchestrate the test runs for faster iterations, and also in a more automated way store the necessary data in experimentation tracking systems, model stores, store metadata, etc.".* The MLOps framework is criticised for providing more importance to model development than creating pipelines for continuous re-training, deployment, inference, and monitoring. In contrast to their observation, they also acknowledged that these are partially included in the "Operations and Infrastructure" dimension of our MLOps framework and maturity model. Instead of focusing more on model development, they suggested focusing on the development of SW (software) components/workloads (together building the ML pipeline) for automated re-training in a production environment and integration into an (automated) SW delivery pipeline. They pinpoint the need to be specific about roles needed during each stage and raise concerns that roles may need to match when they start to focus on automation. In Company I, *VP4* points out a disconnect with the sequential approach of maturing DataOps before MLOps. This may arise due to the diverse data sources in Company I, ranging from flat databases to graphs. As a result, there may be a need to adopt DataOps practices from scratch when introducing a new data source. So, they recommend that the maturation of DataOps be parallel with that of MLOps. They highlight this by saying *"We don't see that working in practice. I can give you an example here. So even if you spend all the time maturing DataOps, there could be a new data source that is so different that you have to essentially do a new cycle or a new round of DataOps. The maturity is very dependent upon sources of data. I think like no matter how mature we are, we will always be challenged by a new source.....So what we do is so that the DataOPS maturing process needs to be in parallel with the MLOps maturing process".* *VP4* at Company I points out that constant development after deploying the model (post-deployment) is significantly ignored. This is particularly true for cases where models are used for task-solving purposes other than predictive purposes. VP4 mentions *"There is post training and post deployment development that happens which most MLOPS pipelines in literature today do not really talk about".* VP6 at Company J indicates the difficulty differentiating between well-defined and manual CI/CD pipelines. According to the opinion of VP6, the MLOps framework is tool-agnostic, so they advocate for placing strict requirements. VP6 pinpointed the following example and said, *"It doesn't necessarily provide like specific requirements. Just as an example, the DataOps one for the well defined CI/CD pipelines. Does 100% of the CI/CD pipeline need to be automated to go to the next step or like it's now like real hard requirements?"*. Both VP5 and VP6 at Company J thinks that due to the absence of specific requirements, the MLOps framework will be less suitable for external evaluations. VP5 and VP6 also suggested that it is good to add descriptions to each role. This is evident when saying, *"This is just an idea, because if, for example, you in the organisation you have like the role well defined roles, specialised roles. And it could be nice to have examples. Also, a lot of work I can see that, but just the idea"*. According to VP4, it is not achievable to set up a comprehensive data platform and teams to manage it due to the nature of the company as an AI company. VP4 also stressed that there will always be data-related issues that require improvements.

3. Optimisation Feedback: Validation practitioners from Company H propose adding data quality, integrity, and privacy considerations to the "data management" element in the MLOps framework and maturity model. They confirmed that the "Kaizen" stage has raised some internal discussions in their company. They prefer to see a measurable target in the Kaizen stage. This is expressed by saying *"While acknowledging the ambition for continuous improvements, we would in a maturity model like this prefer to see a more concrete and measurable target. E.g. this could be how well the MLOps process integrates with the DevOps process for the overall integration, delivery and deployment of the product SW where the ML-based functionality & ML pipeline is included"*. The validation practitioners at Company H emphasise that continuous re-training depends on the use case. As some use cases do not benefit from continuous re-training, it is recommended to express this as "continuous monitoring including triggering of re-training (when necessary)". VP4 feels a clarity issue in defining the kaizen stage and interprets it as programmable MLOps. In Company I, *VP4* prefer tools with a significant community following and that are open-source. VP4 stressed this by saying, *"It's good if the tool is open sourced. This means improvements are more likely to be added to that tool even if it lacks some functionality at the moment somebody's going to add it to that"*.

Categorising the feedback helps us identify areas needing improvement and take quick action. Furthermore, we tried to revise the MLOps framework and maturity model based on validation feedback. For this, we need to decide whether to (a) Keep, (b) Modify, or (c) Remove each feedback. We choose to keep predominantly confirmatory responses that suggest no need for change. We decide to modify constructive or optimisation responses that need potential optimisations. We remove constructive or add optimisation responses whose removal or addition increases efficiency and simplifies our MLOps framework and maturity model. Below, we summarise the validation feedback and shown in Table 8.

In the "Data" dimension, we reject ConF1, i.e., the suggestion to treat data as a prerequisite rather than a dimension. We neglect this feedback as data-related issues must be addressed when the dataset is subject to change over time. The validation practitioners also confirmed that introducing new data sources challenges them while adopting MLOps practices. We reject ConF2, i.e., the recommendation to include a feature store in the "Model" dimension. The rejection is because the feature store stores features extracted from a dataset, which can be used as input for the ML model. Therefore, we keep the feature store in the "Data" dimension. We accept OF1. In our presented MLOps framework and maturity model, data management includes all data collection, aggregation, and processing activities, and data governance covers regulations, procedures, and policies. To incorporate OF1 feedback, we add data quality and integrity to data management and data privacy to data governance. We reject ConF3, i.e., maturing DataOps in parallel with MLOps. When considering companies with limited practitioners, knowledge, budget or time, sequential maturation of DataOps before MLOps is beneficial. In exceptional cases, if new data sources challenge the company or have a high availability of experienced people in all stages of ML workflow, then maturing DataOps in parallel with MLOps can improve efficiency.

In the "Model" dimension, we accept ConF5, i.e., the lack of emphasis on model quality and assurance. This feedback is accepted and incorporated into the MLOps framework and maturity model as it ensures the performance, trust, and accountability of models. We confirm CF1, i.e., the need to ensure reproducibility when adopting MLOps.

In the "Deployment" dimension, we accept ConF6. There needs to be more differentiation between a well-defined CI/CD pipeline and a manual CI/CD pipeline. This helps to optimise the presented MLOps framework and maturity model.

In the "operations & Infrastructure dimension", we accept ConF7, i.e., there needs to be a focus on the development of SW components/workloads for automated re-training and integration in an (automated) SW delivery pipeline. In addition, we accept ConF8, i.e., Infrastructure and tooling seem non-congruent with the stages in the MLOps framework. We also accept OF2 and ConF9, i.e., the need to clarify continuous re-training as continuous monitoring, including triggering of re-training (when necessary) and significant ignorance of constant development after deploying models.

