



A Joint Study of the Challenges, Opportunities, and Roadmap of MLOps and AIOps: A Systematic Survey

JOSU DIAZ-DE-ARCAYA, ANA I. TORRE-BASTIDA, GORKA ZÁRATE, and
RAÚL MIÑÓN, Tecnalia, Basque Research & Technology Alliance (BRTA), Spain
AITOR ALMEIDA, University of Deusto, Spain

Data science projects represent a greater challenge than software engineering for organizations pursuing their adoption. The diverse stakeholders involved emphasize the need for a collaborative culture in organizations. This article aims to offer joint insights into the role of MLOps and AIOps methodologies for raising the success of data science projects in various fields, ranging from pure research to more traditional industries. We analyze the open issues, opportunities, and future trends organizations face when implementing MLOps and AIOps. Then, the frameworks and architectures that promote these paradigms are presented, as are the different fields in which they are being utilized. This systematic review was conducted using an automated procedure that identified 44,903 records, which were filtered down to 93 studies. These articles are meant to better clarify the problem at hand and highlight the future areas in both research and industry in which MLOPs and AIOps are thriving. Our findings indicate that AIOps flourish in challenging circumstances like those presented by 5G and 6G technologies, whereas MLOps is more prevalent in traditional industrial environments. The use of AIOps in certain stages of the ML lifecycle, such as deployment, remains underrepresented in scientific literature.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence; Machine learning**; • **Software and its engineering** → **Software creation and management**; • **General and reference** → **Surveys and overviews**;

Additional Key Words and Phrases: MLOps, AIOps, challenges, opportunities, future trends, frameworks, architectures, ML, AI, systematic survey, SLR

ACM Reference format:

Josu Diaz-de-Arcaya, Ana I. Torre-Bastida, Gorka Zárate, Raúl Miñón, and Aitor Almeida. 2023. A Joint Study of the Challenges, Opportunities, and Roadmap of MLOps and AIOps: A Systematic Survey. *ACM Comput. Surv.* 56, 4, Article 84 (October 2023), 30 pages.
<https://doi.org/10.1145/3625289>

This work was partially financed by the Basque Government through their Elkartek program (EGIA project, ref. KK-2022/00119). The work of Aitor Almeida was funded by the Spanish Ministry of Science and Innovation (INCEPTION project, ref. PID2021-128969OB-I00).

Authors' addresses: J. Diaz-de-Arcaya, A. I. Torre-Bastida, G. Zárate, and R. Miñón, Tecnalia, Basque Research & Technology Alliance (BRTA), Albert Einstein Kalea 28, 01510, Miñano, Álava, Spain, 01510; e-mails: {josu.diazdearcaya, isabel.torre, gorka.zarate, raul.minon}@tecnalia.com; A. Almeida, University of Deusto, Avenida de las Universidades, 48007, Bilbao, Biscay, Spain; e-mail: aitor.almeida@deusto.es.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

0360-0300/2023/10-ART84 \$15.00

<https://doi.org/10.1145/3625289>

1 INTRODUCTION

Nowadays, **artificial intelligence (AI)** is increasingly commonplace, and disruptive AI solutions and developments are rapidly infiltrating comprehensively into multiple fields of human activities and domains. Beyond expectations, AI service development requires skilled professionals, quality labeled data, and rigorous development processes. In this context, it becomes necessary to highlight two crucial phases in the generation of AI-based solutions: the (i) development of the AI solution as a software project, and the (ii) accurate planning of the implementation and continuous deployment costs. Furthermore, AI models in production environments are a small part of a much larger ecosystem, the deployment and commissioning of AI models from a pure software perspective must be considered, and it is as important as the model itself. In addition, many experts share the vision that the development of AI-based solutions is equal to that of a pure software project. However, this is far from accurate since the development and operationalization of AI-based solutions have their own lifecycle, requirements, and particularities.

In the fields of software development and operations, good practices and methodologies have existed for a long time. The best known being the DevOps philosophy [38], an acronym formed by the words development and operations, which aspires to narrow the gap between these areas of expertise. Its goal is to accelerate the software development lifecycle while providing high-quality continuously delivered software components. This mature and widely adopted methodology encompasses and transcends agile software development and there exist multiple technological solutions implementing it. However, due to the rise of new technological paradigms such as Big Data, AI and 5G, DevOps methodologies need to be adapted to these new lifecycles and processes which expand beyond pure software solutions.

This gap materializes in one of the biggest challenges in AI practices today: the operationalization of AI solutions. Many organizations are desperate to figure out how to convert the insights discovered by data scientists into tangible value for their businesses, which has proven to be a cumbersome endeavor. It requires unifying multiple processes across diverse teams, starting with defining business goals and continuing all the way through data acquisition, and model development and deployment. This unification is achieved through a set of best practices for communication and collaboration between the data engineers who obtain the data, the data scientists who prepare the data and develop the model, and the operations professionals who serve the models. It is in this intersection where MLOps was conceived, aspiring to streamline the process of AI operationalization, whereas AIOps is a viable solution to the upsurge of growing IT infrastructures and data. Both are the main exponents of an explosion of Ops variants related to AI, as they can be LAOps [83] or RLOps [72]. Therefore, we focus this survey on these two, instead of extending it to other methodologies that have not yet proved to be of interest to organizations.

The objective of this **systematic literature review (SLR)** is precisely to analyze the challenges and opportunities of these methodologies, in order to promote their benefits and adoption in both industry and academia. The expected result of this work is that those who read it can advance in their different domains and areas of expertise, due to being clearer about the main concepts of MLOps standardization, and the possibilities of providing intelligence and automation to their processes through AIOps. Finally, the various frameworks and domains in which MLOps and AIOps are thriving are shown in detail.

1.1 Comparison between MLOps and AIOps

Due to the fact that both disciplines are relatively young and their definitions differ depending on the source, it is challenging to establish a comparison between them. Furthermore, in spite of their substantial differences, there are some overlaps in the teams, skills, and challenges required to successfully implement them. First, MLOps is a well-established term in industry, although

Table 1. Comparison Framework between the MLOps and AIOps Paradigms

	MLOps	AIOps
Inception	2020	2019
Concept	Machine Learning Operations.	AI for IT Operations.
Goal	Creating stable products from ML prototypes.	Cope with the growing complexities posed by modern IT systems, reduce human intervention.
Enablers	CI/CD, orchestration, reproducibility, collaboration, continuous monitoring.	Big Data, cloud computing, observability.
Focus	Multidisciplinary.	Information Technology.
Use Cases	Failure prediction, autonomous vehicles, defect detection, electricity.	Spikes in demand, maximize uptime, optimize environments, improve efficiency.