In the "Organisation" dimension, we confirm CF2, i.e., the inclusion of the organisation dimension in the MLops framework and maturity model. Validation practitioners mentioned that it is a new addition compared to existing ones in popular MLOps frameworks and maturity models. We accept ConF10, i.e., mismatch in roles when focusing on automation. When companies steer towards automation, the roles need to be aligned with the focus on automation. We accept ConF11, i.e., providing training to people involved in the project. Since different companies utilise different tools based on their use case and external constraints, we reject ConF12, i.e., the absence of a specific description of tools.

We accept ConF13, i.e., the balance between the level of detail and simplicity in claiming adherence to a particular stage. It ensures a more precise assessment of MLOps maturity and allows detailed analysis of improvement areas. In addition, accepting the validation feedback and incorporating it into the MLOps framework and maturity model increases the level of detail.

To conclude, we confirm CF1, CF2, CF3, CF4, CF5, CF6 and CF7. These feedbacks highlight the value and importance of the presented MLOps framework and maturity model. On the other hand, we reject ConF1, ConF2, ConF3, ConF4, ConF12 and accept ConF5, ConF6, ConF7, ConF8, ConF9, ConF10 and ConF11. Also, we accept OF1 and OF2. We provide revised MLOps framework and maturity model in Table 9.

5.3. MLOps taxonomy

MLOps framework and maturity model provide a comprehensive overview of best practices available for companies that are adopting MLOps. But, in practice, companies manage and offer diverse products to establish themselves as industry leaders and meet the evolving needs of their customers. Consequently, each use case within the companies has its own set of unique requirements and exceptions to be addressed [54]. Therefore, companies gain significant advantages by adopting tailored MLOps practices that are suited to their use cases [54]. To achieve this purpose, we have developed an MLOps taxonomy and mapped various use cases we studied in companies to different desired stages of the MLOps framework. These stages include (a) Ad hoc, (b) DataOps, (c) Manual MLOps, (d) Automated MLOps, and (e) Kaizen MLOps. A taxonomy provides a consistent and transparent perspective [55] for selecting the desired MLOps practices for each use

Table 8
Empirical validation feedback on MLOps framework and maturity model.

Framework dimension	Empirical findings	Example feedback
Data	Inclusion of “Data” into the MLOps framework.	ConF1: Issues with including“Data” as a dimension and suggest treating it as a prerequisite.
	Separation of Data/DataOps (especially data management and data governance) from the MLOps framework	ConF2: Recommends separation and suggest inclusion of “feature store” in the “Model” dimension
	Completeness of “data management” aspect	OF1: Recommend on incorporating data quality, integrity and privacy (if not already included in the “data management”in the MLOps framework and maturity model)
	Sequential maturation of DataOps before MLOps	ConF3: Not possible in practice, and maturation of DataOps should occur in parallel with MLOps
Model	Focus on model development	ConF4: More focused on model development over creating pipelines for continuous operations.
	Emphasis on model quality and assurance	ConF5: Lack of emphasis on model quality and assurance
	Need for reproducibility	CF1: Prioritise reproducibility
Deployment	Clarify on well defined CI/CD pipeline and manual CI/CD pipeline	ConF6: No clarity
Operations and Infrastructure	Developing SW components/workloads (together building the ML pipeline) for automated re-training	ConF7: Advise focus on the development of SW components/ workloads for automated re-training and integration in an (automated) SW delivery pipeline
	Congruence of infrastructure and tooling across MLOps framework stages.	ConF8: Infrastructure and tooling seem non-congruent with the stages in the MLOps framework
	Clarification of “Continuous re-training” terminology.	OF2: Recommend to express it as“ continuous monitoring including triggering of re-training (when necessary)”
	Constant development happens post model deployment	ConF9: Significant ignorance of constant development after deploying models
Organisation	Inclusion of “organisation” in MLOps framework	CF2: Appreciate the inclusion
	Definition of roles in different stages of MLOps framework	ConF10: Mismatch in roles when focusing on automation
	Providing training to people	ConF11: Not covered in the MLOps framework
	Tool agnostic MLOps framework	ConF12: Absence of specific requirements (descriptions)
Overall	Comprehensiveness of the MLOps framework.	CF3: Helpful in adopting MLOps practices.
	Flexibility in determining the MLOps maturity stage	ConF13: Balance between the level of detail and simplicity in claiming adherence to a particular stage
	Clarity and specificity in “Kaizen” stage	OF3: Need for a measurable target OF4: Interpret Kaizen as a programmable MLOps
	Utility of the MLOps framework	CF4: Tool for visualising the current status of an ML project to external stakeholders and identifying improvement areas CF5: More valuable to companies compared to tool-specific maturity models. CF6: MLOps framework can be used for internal projects, team improvement discussions and for providing insights for leadership and organisational management. CF7: MLOps maturity model can also bring values the same way as Software maturity model

case. We have chosen two criteria for standardising the MLOps taxonomy across various companies - (a) Regulatory Compliance and (b) Operational Excellence. Hence, companies can determine whether their specific use case aligns or misaligns with the desired MLOps practices in the defined MLOps taxonomy. If the use case aligns, it indicates that the company has adopted MLOps practices that are consistent with the taxonomy. Conversely, if the use case is misaligned, it helps companies identify the challenges they face due to non-adherence to the desired MLOps practices outlined in the taxonomy. This enables companies to address these challenges and improve their MLOps practices. Below, we detail each criterion in the context of the revised MLOps framework.

5.3.1. Regulatory compliance

Regulatory compliance refers to legal/regulatory requirements [55] applicable to a particular use case. It influences the practices related to data collection, development, deployment and operationalisation of

Table 9
Revised MLOps framework.

Dimensions	Stages				
	Ad hoc	DataOps	Manual MLOps	Automated MLOps	Kaizen MLOps
Data	Ad hoc processes for data management	Standardised and automated processes for data management	Standardised and automated processes for data management	Standardised and automated processes for data management	Data management is continuously improved through an iterative process
	Ad hoc processes for data governance	Standardised and automated processes for data governance	Standardised and automated processes for data governance	Standardised and automated processes for data governance	Data governance is continuously improved through an iterative process
	Ad hoc data versioning	Standardised and automated data versioning	Standardised and automated data versioning	Standardised and automated data versioning	Continuously improved and optimised data versioning
	Ad hoc feature store	Standardised and automated feature store	Standardised and automated feature store	Standardised and automated feature store	Continuously improved and optimised feature store
Model	Ad hoc processes for model development with a lack of emphasis on model quality and assurance	Ad hoc processes for model development with a lack of emphasis on model quality and assurance	Standardised and manual processes for model development with an emphasis on manual model quality and assurance	Standardised and automated processes for model development with an emphasis on automated model quality and assurance	Continuous improvement of model quality and assurance as part of the iterative process for model development
	Ad hoc code versioning and code review	Basic code versioning and code review	Highly structured and manual code versioning and code review	Highly structured and automated code versioning and code review	Continuously improved and optimised code versioning and review
	Ad hoc model metadata management	Ad hoc model metadata management	Standardised and manual model metadata management	Standardised and automated model metadata management	Continuously improved and optimised model metadata management
	No reproducible experimentation setup	Limited reproducible experimentation setup	Manual reproducible experimentation setup	Automated reproducible experimentation setup	Reproducible experimentation setup is continuously refined and improved
Deployment	Adhoc processes for model deployment	Ad hoc processes for model deployment	Standardised and manual processes for model deployment	Standardised and automated processes for model deployment	Model deployment is continuously improved through an iterative processes
	Infrequent model deployments	Less frequent model deployments	Frequent model deployments	Highly frequent model deployments	Continuous and iterative model deployments
	No/basic CI/CD pipeline	Initial CI/CD pipeline	Manual CI/CD pipeline	Highly automated CI/CD pipeline	CI/CD pipeline is continuously improved and optimised
Operations and Infrastructure	Ad hoc processes for model monitoring including triggering of ad hoc re-training (when necessary)	Ad hoc processes for model monitoring including triggering of ad hoc re-training (when necessary)	Standardised and manual processes for model monitoring including triggering of manual re-training (when necessary)	Standardised and automated processes for model monitoring including triggering of automated re-training (when necessary)	Model monitoring and triggering of model re-training (when necessary) is continuously improved through an iterative processes
	Infrequent model retraining	Less frequent model retraining	Frequent model retraining	Highly frequent model retraining	Continuous and iterative model retraining
	Limited infrastructure and tooling	Infrastructure and tooling focus on data management and governance	Infrastructure and tooling focus on manual ways to manage ML lifecycle	Infrastructure and tooling focus on automated ways to manage ML lifecycle	Infrastructure and tooling support continuous improvement
Organisation	No well-defined roles	Clearly defined roles	Specialised roles	Roles that specialise in automation of model development, deployment, and operations.	Roles with a focus on continuous improvement, optimisation, and automation
	Informal communication with stakeholders when required	Regular communication with stakeholders to inform data-related activities	Regular communication with stakeholders to inform model performance updates	Regular communication with stakeholders to inform model performance updates with automated reports/dashboards	Regular communication with stakeholders to focus on continuous improvement
	Informal communication with team when required	Regular communication with team to focus on data management and governance	Regular communication with team to focus on manual deployment and management of models	Regular communication with team to focus on automated deployment and management of models	Regular communication with team to focus on continuous improvement

ML/DL models for that use case. Within the context of the revised MLOps framework, adhering to regulatory compliance [56] involves (a) Implementing secure data management (to address collection, aggregation, processing and sharing of data) and data governance (to establish procedures and policies for data management) practices, (b) Addressing issues related to bias, fairness, transparency, and reproducibility when developing models, (c) Providing understandable explanations for outcomes of the ML model, (d) Ensuring continuous monitoring

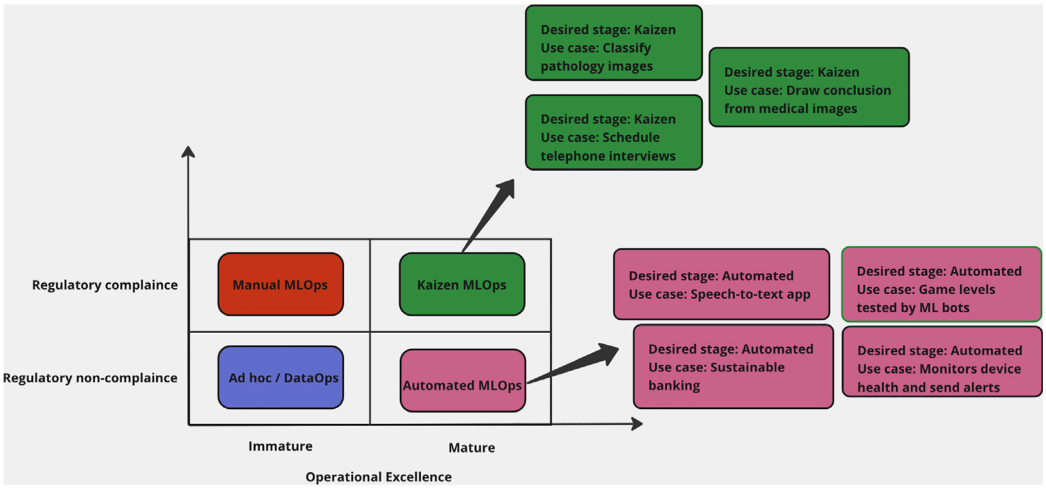

Fig. 3. MLOps taxonomy.

and retraining of model, and (e) Engaging with stakeholders, teams and legal/regulatory experts to ensure regulatory compliance across the ML workflow.

5.3.2. Operational excellence

Operational excellence refers to the efficient and automated deployment and operationalisation of ML models in companies. Within the revised MLOps framework, operational excellence involves (a) Mechanisms for versioning data, features, and models, (b) Streamlining the development, deployment and retraining of models, (c) Setting up CI/CD pipelines (d) Establishing alert mechanisms for drifts, and (e) Ensuring the availability of tools and infrastructure.

We have created a two-dimensional graph to visually represent the MLOps taxonomy as shown in Fig. 3. The graph has been designed to represent the level of operational excellence on the X-axis, with “Immature” on the left and “Mature” on the right. Similarly, the Y-axis represents the level of regulatory compliance ranging from “Regulatory non-compliance” at the bottom and “Regulatory compliance” at the top. This graphical representation allows us to assess each use case studied in companies against these two criteria and plot its position in the graph. In addition, we have employed a quadrant matrix chart to represent how different combinations of these two criteria result in the desired MLOps stage.