it was first mentioned in scientific literature in 2020, a contemporary definition is provided by the authors of [69], in which the authors argue that MLOps leverage machine learning, DevOps, and data engineering, with the goal of productionizing machine learning systems by facilitating the creation of machine learning products. It stands on principles such as CI/CD, collaboration, orchestration, reproducibility (data, model, and code versioning), and continuous monitoring. The use cases in which the MLOps paradigms are applied vary greatly, from failure prediction, self-driving vehicles, and defect detection [63] to the electricity market [106]. On the other hand, the first mention of AIOps in the scientific literature dates back to 2019. The authors in [99] argue there is no generally accepted definition of AIOps in the scientific literature and highlight the importance of artificial intelligence to create self-learning and self-healing applications and infrastructure in the field of IT. Furthermore, AIOps is regarded as the only viable solution to deal with the expanding IT infrastructure, according to the findings of [25]. In this regard, observability is crucial for the adoption of AI-based solutions, and is being conducted in various areas, including distributed cloud applications [114]. Due to the scarcity of manuscripts on AIOps in the scientific literature, we have explored the gray literature and how the major providers in the market approach AIOps to gain a better understanding of the evolution of this paradigm. IBM provides various AIOps solutions for the following case studies: scaling performance and sustainability, optimizing IT environments, maximizing uptime, and managing spikes in demand [55]. For AWS, the goal of AIOps is to reduce human intervention in **Information Technology (IT)** operations processes. Microsoft claims that AIOps helps achieve high quality and efficiency with less human intervention [79]. In this study by Forrester, commissioned by Google [27], they elaborate that AIOps is key to solving cloud operation challenges and improving efficiency and productivity.

On the basis on these findings, a comparative framework between MLOps and AIOps is proposed in Table 1, which aims to serve as the stepping stone for the reader to dive into this survey. It is noteworthy to mention that while the first appearance of these two terms in scientific literature is very recent, earlier research has examined the challenges posed by them.

1.2 Objectives

This SLR is framed in the fields of MLOps and AIOps. The motivation for doing this SLR are: to determine the causes that are preventing these fields from having a more widespread adoption in industry and academia, to understand how to overcome these issues by revealing the opportunities that are currently arising for MLOps and AIOps, to discover the architectures and frameworks that may aid to embrace these methodologies, and to have a clear picture of which are the current and

Table 2. Research Questions and their Motivations

	Question	Motivation
Q1	What are the open issues, challenges and particularities in MLOps and AIOps?	To have a clear view of the inherent difficulties of embracing MLOps and AIOps, so adopters can prepare accordingly.
Q2	What are the opportunities and future trends in MLOps?	To understand which are the areas that may be interesting to explore in the field of MLOps.
Q3	What are the opportunities and future trends in AIOps?	To understand which are the areas that may be interesting to explore in the field of AIOps.
Q4	What frameworks and architectures facilitate MLOps and AIOps?	To have a better understanding of how industry and academia are utilizing MLOps and AIOps from a practical standpoint.
Q5	What are the current and future fields in which MLOps and AIOps are thriving?	To determine the areas in which MLOps and AIOps are being applied and considered.

future areas in which these methodologies are being applied. Due to this, in Table 2 we have elaborated a set of research questions and their purpose.

The rest of this paper is structured as follows: Section 2 covers the methodology utilized in this systematic literature review, diving into the details about how the manuscripts have been retrieved and selected. An overview of the selected studies can be found in Section 3. In Section 4 the studies that have been included in the review are used to offer insights into the selected research questions displayed in Table 2. Finally, the conclusions of this research are summarized in Section 6.

2 METHODS

In [65] the authors summarize the required phases for performing an SLR, which we have taken as a reference in this manuscript. The research questions are specified, and a review protocol is developed during the planning phase. Next, the primary studies that are to be included in the review and a quality assessment on them is performed during the conducting phase. Finally, the formatting of the main report takes place in the reporting phase.

2.1 Article Retrieval

Automatic retrieval of manuscripts provides a more comprehensive, less error-prone, and unbiased search. To do this, we have considered and tested various repositories, such as arXiv [112], Springer [105], and IEEE [57]. However, not all the necessary features were available, and others, such as MDPI and ACM do not expose their databases through dedicated services. This is when initiatives like Crossref [28] come in handy because they provide a unified point of access for the published manuscripts of the various editorials; hence, only a single client is necessary for performing the automated search, which streamlines the development process. On the downside, not all the editorials deposit all the metadata for their manuscripts; vital information used in this SLR, such as the paper abstract, is not always available. Finally, Elsevier [19] through its renowned Scopus service, provides metadata search for a wide variety of editorials; abstract information is available; and its API is feature rich. For this reason, we have developed a client service that interacts with the Scopus API to gather all the manuscript metadata that has been used in the selection process. The source code for this client, along with the documentation for its use, is publicly available on GitHub [34].

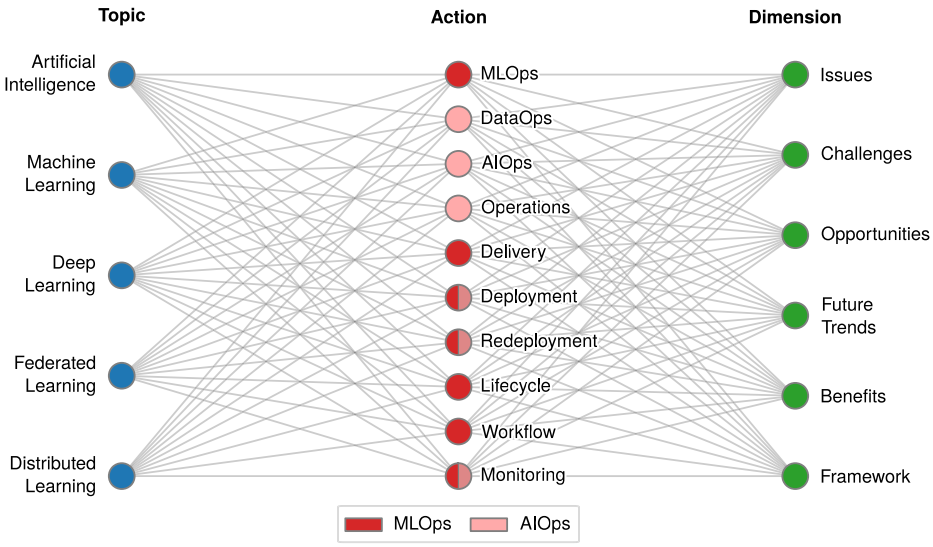


Fig. 1. Search queries utilized to identify the manuscripts that are included in the SLR.

```
'AIOps' AND ( 'Failure Prevention' OR 'Failure Prediction' OR 'Failure Detection' OR
'Root Cause Analysis' OR 'Remediation' OR 'Resource Consolidation' OR
'Scheduling' OR 'Power Management' OR 'Service Composition' OR
'Workload Estimation' OR 'Event Correlation' OR 'Failure Mitigation' )
```

Fig. 2. Search string utilized to identify high-quality studies on AIOps.

2.2 Search Terms

The search process has been the result of the combination of concepts in the different columns defined in Figure 1. Each of them has been selected by the authors based on their expertise in the field. The blue column stands for the main topic and situates the search in the right field. The purpose of the red column is to narrow the search towards actions addressing MLOps, depicted in a darker shade of red, or AIOps, depicted in a lighter shade of red. Additionally, some actions pertinent to both fields contain both colors. Finally, the green column is the dimension, and it serves the purpose of providing a better alignment of the search string with the research questions outlined above. Following this method, t300 search queries are constructed, and we utilize the search capabilities provided by Scopus to find articles that contain all three terms of the search query in the title, abstract, or keywords. The search has been narrowed to articles published between 2018 and 2023. One of the benefits of performing an automated search of scientific databases is that the searched space is considerably wider than that of a manual approach. For this reason, the three separate groups that cover topic, action, and dimension make up for 600 different search strings.

While the search procedure described above reveals a vast array of studies, works about AIOps are significantly underrepresented in the scientific literature in comparison to MLOps. Due to this, yet another query specifically tailored to discover those studies addressing AIOps has been applied. The authors chose the search terms for this procedure using both their professional knowledge and relevant scientific literature [25, 84]. Figure 2 depicts the search string utilized to discover high-quality studies in the field of AIOps.

Table 3. Study Selection Criteria

Inclusion Criteria	
I1	Published between 2018 and 2023.
I2	Identified by the search queries.
I3	Articles presenting new interesting ideas.
I4	Manuscripts closely aligned with the topic.
Exclusion Criteria	
E1	Publication not in English.
E2	Retracted publications.
E3	Publisher not aligned with the study.
E4	Publications behind a paywall that cannot be retrieved.
E5	Articles with insufficient citations.

2.3 Selection Criteria

Table 3 summarizes the inclusion and exclusion criteria that has been applied in this systematic review to narrow down the number of articles and analyze the most relevant ones for the topic in hand. Considering that both MLOps and AIOps are recent concepts, and the upsurge of new paradigms such as cloud or edge, which introduce new concepts in the field of AI operationalization, only studies from 2018 onward have been considered.

Therefore, the list of articles included must have been identified in the automatic search and hold interesting or groundbreaking ideas aligned with the topic of the review. On the other hand, articles not written in English and retracted publications are not included. In addition, studies being released by publishers clearly not aligned with the main topic are excluded. We have made every effort to obtain the articles with the subscriptions available by the authors. However, any article behind a paywall or unavailable becomes excluded as well. Finally, an automated minimum threshold based on the number of citations is applied to the articles. Manuscripts from 2022 onward are not filtered as we understand this would leave out interesting research that has not yet reached a wider audience; articles from 2021 must be cited at least once, two cites for articles from 2020, three for 2019, and four for 2018. This filter promotes the inclusion of newer ideas and preserves the freshness of this systematic review.

2.4 Quality Metrics

One of the defining features of a systematic review is to give an appraisal of the quality of the studies included [6]. To this end, in Table 4 we have listed a set of quality metrics to measure the included manuscripts and rank them. These metrics have been gathered from various sources, such as PRISMA [91, 92], other literature reviews [9, 113], and our expertise on the field.

After evaluating these metrics upon the included articles, each of them ends up with a total score between 0 and 10. We classify them as deficient (0-2), sufficient (3-4), good (5-6), very good (7-8), and excellent (9-10). Please note that the mark does not directly relate to the quality of the article, instead it provides an approximation of its alignment with the goals of this SLR.

3 RESULTS

This section is devoted to supplying an overview of the selected studies. In Section 3.1 the methodology for the retrieval and selection of the manuscripts included in this SLR is offered. Next, the risk of bias and validity threats are explained in Section 3.2. Finally, Section 3.3 discusses a high-level overview of the selected studies.

Table 4. Quality Metrics for Evaluating the Articles

Quality Metrics	Value	Weight
M1 It provides a comprehensive state of the art aligned with this SLR, and identifies the knowledge gaps to justify the reason of the study.	0/1	1
M2 The research is validated against at least one use case aligned with the objectives of this SLR.	0/1	2
M3 The number of research questions that are tightly coupled with the manuscript.	[0..5]	2
M4 The manuscript has been published under an open license.	0/1	1
M5 Publication type (Other / Journal).	0/1	1
M6 The manuscript has been published in a journal in the first JIF quartile.	0/1	1
M7 Citation count.	[0..3]	2

3.1 Study Selection

The PRISMA statement [91, 92] emphasizes the need to describe the results of the search and selection process potentially by using a flow diagram. In [50], the authors highlight again the importance of such a diagram in systematic reviews to facilitate for the reader a rapid understanding of the core procedures utilized. In addition, they also provide an application which we have used to generate the diagram in Figure 3. Two phases have been conducted to obtain the final number of articles included in this systematic review: an identification phase, and a screening phase, after which the remaining studies are selected for further analysis.

As mentioned in the previous section, the search process, retrieval, and storage of the meta-data associated with the studies has been made programmatically. Due to this, a daunting total of 44,903 articles have been identified by the search queries detailed in Table 1. It is worth noting that duplicated records were already filtered during the retrieval process, hence they are not included in this number. Then, a series of filters have been applied before the screening phase. There are 453 records for which no title has been obtained, 9 retracted articles, and 36 erratum articles. The publisher not aligned with the field of the study has resulted in 12,863 manuscripts being excluded. 1,057 full conference proceedings are excluded as we understand the meaningful articles should have been already identified in this research. Finally, we have applied an automated filter to prioritize recent articles over older ones based on the citation count. This results in 23,508 being handed over to the screening phase, in which the inclusion and exclusion criteria detailed in Table 3 is to be applied. This results in 21,940 studies being excluded due to their title, and 1,111 due to their abstract. Up to this point, the analysis has been performed purely on the metadata being retrieved from Scopus by the automated search client. The remaining 452 papers have been downloaded for deeper analysis; after screening all the papers the authors selected 93 articles for further analysis in this systematic review.