As shown in the above MLOps taxonomy, we categorised the use cases from Company A to G using the quadrant matrix chart. Our analysis highlighted that the desired and actual MLOps stage of two of the use cases, “sustainable banking” and “game levels tested by ML bots”, almost align with MLOps Taxonomy. However, the desired stage of the use case, “schedule telephone calls”, is Kaizen, but in practice, it falls under the DataOps stage. This is because of the need for more specific mention of adherence to regulatory compliance and the fact that the model deployments are ad hoc and infrequent without a CI/CD pipeline. Similarly, the desired stage for the use case “classify pathology images” is “Kaizen MLOps”, whereas, in practice, it is in the DataOps stage due to the need for help with data synchronisation to AI platforms and systematic data versioning. Moreover, the lack of standard processes for model development, an experimental deployment stage, a practical CI/CD pipeline and mechanisms for monitoring and retraining contributed to the difference between the desired and actual stages. The desired stage for the use case “concludes medical images” is Kaizen. Still, in practice, it is in the “Manual MLOps stage” because they lack experimentation reproducibility and standardised processes for model deployment with less degree of automation in monitoring and retraining.

The desired stage for the use case “speech to text app” is “Automated MLOps”, whereas, in practice, the use case is in the “Manual MLOps” stage. This use case lacks feature store and data versioning in model release, automated model deployment, monitoring and retraining processes, and the ability to set up a CI/CD pipeline. Finally, the desired stage of the use case “Monitors device health and send alerts” is “Automated MLOps”. However, in practice, it is under the “DataOps” stage because they lack knowledge in operationalising models. As a result, they need a use case in the production stage.

6. Discussion

The multi-case study conducted in collaboration with AI Sweden and Software Center (i.e., embedded systems and IT-intensive companies) highlights several important insights into the field of MLOps. By developing the MLOps framework, maturity model and taxonomy through partnership with practitioners from companies, we address the need for co-innovating and co-creating solutions for MLOps adoption [57]. Our study reduces the shortage of scientific articles focused on designing MLOps solutions [12,16] and helps to reduce existing doubts or distrust regarding MLOps adoption [13].

To address RQ1, we conducted semi-structured interviews and found that companies under study share several similarities and differences when adopting MLOps practices. Based on our analysis, we observe strong adherence to data practices across companies such as data management (D1, D3, D4 and D5). However, there exists a difference in the adoption of advanced MLOps practices such as feature stores (D6) and automated deployments (DY1). Interestingly, most companies show robust approaches to code and data versioning (M2 and D2), but there are notable gaps in reproducible experimentation (M3) and flexible deployment options (DY3). Companies that excel in deployment and operations (For instance, Company B and Company F) also tend to have strong infrastructure for monitoring and retraining (OI2, OI3). This suggests a correlation between advanced operational and infrastructure capabilities, as well as deployment flexibility. Organisation dimensions, for instance, defined roles and regular communication (OG1, OG2) are crucial for all companies exhibiting maturity in MLOps adoption.

Our findings challenge the hypothesis suggested by [33]. It proposes that companies typically proceed from a data-centric setup towards a model-centric setup before starting to operationalise models. However, we observe that companies tend to adopt practices based on their available tools and skills. This demands flexibility in the MLOps frameworks. In a previous paper, we developed a maturity model with four stages [18], derived from academic literature and validated in companies. Since MLOps is an evolving field and companies are constantly adopting MLOps practices, new or revisited maturity models are needed.

The maturity model we proposed addresses the limitations of existing maturity models (mentioned in the previous section (3.2)) bound to specific platforms or domains. This is beneficial for companies that have multi-cloud settings or want to avoid vendor lock-in. Our framework and maturity models align with most of the fundamental MLOps principles outlined in existing maturity models. However, our framework includes all elements required for successful MLOps adoption, breaks them into five dimensions and progresses through different maturity stages. The proposed framework begins with a fragmented stage (Ad hoc), advances to standardised processes (DataOps and MLOps), gradually progresses to a fully automated stage (Automated MLOps), and finally focuses on continuous improvement through an iterative process (Kaizen stage). In this way, our framework and maturity model assist companies in the early stages of MLOps adoption, where processes and teams have not yet been established. Therefore, the proposed framework and maturity model differs from existing models that assume a certain level of standard processes already in place in companies. Furthermore, our framework outperforms others through the "Kaizen" stage, which promotes continuous improvement or evolution of various processes related to data, models, deployment, operations, and organisation. Unlike other existing models that mention basic collaboration among the teams, our framework focuses on the ongoing optimisation of pipelines, collaboration and communication between teams, and the refinement of roles.

We observe that some companies involved in the study assess their MLOps maturity using existing maturity models, whereas other companies rely on in-house maturity models. Additionally, it is noted that only a few company domains have attempted to assess the maturity of their MLOps projects. For example, automotive and manufacturing domains [58,59]. Our study can be an inspiration for more companies to assess and advance their MLOps practices. When companies in AI Sweden are mapped to the MLOps maturity model (based on the MLOps framework), we see that most companies struggle in the initial stages of the MLOps maturity model. As an exception, only a few companies have reached the automated MLOps stage. Also, we have come across use cases that are already in production, have yet to be deployed, or lack a specific use case.

The related work (Section 3.1), which details the methods, process, tools, and adoption challenges, is not comprehensive; instead, it provides an overview of existing knowledge. The MLOps practices described in our MLOps framework resonate with the methods highlighted in the related work. This indicates that our framework is built on established methods. Our framework provides companies with the independence to use any tools mentioned in related work based on their needs. The MLOps workflow depicts steps involved in the development, deployment and evolution of models, whereas the MLOps practices try to automate these ML workflow. In our MLOps framework, we emphasise the importance of including organisational perspectives (one of the challenges mentioned in Section 3.1).

Validating the developed MLOps framework and maturity model with a set of different companies enhances the credibility of the work (RQ3). The feedback is accepted or rejected after forming a consensus among all authors. The validation process ensures that the framework is comprehensive and covers all the necessary MLOps practices for successful MLOps adoption.