Table 5 showcases the various stages executed during the selection process to obtain the included studies. During stage 1, we have solely used Scopus which takes away the nuances between the different databases and their APIs (e.g., some databases do not provide the possibility of performing complex queries, others do not offer an API, some establish a threshold on the maximum queries). We deem this appropriate as Scopus integrates a large database of the most relevant publishers in the field of this SLR. Stage 2 has been performed entirely by directly applying the filters on the database itself. Stages 3 and 4 have been executed by the authors by showing merely the information required for that particular stage (i.e., the title in stage 3, and the abstract in stage 4). Finally, a full reading of the article has been done in the final stage, which results in the studies finally included as part of this review. These researches are then fully

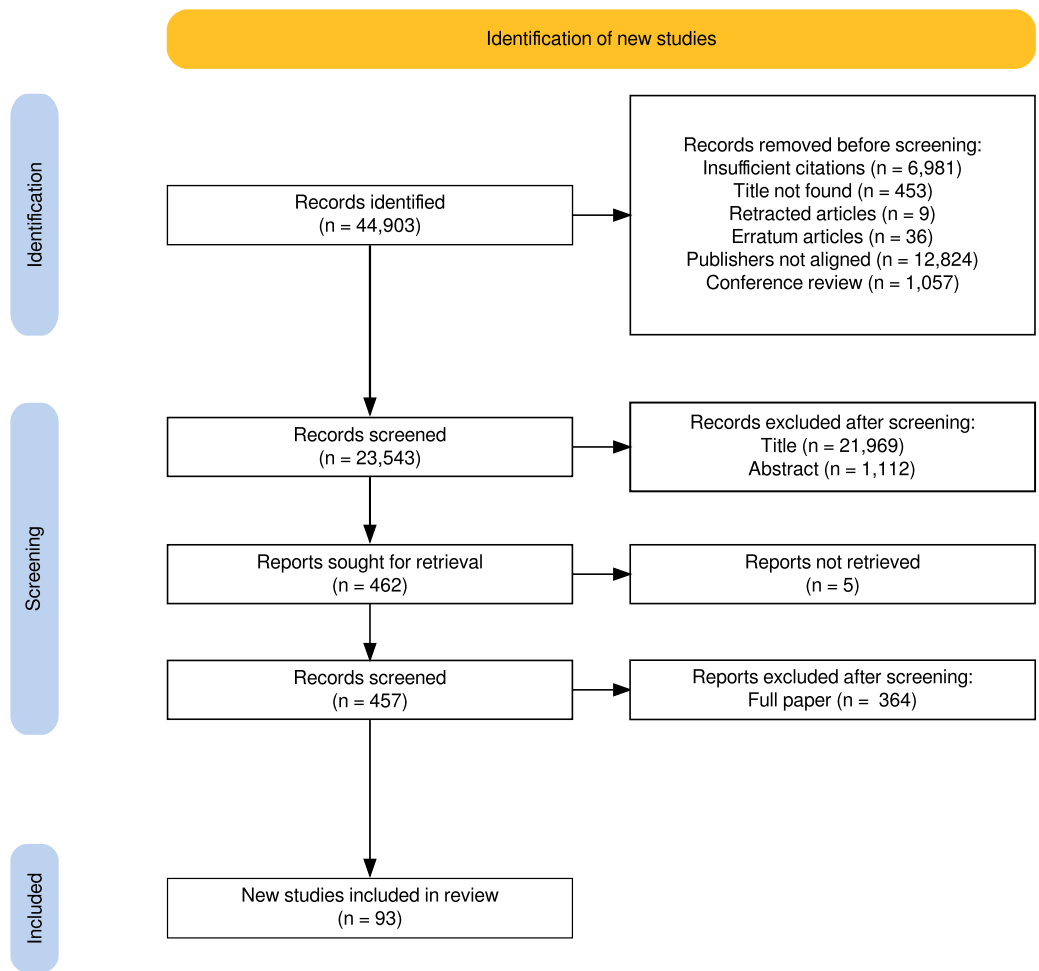


Fig. 3. Flow chart showing study selection.

Table 5. The Manuscripts Filtered at the Various Stages Broken Down by Editorial

Stage	Process	Selection Criteria	ACM	Elsevier	Springer	IEEE	MDPI	Others	Total
1	Identification	Search terms	1,612	6,666	5,729	11,959	3,463	15,474	44,903
2	Identification	Filtering	1,064	6,121	3,961	8,316	3,098	963	23,543
3	Screening	Title	115	344	287	618	158	52	1,574
4	Screening	Abstract	42	85	87	192	37	19	462
5	Screening	Full paper	10	9	10	50	7	7	93

read and ranked against the metrics defined in Table 4, the result of this work is showcased in Table 9.

During the final screening phase, there are various manuscripts that even though they appear to comply with the criteria in Table 3 they were finally not included in this SLR. In [127] the

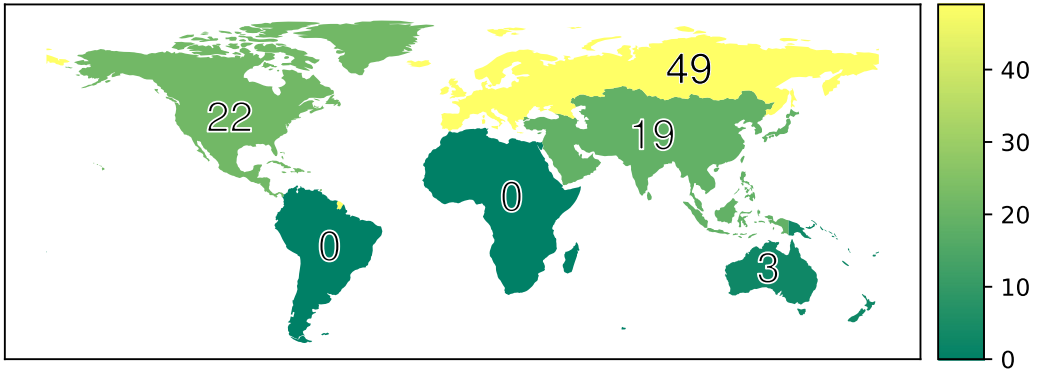


Fig. 4. Geographical distribution of the selected studies.

authors provide a survey on ML testing, and outline research trends and challenges in the field. A DNN tuning framework on mobile frameworks is presented in [122]. In [26] the authors propose a framework to provide forecasting both at the edge and cloud layers. Future directions at the intersection of machine learning and operations management are highlighted in [13]. Even though we found these manuscripts fascinating they were not included as part of this SLR due to the impossibility to fit their findings within the research questions described in Table 2. However, we cite them because we reckon they mention insightful ideas and may be suited for similar researches.

3.2 Risk of Bias

The fact that the initial search process has been performed programmatically diminishes the risk of the researcher bias in the selection of the studies. Even though we acknowledge the chance of relevant manuscripts not being identified by the automated process, the sheer number of studies identified by the wide search queries outlined in Figure 1 constitute sufficient research material for this SLR. In addition, the reason for not having included non-academic sources such as gray literature into this study is that this research does not suffer from low volume nor quality of evidence [41]. Finally, the selected subset of studies has been qualitatively analyzed following the criteria in Table 4, for which we have positively valued quantitative statements. Furthermore, several rounds have been performed by the authors of this study during the various stages to minimize personal preferences and biases.

3.3 Overview

In terms of geographical distribution, a significant number of the selected papers have been published in Europe ($\approx 51\%$) as shown in Figure 4, followed by North America ($\approx 23\%$) and Asia ($\approx 20\%$). Only three papers out of the 93 have been finally selected in Oceania. Unfortunately, no manuscript from South America nor Africa has reached the final stage of selected papers. We reckon that a plausible explanation for this distribution is that considering that all the authors of this systematic review reside in Europe, the search queries defined in Figure 1 have a closer alignment with the research promoted in this continent. Next, MLOps, AIOps, and AI operationalization are in fact hot topics in North America, Europe, and Asia.

With regards to the nature of the selected manuscripts, the majority of them are conference papers ($\approx 57\%$), followed by journal articles ($\approx 41\%$). This distribution is pertinent to this research since one of the quality metrics defined in Table 4 is in fact the publication type, for which we favor research published in journals as they tend to be longer pieces of work with more insightful results.

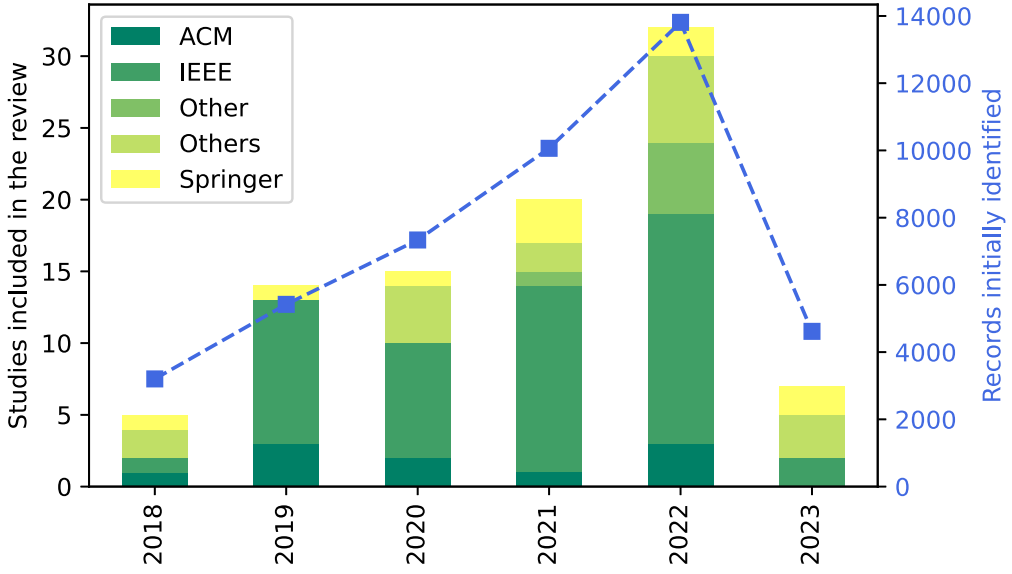


Fig. 5. Cumulative number of manuscripts by year and publisher.

The data presented in Figure 5 is broken down by year and includes two pieces of information: the (i) number of manuscripts included in this research aggregated by publisher represented in the left-hand vertical axis, and the (ii) total records initially identified showcased in the right-hand vertical axis. The former shows an increasing trend in the interest in the topics discussed as part of this research in recent years, with a steady upwards slope between 2018 and 2020, and a milder slope between 2020 and 2022. The latter offers a higher granularity on the number of articles published by the different editorials over the years.

Finally, in Table 9 there are five manuscripts that stand out from the rest after evaluating the quality metrics defined in Table 4. The authors in [29, 59] circumscribe their work to AI operationalization in 5G and 6G environments, respectively. Similarly, the deployment of ML solutions for network traffic classification is described in [90]. In [18, 118] the authors introduce their frameworks for the deployment and redeployment of smart algorithms in production systems.

4 DISCUSSION

This section thoroughly examines the studies included in this review and classifies their insights against the research questions described in Table 2. The open issues and challenges of MLOps and AIOps are discussed in Section 4.1. Next, the opportunities and future trends of MLOps and AIOps are outlined in Section 4.2, and Section 4.3, respectively. In Section 4.4, different frameworks and architectures for both paradigms can be found. Finally, in Section 4.5 the fields in which MLOps and AIOps thrive are described.

4.1 What are the Open Issues, Challenges, and Particularities in MLOps and AIOps? (RQ1)

Due to the relative youth of these two paradigms, scientific literature concerning their issues and challenges exhibits many similarities. For this reason, this section provides insights from a joint perspective. The implementation of MLOps and AIOps based projects in production environments

is cumbersome, and a combination of cross-domain skills and a collaborative culture is necessary to accomplish such an endeavor [21, 104]. One of the main problems is that data scientists are unfamiliar with the unique requirements and characteristics of some environments [17] and the level of understanding they must acquire in order to build a solution that includes business value, data, system, and process integration considerations [30], and infrastructure [77]. Unlike in research, industry requires a balance between accuracy and processing, and systems tend to be more complex [81]. Next, software developers struggle to acquire ML skills due to its interdisciplinary nature [2] and their lack of ability in feature engineering, parameter tuning, and model selection; combined with the myriad ML libraries and frameworks [14]. Finally, the complexity of ML workflows forces operations engineers to have a high degree of application and platform expertise to size, provision, and operate the required resources [22, 45]. In summary, this complexity is stemmed by the difficulty to find within teams the large variety of skills required to operationalize data science projects [5, 74], and the different backgrounds and knowledge of the involved stakeholders [7, 64, 99, 129]; hence, the high value of having a strong background in both software engineering and data science [49] for the project to succeed.

4.1.1 Computer Engineering. On the software side, a more streamlined and systematic approach to AI application development and lifecycle management is needed [54], and the interaction between software engineering and ML workflow activities needs to be better reported [75]. However, data scientists are not necessarily computer engineers by training [109], which threatens the maintenance and evolution of AI solutions. In fact, they often fail to understand the different frameworks utilized during the development process [47] and the unique requirements and characteristics of certain environments [17]. In this regard, experienced IT professionals still question the efficiency of ML-based solutions [99], partly due to the lack of explainability of such systems [1]. In addition, the patterns and practices used in software engineering do not fit those required in data science [30] and introduce the challenges of data availability, concept and data drift [60], and scalability [102], leading to failures and reduced model accuracy in ML deployments [93]. For instance, the complexity of AI designs represents a bigger challenge [39], such as in terms of modularization [5], than traditional software components and requires a working understanding of ML principles and proficient technical expertise [43]. In addition, common frameworks utilized in software engineering, such as source code versioning tools, fail short when being used in data science projects due to their close relationship with the associated data [74]. For instance, in spite of the popularity of GitHub as a hosting platform, there is a lack of open-source ML-enabled projects leveraging GitHub Actions [20]. In addition, version control management, metainformation definition, and data and model governance need to be addressed [126].