While developing and validating the MLOps framework and maturity model, we noticed that the domain of a few companies (for instance, in health care) hinders them from reaching the Kaizen stage mentioned in the MLOps maturity model. This motivated us to think about an MLOps Taxonomy (RQ2) for classifying ML use cases in various company contexts. First, we planned to add three criteria: (a) Regulatory compliance, (b) Operational Excellence and (c) Organisational maturity. Then, we finalised that two criteria would be good as we see an overlap of organisation in the first two criteria. In addition, when trying to develop an MLOps taxonomy, we initially considered classifying the use cases based on companies or domains. However, we rejected this idea since companies deal with varying use cases. For example, Company B handles use cases like sustainability banking and financial crime prevention. Similarly, Company C focuses on the classification of pathology images and the prediction of sickness leave of hospital staff. This highlights that most companies deal with a wide variety of use cases, including sensitive and non-sensitive use cases. Based on our MLOps taxonomy, most companies are not in their desired adoption stage of MLOps practices. However, the MLOps taxonomy provides an exact idea about what should be the possible MLOps stage to achieve based on their constraints and thus spend time and resources wisely.

Based on the study, we have noticed that each company has a unique approach to adopting MLOps practices. This presents an opportunity, as mentioned in [57], to standardise and benchmark the adoption of MLOps practices across companies through the development of an MLOps framework, maturity model, and taxonomy.

7. Validity threats

We addressed potential validity threats in our study [60]. To improve construct validity, we conduced additional interviews with practitioners from the same company whenever possible. Moreover, we sent the study objective to interviewees in advance and validated the interview guide by a person at AI Sweden during Phase 1 to ensure accurate observation. To ensure internal validity, we gathered feedback on our developed MLOps framework and maturity model to ensure correctness. Furthermore, we avoided incorrect groupings of codes and wrong results by collecting evidence from practitioners in interviews and workshops. To consider the external validity of our research, we carefully framed our research questions. When selecting companies for the study, we tried to include companies in their beginning, intermediate, and advanced stages of MLOps adoption and of varying sizes and different domains. This allowed us to develop an MLOps framework, maturity model, and taxonomy that could be generalised beyond the scope of our study. Additionally, the involvement of IT-intensive and embedded systems companies supports the generalisation of our findings. To ensure reliability, the observations from interviews were reviewed by the second and third authors to minimise researcher bias. We also used structured questions during the interviews and created transcripts for analysis.

8. Conclusion

To address the increasing need among companies for structured approaches to advancing the adoption of MLOps practices, we have developed an MLOps framework, maturity model, and taxonomy. The findings of the study were derived from a multi-case study conducted across several companies. These approaches help companies to assess their adopted MLOps practices, benchmark against industry best practices, and improve to reach the desired stages. Moreover, the insights from the study provide valuable learnings for companies on how to successfully adopt MLOps or avoid common pitfalls. As a result, it can be utilised as a guide and source of inspiration for companies regardless of their current MLOps status. In the future, we plan to explore various strategies for transferring MLOps knowledge across teams within companies. In addition, we emphasise the significance of quantitative metrics in providing measurable insights into the effectiveness and efficiency of MLOps systems, particularly in terms such as automation and deployment success rate [61]. For instance, metrics that represent decisions related to integration and delivery as well as pipeline triggers [62]. Recent research highlights the importance of Continuous Integration and Continuous Deployment (CI/CD) pipelines in automating the training and deployment of ML models [63]. In this context, combining the evaluation of quantitative metrics with a qualitative framework for MLOps adoption provides a comprehensive perspective into how effectively a company adopts and advances MLOps practices. Furthermore, we will explore the potential of a Kubernetes-based MLOps framework [64] in achieving the MLOps practices outlined in the Kaizen stage of the maturity model.

CRediT authorship contribution statement

Meenu Mary John: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Helena Holmström Olsson:** Writing – review & editing, Validation, Supervision, Project administration, Conceptualization. **Jan Bosch:** Writing – review & editing, Validation, Supervision, Project administration, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used [Grammarly] in order to [correct grammar and rephrasing sentences]. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank all of the practitioners involved in interviews and workshops from different companies in AI Sweden and Software Center.

Appendix A

A.1. Interview guide - Explore the adoption of MLOps practices across companies

1. *Objective*: Gain a deeper understanding of the techniques and practices of applying SE and operations to the development, deployment and operationalisation of ML models. We are eager to hear about your personal experiences and perspectives on the current use of ML/DL in your company and how it is being integrated into your organisation.
2. *Introduction Questions*
    - Can you tell us about your current role and responsibilities within the company?
    - Please elaborate on typical roles/positions found in your AI team.
    - Can you give us an overview of the projects and initiatives you are currently working on?
    - Do you use ML/DL technologies? What ML/DL domains do your task(s) belong to?
    - Can you walk us through a specific project where you/your team has collected data OR developed models OR deployed models into production? Different use cases?
3. *Main Questions*
    - Data
        - How do you currently collect data? What are your data sources? What kind of data do you use? Have you used APIs or web scraping to collect data? What is the size of your dataset?
        - How do you aggregate, store and access data from different data sources?
        - How do you handle data security and privacy when working with sensitive or confidential data?
        - How do you deal with -
            * Missing or inconsistent data?
            * Unstructured data such as text or images?
            * Imbalanced data?
        - How do you ensure that the data you use is accurate, unbiased and reliable?
        - Can you recall a scenario where you had to manage a large volume of data? Did you implement any measures to accommodate the size of the data?
        - Can you walk me through the process of designing and building a data pipeline in your company? Use case?
        - How do you explore and understand data? Employed approaches? Do you use any data visualisation tools?
        - How have you previously handled situations where data practitioners need to quickly answer questions about their data and what they can do with data?
        - How do you label data and ensure labelling accuracy?
        - Have you versioned your data? If yes, when do you version changes related to data? What is your company's approach to data versioning? Do you store versioned data? If yes, how is it stored?
        - How do you visualise the flow of data in a production pipeline?
    - Feature
        - How do you extract features? Tools used for feature engineering? Could you share your experience with feature engineering and how it has positively impacted the results of previous ML projects?
        - How do you make the decision on which features to extract/create for a given dataset? Could you share your approach used to ensure that the features you extract/create are relevant and are useful for the model performance?
        - How do you keep track of the extracted features and their performance over time? How do you store features? Do you use a feature store?
        - Do you have the same feature manipulation during the training and inference stage?
    - Model
        - How do you divide your data into training, validation and test sets for the model and what is your methodology for doing so?
        - How do you select a specific model for experimentation?
        - Which ML frameworks do you utilise for developing models? Can you provide examples of previous projects where you used a specific framework, and what made it a good fit for the project requirements?
        - What is your approach to setting specific benchmarks and metrics to evaluate the performance of a model? What metrics do you typically use to evaluate the quality of your ML models? How do you assess the success of a model through case studies?