4.1.2 Data Management. The importance of data management cannot be undermined in both methodologies. Aspects such as data quality [77, 99], data access [15, 49, 77], data preparation and labeling [2, 74, 121], data validation and cleaning [97], limited availability of datasets [1], or having to integrate the data from different sources [76, 129]; are often overlooked and may consume a large part of the project's time and budget [49, 74] and may require the involvement of experts in the field [121]. In addition, not only is there a tight dependency between the developed models and the associated data [94], but also the data discovery, management, and versioning are more complex in ML-based scenarios [5, 74]. The Big Data paradigm [4, 14, 30, 40, 62] adds yet another layer of complexity to data management in ML projects: aspects such as task distribution and data movement [62], or batch processing [14], become even more relevant in these scenarios. In addition, the underlying complexity of ML and Big Data workflows resorts to time-consuming, convoluted coding [4]. Due to this, the joint work of industry and academia is beneficial in these endeavors [30].

4.1.3 Orchestrating the ML Lifecycle. An efficient distribution, parallelization, and orchestration of the data and tasks of an ML solution are of paramount importance for the success of any project [62, 87]. In this regard, innovative technologies for the execution of distributed tasks, such as serverless, promote the use of multistage machine learning solutions by guaranteeing optimal response and running time [8]. On the other hand, opting for a distributed deployment requires dealing with the traffic overhead, bandwidth, and latency delays of the deployed services [123], as well as the scalability of the computation resources associated with the deployment [16, 90]. These pros and cons are particularly relevant when utilizing different computational layers for the deployment of data science projects. For instance, metaheuristic techniques can be used in cloud computing deployments to maximize the degree of isolation and resource sharing among the components [88]. In addition, cloud deployments can be extended to the edge of the network, but the communications (e.g., bandwidth, latency, connectivity) [39, 52, 119, 123, 125], scarce computing resources [40, 47, 52, 71, 111, 123], storage space [111], memory [40, 47, 78, 123], security [42], privacy [42, 125], and energy consumption [47, 52, 78, 78, 87, 119, 125] where GPU dependent applications tend to be particularly energy-hungry [39]; need to be thoroughly scrutinized in order to achieve the best performance. The use of edge computing and IoT [66] technologies keeps rising, but represents a challenge due to the restrained resources found in this computational layer, hence they need to be used efficiently [52]. In addition, protecting the intellectual property and integrity of the deployed models is complex due to the distributed nature of edge devices. The aforementioned drawbacks lead to data often having to be transferred over to more resourceful infrastructural devices [52]. In addition, edge nodes often operate in unreliable environments [70] and need to undergo cumbersome manual optimizations [18]. Containerized technologies are often utilized for delivering the ML solution, but more emphasis is required on dependency and filesystem management [89]. In general, the lifecycle of data science projects must leverage the following aspects: (i) ML workflows often operate in unreliable, harsh, and constrained environments [70, 87, 117, 120], and (ii) the trade-off between the benefits of using high computational resources and more humble architectures [118].

4.1.4 Hardware and Architectures. Another key aspect to consider in MLOps and AIOps is the infrastructural devices available during the various stages of the ML lifecycle, where an appropriate configuration of the hardware represents a challenge in DNN solutions [18, 107]. The large heterogeneity of the hardware platforms [14], their diverse characteristics [115], and restrictions [120] such as limited performance or memory [40], makes hardware configuration an important task to consider in every data science project. On the other hand, the use of technical software [16], incompatibilities between ML libraries [124], and the different architectures and resources imposed by the models [116]; pose yet another threat to the materialization of the project. In addition, yet another challenge is the differences in architecture designs and implementation of the various framework found in the development and deployment of the data science solution [47]. The security and reliability of the ML workflow needs to be assessed from a data perspective [81], and from the degree of isolation of the different components forming it [88]. In the past, computational power has slowed down the development of **Large Language Models (LLMs)** and conversational agents, whereas now the challenges are more related to the trustworthiness of the results [101].

4.1.5 Monitoring. The different components that comprise data science projects require continuous monitoring in order to detect deviations during their runtime [53, 74]; hence, the use of dedicated tools is mandatory [56, 81]. The static nature of trained ML models severely affects dynamically changing environments [98]. Due to this, ML models must be responsive to changes such as component wear and aging [94], and data biases [81] which can result in models becoming outdated due to concept drift [96]. However, the definition of performance metrics

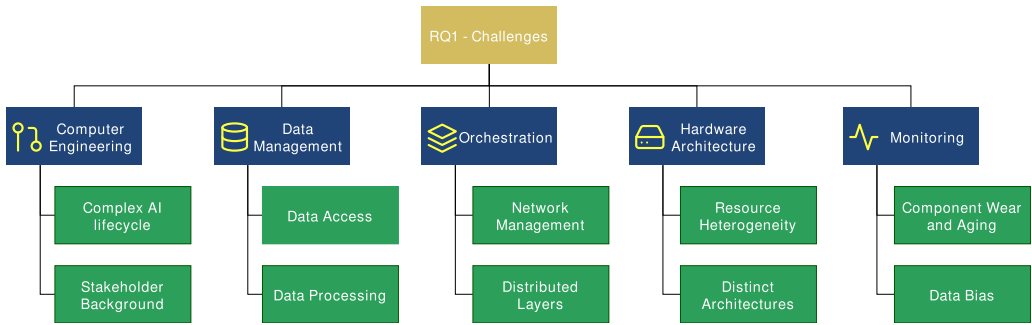


Fig. 6. A summary of the challenges associated with the adoption of MLOps and AIOps methodologies classified by category.

to monitor the deployment is problem-specific [49], and some authors propose a method for inferring application KPI degradation, without having to wade into application-specific metrics, for application engineers to rely on [45]. The recent upsurge of LLMs raises several concerns. In terms of fairness, AI applications at various stages of development may be biased by using training data in dubious ways. Mental health risks are associated with human participation in the MLOps cycle, which has implications for transparency, and the use of frameworks for developing reliable generative AI applications is crucial for explainability [51].

To top it all off, data science endeavors very often suffer from unrealistic expectations [77], and business, architectural, process, and organizational challenges are to be expected while adopting ML-based solutions in traditional industries [37], which in conjunction with the aforementioned challenges results in a degraded productivity of the stakeholders due to the interdisciplinary nature of the projects and management complexity [22]. The summary of the main areas of interest related to this research question is depicted in Figure 6.