- What steps do you take to optimise the performance of a model, and can you give examples of techniques you have used in the past to improve model performance? How do you determine if further improvements are necessary? Any tools used for hyper-parameter tuning?
- Do you version your code? If yes, how? How do you store your versioned code?
- How do you store information (metadata) related to a model? What information do you store?
- Do you conduct code reviews in your projects, and if so, what is your frequency of code review?
- What is your current method for tracking the various experiments in projects? Do you believe that your current experimentation setup can be completely reproduced? If not, what are the reasons for this?
- How do you validate/test your model? Steps involved?
- How do you handle situations where the model is not suitable for production deployment? How feasible is it to evaluate a completely fresh model in such cases?
- Can you provide a real-life example of an ML project where explainability played a crucial role in its success?
- How do you handle potential bias issues in models and address them?
- How do you keep track of different versions of models ready to be deployed into production?
- How do you ensure the quality of a production-ready model?

• Deployment

- What processes do you follow for promoting a model from the development phase to production? Use Case? Do you have separate development and production environments?
- How close do you deploy your models out to the real-world (once or periodically)? Deployment techniques used?
- Can you describe how the ML solution integrates with other systems and technologies in the organisation? Can you describe some of the infrastructure choices that are necessary for the deployment, hosting, evaluation and maintenance of models? Are these existing in your company? Have you ever used containerisation and orchestration tools such as Docker and Kubernetes?
- Do you have a continuous integration and delivery pipeline for ML models? If yes, how did you set up that?
- In your experience, what are the possible reasons for model failure in production?
- Have you ever had to rollback a ML model in production? If yes, why? How did you resolve the issue?
- When to deploy models after retraining? How do you compare the new and deployed model?
- Have you come across a situation where you had to redeploy a model based on a change in client requirements?

• Monitoring

- How do you monitor and maintain the performance of a deployed machine learning model over time? What do you exactly monitor? What information do you log?
- Have you used any triggering mechanisms and tools for diagnostics and performance monitoring? Are you utilising any visual tools and centralised dashboards for easy monitoring? If yes, which are they?
- How did you address drifts(data/feature/model) in production? Do you use any automation scripts in managing and monitoring models based on drift? Is this process manual in your company?
- How do you approach monitoring and troubleshooting data pipeline issues?
- Can you describe a specific example of a project where you had to handle abrupt changes in data due to external factors and how you overcame any challenges that arose?
- How do you provide alerts when performance of ML models degrades? Is that automated in your company?
- Have you ever encountered an instance where an ML model failed to generalise outside of the training data distribution? If so, how did you handle it?
- How do you handle the integration of new data in your production machine learning models? Mechanisms involved? How do you label new data?
- How do you decide when to retrain a model? How often do you retrain your models? How much time does it require?
- Do you have any mechanism for automatically re-training models in production using fresh data based on live pipeline triggers and feedback loops? Is this process manual or automated in your company?
- Can you explain a scenario where you had to capture additional retraining data for models? In your experience, how much data is needed for retraining?
- How do you ensure data quality and consistency when dealing with frequently updated or changing data sets?

• Organisation

- Do you have any certification procedures in place for your machine learning models and if so, can you provide details on the certification process and why certifying these models is important?
- How do you organise the development and operations team in your company? What competence do they need?
- How do you handle collaboration and communication between your team and other stakeholders?
- How do you share computation resources among team members?
- How do you/your team stay updated on the latest best practices and tools?
- How are you expecting your team to change in the next three months?

4. *Supplementary Questions*

- How do you approach quality assurance of data/model/deployment pipelines? How do you make sure your pipeline is reproducible, explainable and scalable? Can you give an example of a time when you had to scale/reproduce/explain a pipeline?
- What are the different tools used in your pipelines? What is your process for selecting the appropriate tools and platforms from a large pool of options? Can you mention a specific MLOps tool or practice that you have found particularly beneficial in your work on ML models?

- Can you describe any improvements or changes your company has made to its pipeline in recent years?
- Can you tell me about key challenges you have faced when working on machine learning projects? How do you react? And how did you address them?

Appendix B

B.1. MLOps framework and maturity model - Validation interview guide

1. *Objective*: Validate the effectiveness of the developed MLOps framework and maturity model in various organisations.
2. *Background*
    - Can you describe your current role and experience within your organisation?
    - Have you recently adopted MLOps practices in any of your ML projects? If yes, can you share the details?
3. *Assess the current practices in the organisation:*
    - Could you elaborate on the MLOps practices and processes your team currently follows?
    - What are the primary tools and technologies used to facilitate MLOps adoption in your projects?
4. *A brief presentation on the MLOps framework and maturity model*
    - Provide a quick overview of the developed MLOps framework and Maturity model.
5. *Validation:*
    - What is your initial impression of the presented MLOps maturity model and framework? Point out: -
        - Any immediate strengths or weaknesses?
        - Are any components or practices in MLOps missing or underrepresented?
        - Are there any specific aspects that you recommend improving or expanding?
        - Any aspects of the framework resonate with your company? If not, which are they?
    - How does your organisation standardise and automate processes related to data? At what stage is this implemented in your company?
    - Can you elaborate on how your organisation ensures data quality, integrity, and privacy? Should these aspects be explicitly included in the MLOps framework?
    - How do you ensure model quality and assurance? Is it advisable to include these aspects in the MLOps framework?
    - How do you integrate MLOps with DevOps to deliver and deploy product software that includes ML functionality?
    - Can you provide an example of how your organisation has implemented an ML pipeline for automated deployment, inference and retraining in a production environment?
    - How does continuous retraining fit into your use cases? What are your criteria for deciding when to retrain your model?
    - How do you perceive the concept of 'Kaizen' in the context of MLOps? Do you think the continuous improvement in MLOps should have more concrete targets?
    - Could you describe the process of selecting and orchestrating tools as you advance through different stages of MLOps maturity?
    - What roles within your organisation are essential at different stages of MLOps maturity? How are these roles affected by the emphasis on automation?
    - Do you have any final thoughts or suggestions regarding the MLOps framework and its adoption you would like to share?
6. *Conclusion:*
    - Mention any consent for follow-up steps if applicable.

Data availability

The data that has been used is confidential.

References

[1] I. Lee, Y.J. Shin, Machine learning for enterprises: Applications, algorithm selection, and challenges, Bus. Horiz. 63 (2) (2020) 157–170.
[2] B. Akshay, S.R. Pulari, T. Murugesh, S.K. Vasudevan, Machine Learning: A Comprehensive Beginner's Guide, CRC Press, 2024.
[3] What is AI/ML and why does it matter to your business? 2024, https://www.redhat.com/en/blog/what-aiml-and-why-does-it-matter-your-business. (Accessed 30 January 2024).
[4] J. Westenberger, K. Schuler, D. Schlegel, Failure of AI projects: understanding the critical factors, Procedia Comput. Sci. 196 (2022) 69–76.
[5] T. Ermakova, J. Blume, B. Fabian, E. Fomenko, M. Berlin, M. Hauswirth, Beyond the hype: why do data-driven projects fail? 2021.
[6] S. Tatineni, V. Boppana, AI-powered DevOps and MLOps frameworks: Enhancing collaboration, automation, and scalability in machine learning pipelines, J. Artif. Intell. Res. Appl. 1 (2) (2021) 58–88.
[7] M. Zaharia, A. Chen, A. Davidson, A. Ghodsi, S.A. Hong, A. Konwinski, S. Murching, T. Nykodym, P. Ogilvie, M. Parkhe, et al., Accelerating the machine learning lifecycle with MLflow, IEEE Data Eng. Bull. 41 (4) (2018) 39–45.
[8] D. Kreuzberger, N. Kühl, S. Hirschl, Machine learning operations (mlops): Overview, definition, and architecture, IEEE Access 11 (2023) 31866–31879.
[9] M. Testi, M. Ballabio, E. Frontoni, G. Iannello, S. Moccia, P. Soda, G. Vessio, MLOps: a taxonomy and a methodology, IEEE Access 10 (2022) 63606–63618.
[10] M. Treveil, N. Omont, C. Stenac, K. Lefevre, D. Phan, J. Zentici, A. Lavoillotte, M. Miyazaki, L. Heidmann, Introducing MLOps, O'Reilly Media, 2020.
[11] MLOps: Continuous delivery and automation pipelines in machine learning, 2024, https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning. (Accessed 09 February 2024).
[12] S. Idowu, D. Strüber, T. Berger, Asset management in machine learning: A survey, in: 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice, ICSE-SEIP, IEEE, 2021, pp. 51–60.
[13] S. Moreschi, D. Hästbacka, A. Janes, V. Lenarduzzi, D. Taibi, Initial insights on MLOps: Perception and adoption by practitioners, 2024, arXiv preprint arXiv:2408.00463.
[14] A. Lima, L. Monteiro, A.P. Furtado, MLOps: Practices, maturity models, roles, tools, and challenges-a systematic literature review, in: ICEIS (1), 2022, pp. 308–320.
[15] J. Diaz-De-Arcaya, A.I. Torre-Bastida, G. Zárate, R. Miñón, A. Almeida, A joint study of the challenges, opportunities, and roadmap of mlops and aiops: A systematic survey, ACM Comput. Surv. 56 (4) (2023) 1–30.
[16] S.J. Warnett, U. Zdun, Architectural design decisions for machine learning deployment, in: 2022 IEEE 19th International Conference on Software Architecture, ICSA, IEEE, 2022, pp. 90–100.
[17] G. Recupito, F. Pecorelli, G. Catolino, S. Moreschini, D. Di Nucci, F. Palomba, D.A. Tamburri, A multivocal literature review of mlops tools and features, in: 2022 48th Euromicro Conference on Software Engineering and Advanced Applications, SEAA, IEEE, 2022, pp. 84–91.
[18] M.M. John, D. Gillblad, H.H. Olsson, J. Bosch, Advancing MLOps from ad hoc to Kaizen, in: 2023 49th Euromicro Conference on Software Engineering and Advanced Applications, SEAA, IEEE, 2023, pp. 94–101.
[19] C. Hill, R. Bellamy, T. Erickson, M. Burnett, Trials and tribulations of developers of intelligent systems: A field study, in: 2016 IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC, IEEE, 2016, pp. 162–170.
[20] Machine learning workflow, 2024, https://cloud.google.com/ai-platform/docs/ml-solutions-overview. (Accessed 26 September 2024).
[21] R. Ashmore, R. Calinescu, C. Paterson, Assuring the machine learning lifecycle: Desiderata, methods, and challenges, ACM Comput. Surv. 54 (5) (2021) 1–39.
[22] S. Amershi, A. Begel, C. Bird, R. DeLine, H. Gall, E. Kamar, N. Nagappan, B. Nushi, T. Zimmermann, Software engineering for machine learning: A case study, in: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice, ICSE-SEIP, IEEE, 2019, pp. 291–300.
[23] S. Schelter, F. Biessmann, T. Januschowski, D. Salinas, S. Seufert, G. Szarvas, On challenges in machine learning model management, 2015.
[24] A. Paleyes, R.-G. Urma, N.D. Lawrence, Challenges in deploying machine learning: a survey of case studies, ACM Comput. Surv. 55 (6) (2022) 1–29.
[25] N. Gift, A. Deza, Practical MLOps, O'Reilly Media, Inc. 2021.

[26] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, D. Dennison, Hidden technical debt in machine learning systems, Adv. Neural Inf. Process. Syst. 28 (2015).

[27] C. Ebert, G. Gallardo, J. Hernantes, N. Serrano, DevOps, IEEE Softw. 33 (3) (2016) 94–100.

[28] R. Jabbari, N. bin Ali, K. Petersen, B. Tanveer, What is DevOps? A systematic mapping study on definitions and practices, in: Proceedings of the Scientific Workshop Proceedings of XP2016, 2016, pp. 1–11.

[29] K. Salama, J. Kazmierczak, D. Schut, Practitioners guide to MLOps: A framework for continuous delivery and automation of machine learning, 2021, Google Could White paper.