Implementing MLOps and AIOps is cumbersome and requires a combination of cross-functional skills and a collaborative culture within the organization. To be successful, projects require a combination of software engineering, data science, and operations expertise that is difficult to find. The importance of proper data management is paramount in data science projects, which is only exacerbated in big data ecosystems. Tools and frameworks for efficient distribution, parallelization, and orchestration of data and tasks are necessary due to the different computing tiers (i.e., cloud, edge) in which these projects operate, but also present multiple challenges (e.g., scarce computing resources, security, privacy). In addition, these environments often consist of heterogeneous infrastructure devices with different characteristics and requirements. Next, monitoring the production environment promotes the responsiveness of such models to the various drifts. The scientific literature on this topic is in a state of ongoing development, and more extensive research will be required in the future to identify distinct challenges for both methodologies.

4.2 What are the Opportunities and Future Trends in MLOps? (RQ2)

This subsection provides an analysis and categorization of the various studies that address the potential applications and emerging trends of MLOps studies. Figure 7 depicts the primary areas on which academic studies are focused. The cloud continuum and edge computing, followed by harnessing the AI lifecycle, are the main opportunities identified by researchers.

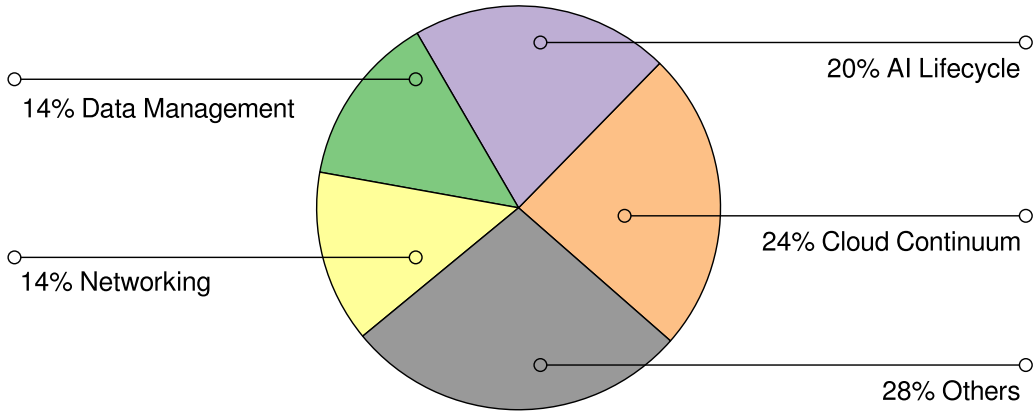


Fig. 7. The principal areas where opportunities and future trends for MLOps are trending.

4.2.1 Industry. There is a need for accurately measuring the business impact of AI solutions on businesses [37]. Data science projects require the involvement of different business layers to succeed, and leadership, executives, and stakeholders need to be on board with the MLOps strategy [21]. In this regard, education and training are required for ML activities [5, 21]. The use of ML algorithms has shown enormous potential for complex critical systems and processes [4, 75].

4.2.2 AI Lifecycle. In order to do this, the machine learning lifecycle needs to be revisited and requires more research to aid practitioners [49]. Some authors point to the continuous delivery of MLOps [64], and to the end-to-end automation of the various stages of the machine learning workflow [5]. There is a need to apply software engineering principles in the ML workflow [5], and the interaction between MLOps and existing practices within companies needs to be taken into account [64]. In addition, the opportunity to incorporate LLMs into the MLOps lifecycle has arisen because of their recent growth [51]. An appropriate data management remains a short term goal for organizations to handle [21]. In this regard, actions such as data availability [5], data standardization [21], data sharing [21], data integration [49], data collection [5], data cleaning [5], and data analysis [15] become of paramount importance.

4.2.3 Cloud Continuum. Data science projects promote the continuous evolution of hardware capabilities to provide the computational power and energy efficiency required [115]. For instance, the power consumption, memory, and real time constraints of microcontrollers require attention in order to deploy neural networks at the edge of the network [87]. On the other hand, FPGAs can be the cornerstone for designing the next-generation AI processors for consumer devices [115]. In fact, the use of edge and IoT devices in data science projects may lead to improvements in latency [71, 111], reliability [71], performance [4], safety [4], economy [4], privacy [71, 78, 111], energy consumption [78], and networking [78]. Yet another twist is tiny robot learning, which lies at the intersection of ML, embedded systems, and robotics, hence having to deal with their combined requirements [82]. In addition, the impact of reduced computational power can be mitigated by executing the most computationally intensive tasks on more resource-rich devices [40]. For instance, the training can be shared between both cloud and edge devices [44, 98]. In this regard, continuous learning yields the opportunity to periodically retrain the ML models based on the continuous stream of data [60], in order to raise the efficiency of the ML models [108]. In this regard, existing **High Performance Computing (HPC)** infrastructures represent one of the most cost-effective solutions [16], and containers promote the deployment of scalable code on different operating systems and hardware architectures [17, 89].

4.2.4 Networking. There is a rising interest in networking among the research community in MLOps [40, 78]. For instance, the use of lightweight networks as reliability estimators may be used to predict potential task failures [70]. Some authors identify **Function-as-a-Service (FaaS)** technologies as enablers of MLOps patterns [109], and the deployment of already trained ML models on FaaS may unleash the use of these technologies for event-driven AI solutions [23, 120]. The continuous monitoring of ML-based applications is required for reliable performance on critical systems [75]. The deployment and redeployment of intelligent algorithms onto heterogeneous hardware and software architectures can be alleviated by using a domain-specific language [35, 118]. Finally, traditional version control systems utilized in software engineering often lack the ability to distinguish between ML specific components such as models and datasets [56], and more research needs to be conducted in this field to better support the machine learning lifecycle [49].

Data science projects require the involvement of different business units, and their education and training are of paramount importance. More focus on the ML lifecycle is required, and software engineering principles such as continuous delivery need to be applied. Data management should be a pivotal point for organizations willing to benefit from AI. Being able to use hardware platforms such as FPGAs and IoT devices leads to improvements in areas such as networking and privacy. Technologies like containers and FaaS promote the deployment of scalable ML projects on different platforms and architectures. Finally, data science endeavors require additional versioning tools and frameworks for models and datasets than those of traditional software engineering.

4.3 What are the Opportunities and Future Trends in AIOps? (RQ3)

This section focuses on opportunities and emerging trends in AIOps, where the inclusion of cutting-edge solutions such as DL and Transformers can be advantageous due to the complexities of these scenarios [1]. The primary topics are the AI lifecycle and its various stages, networking-related opportunities, and the various goals pursued by AIOps endeavors.