[30] G. Araujo, M. Kalinowski, M. Endler, F. Calefato, Professional insights into benefits and limitations of implementing MLOps principles, 2024, arXiv preprint arXiv:2403.13115.

[31] Y. Zhou, Y. Yu, B. Ding, Towards mlops: A case study of ml pipeline platform, in: 2020 International Conference on Artificial Intelligence and Computer Engineering, ICAICE, IEEE, 2020, pp. 494–500.

[32] Continuous delivery for machine learning, 2024, https://martinfowler.com/articles/cd4ml.html. (Accessed 23 April 2024).

[33] S. Mäkinen, H. Skogström, E. Laaksonen, T. Mikkonen, Who needs MLOps: What data scientists seek to accomplish and how can MLOps help? in: 2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI, WAIN, IEEE, 2021, pp. 109–112.

[34] D.A. Tamburri, Sustainable mlops: Trends and challenges, in: 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC, IEEE, 2020, pp. 17–23.

[35] T.M. Mucha, S. Ma, K. Abhari, Beyond MLOps: The lifecycle of machine learning-based solutions, in: AMCIS, 2022.

[36] Machine learning operations maturity model, 2024, https://docs.microsoft.com/en-us/azure/architecture/example-scenario/mlops/mlops-maturity-model. (Accessed 23 April 2024).

[37] MLOps foundation roadmap for enterprises with amazon SageMaker, 2024, https://aws.amazon.com/blogs/machine-learning/mlops-foundation-roadmap-for-enterprises-with-amazon-sagemaker/. (Accessed 23 September 2024).

[38] J.W. Creswell, J.D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Sage publications, 2017.

[39] D.R. Hancock, B. Algozzine, J.H. Lim, Doing Case Study Research: A Practical Guide for Beginning Researchers, Teachers College Press, 2021.

[40] S. Baskarada, Qualitative case study guidelines, Qual. Rep. 19 (40) (2014) 1–25.

[41] G. Thomas, A typology for the case study in social science following a review of definition, discourse, and structure, Qual. Inq. 17 (6) (2011) 511–521.

[42] J.M. Bass, S. Beecham, J. Noll, Experience of industry case studies: A comparison of multi-case and embedded case study methods, in: Proceedings of the 6th International Workshop on Conducting Empirical Studies in Industry, 2018, pp. 13–20.

[43] J. Gerring, What is a case study and what is it good for? Am. Political Sci. Rev. 98 (2) (2004) 341–354.

[44] R.K. Yin, Case Study Research: Design and Methods, vol. 5, sage, 2009.

[45] AI Sweden, 2024, https://www.ai.se/en. (Accessed 16 January 2024).

[46] M.D. Myers, M. Newman, The qualitative interview in IS research: Examining the craft, Inf. Organ. 17 (1) (2007) 2–26.

[47] S.E. Hove, B. Anda, Experiences from conducting semi-structured interviews in empirical software engineering research, in: 11th IEEE International Software Metrics Symposium, METRICS'05, IEEE, 2005, p. 10.

[48] J. Seawright, J. Gerring, Case selection techniques in case study research: A menu of qualitative and quantitative options, Political Res. Q. 61 (2) (2008) 294–308.

[49] M.M. John, H.H. Olsson, J. Bosch, D. Gillblad, Exploring trade-offs in mlops adoption, in: 2023 30th Asia-Pacific Software Engineering Conference, APSEC, IEEE, 2023, pp. 369–375.

[50] Software center, 2024, https://www.software-center.se/. (Accessed 16 January 2024).

[51] J.A. Holton, The coding process and its challenges, Sage Handb. Grounded Theory 3 (2007) 265–289.

[52] W. Virginia, Research methods: Triangulation, Evid. Based Libr. Inf. Pr. 11 (1) (2016) 66–68.

[53] M.M. John, H.H. Olsson, J. Bosch, Towards mlops: A framework and maturity model, in: 2021 47th Euromicro Conference on Software Engineering and Advanced Applications, SEAA, IEEE, 2021, pp. 1–8.

[54] Getting started with MLOps: Selecting the right capabilities for your use case, 2024, https://cloud.google.com/blog/products/ai-machine-learning/select-the-right-mlops-capabilities-for-your-ml-use-case. (Accessed 28 March 2024).

[55] Creation of a taxonomy for the European AI ecosystem, 2024, https://eit.europa.eu/sites/default/files/creation_of_a_taxonomy_for_the_european_ai_ecosystem_final.pdf. (Accessed 05 February 2024).

[56] Compliance considerations when building machine learning models, 2024, https://www.linkedin.com/pulse/compliance-considerations-when-building-machine-rishikesh-psafc/. (Accessed 28 March 2024).

[57] B. Dhamodharan, Driving business value with AI: A framework for MLOps-driven enterprise adoption, Int. J. Sustain. Dev. Comput. Sci. 5 (4) (2023) 1–10.

[58] M.A. Mateo Casalí, F. Fraile Gil, A. Boza, A. Nazarenko, An industry maturity model for implementing machine learning operations in manufacturing, Int. J. Prod. Manag. Eng. 11 (2) (2023) 179–186.

[59] M. Weber, J. Schniertshauer, L. Przybilla, A. Hein, J. Weking, H. Krcmar, Streamlining the operation of AI systems: Examining MLOps maturity at an automotive firm, 2024.

[60] E.-M. Ihantola, L.-A. Kihn, Threats to validity and reliability in mixed methods accounting research, Qual. Res. Account. Manag. 8 (1) (2011) 39–58.

[61] A. Bodor, M. Hnida, N. Daoudi, Machine learning models monitoring in MLOps context: Metrics and tools, Int. J. Interact. Mob. Technol. 17 (23) (2023).

[62] S.J. Warnett, E. Ntentos, U. Zdun, A model-driven, metrics-based approach to assessing support for quality aspects in MLOps system architectures, J. Syst. Softw. 220 (2025) 112257.

[63] P. Liang, B. Song, X. Zhan, Z. Chen, J. Yuan, Automating the training and deployment of models in mlops by integrating systems with machine learning, 2024, arXiv preprint arXiv:2405.09819.

[64] A.M. Burgueño-Romero, A. Benítez-Hidalgo, C. Barba-González, J.F. Aldana-Montes, Towards an open-source MLOps architecture, IEEE Softw. (2024).