4.3.1 AI Lifecycle. The seamless support of the data science lifecycle is critical in ML-based solutions [94]. In [61], the authors discuss **Deep Neural Network (DNN)** deployment on heterogeneous hardware platforms. Multi objective optimization algorithms can be leveraged in the deployment of analytical pipelines in heterogeneous environments [35]. ML techniques can boost the efficiency of the monitoring phase and liberate IT professionals to do more innovative tasks [99]. Some authors suggest the use of generative adversarial networks for the implementation of this stage [53], others propose the automated inference of application KPIs without application-specific knowledge [45]. In addition, AIOps solutions can help monitor concept drift and suggest suitable model retraining methods [96]. A significant amount of the existing literature is focused on data management, where the continuous access to data often found in big data systems facilitates the proactive retraining of models when necessary [62]. The quality and scale of data are crucial for the observability of the system [103]. In [61], the authors identify that AIOps can be of essential value for data lakehouses due to the complex operational challenges for SRE such as disaster recovery, backup, and restore. The interoperability between distinct components, including data collection, processing, and summarization, facilitates the use of predictive analytics over data streams [12]. The authors of [67] argue that data-driven anomaly detection is a crucial component of AIOps. In the field of root cause analysis, the use of logs to analyze certain events can become an integral part of IT operations [86]. A more ambitious endeavor is proposed in [24], where the authors address model training, packaging, and deployment. They employ AIOps to autonomously detect the state of the system, allocate resources, warn, and detect anomalies.

4.3.2 Networking. The complexities posed by networking technologies, such as the increasing number of devices and the growing number of services relying on connectivity, pave the way for the application of AI-based solutions to address them [81]. The myriad objectives of these network optimizations might range from minimizing energy consumption to reducing network delays [117]. Transfer learning is a promising technique for estimating the **Quality of Transmission (QoT)** of optical links [94]. Similarly, **Deep Reinforcement Learning (DRL)** optimization can raise the robustness of the network topology to tackle the inherent vulnerabilities of IoT to network failures and malicious attacks [95]. The authors in [100] use graph theory for optimizing network lifetime, which leverages the capabilities of edge devices and **Software Defined Network (SDN)** controllers. Similarly, deployment performance can be optimized by leveraging SDN and serverless architectures [123]. Promising results are obtained by applying genetic algorithms for minimizing network delays [36].

4.3.3 Objectives. The definition of certain objectives that the system will attempt to optimize is a common approach for AIOps solutions. In [18], the authors propose a system to take the burden of manual optimization of neural networks off the programmer. It outperforms existing solutions in the field in terms of performance, inference, and energy efficiency. Similarly, Hazra et al. address the difficulties posed by energy consumption and processing delays in edge environments [52]. Clustering techniques can reduce the communication load and enhance the energy efficiency of devices at the edge of the network [66]. In [36], the authors propose the use of multi objective optimization to leverage multiple converging objectives, including cost, performance, resilience, and network; subject to a set of predefined constraints. Alternatively, AIOps can serve as a manner to improve software quality and engineering productivity [30, 75].

AIOps solutions can boost the various stages that comprise the data science lifecycle, including data management, monitoring, packaging, and deployment. A significant amount of the literature is devoted to data management, whereas other phases, such as monitoring and retraining of ML models, receive less attention. In this regard, the inference of metrics utilizing AI simplifies the monitoring of ML models, and the continuous stream of data facilitates their retraining. A significant portion of the literature deals with the difficulties posed by networking technologies, such as connectivity and network delays. Finally, optimizing the performance, energy consumption, and cost of existing architectures are also popular topics in academia.

4.4 What Frameworks and Architectures Facilitate MLOps and AIOps? (RQ4)

This section elaborates on the various MLOps and AIOps frameworks discovered in scientific literature. A taxonomy is presented in Table 6, which serves to identify the differences between both methodologies with respect to their focus areas.

4.4.1 Data Management. On top of the inherent difficulties of implementing MLOps and AIOps in production environments, Big Data ecosystems need to overcome the particularities associated with the volume, variety, and velocity of the data. In this regard, some authors opt to simplify the complexity of big data workloads by characterizing them into subsets, which they later train independently [62]. Similarly, the complexity of integrating big data, machine learning, and IoT solutions has raised attention towards the training and inference orchestration of the underlying ML solutions [4] and lowered the burden for developers to work in such sophisticated environments [14]. An ML-based solution must also efficiently manage the dataset and model versions through dedicated tools and architectures [126]. The authors in [121], address the dearth of labeled datasets in ML-based solutions and propose a novel method for automating the labeling of

Table 6. A Taxonomy of the Various Frameworks and Architectures that Facilitate the Adoption of MLOps and AIOps Classified According to their Focus Areas

	MLOps		AIOps
	HPC	[16, 17]	–
Infrastructure Management	Edge	[3, 4, 8, 18, 33, 44, 80, 87, 98, 125]	[4, 18, 78, 98, 123]
	Cloud	Generic	[36, 44, 54, 62, 80]
		Serverless	[8, 22]
		as a Service	[14, 43, 59]
	Networking	[16, 44, 59, 68]	[29, 123]
Data Management	DataOps	[126]	[76, 121]
	Big Data	[4, 14]	[4]
Lifecycle Management	Generic	[4, 31, 32, 54, 129]	[4, 54]
	Monitoring	[53]	[67, 76, 96]
	Deployment	[3, 16–18, 36, 68, 80]	[18, 36]
	Training	[98]	[98]

log messages without manual labor or expert intervention. Similarly, data-driven proactive incident triage using data is utilized in [76].

4.4.2 Infrastructure Management. **High Performance Computing (HPC)** represents an interesting alternative for dealing with such complex workloads, but it is challenging for data scientists due to the unique requirements and characteristics of these environments [17]. The scalability of computational resources is key for solving these large workloads. Due to this, the use of containerization solutions has yielded promising results for deploying AI on HPC systems [16]. An approach for leveraging this paradigm is the use of cloud computing resources, for which some authors highlight the importance of guaranteeing multitenancy isolation [88]. In addition, deep learning components can benefit from the deployment of training tasks over Cloud and Edge infrastructural devices [44]. Similarly, the problem of cloud resource allocation can be tackled with a hybrid multi-objective genetic algorithm [48]. In order to simplify this process, the serverless computing paradigm is an attempt to simplify the use of such resources, in which the management of the infrastructural devices is accomplished by the provider. In addition, it aspires to reduce the cost while ensuring scalable resources and inference latency [11]. To this end, a framework for the end-to-end management of the necessary resources required by ML workflows is of paramount importance [22]. The combination of serverless functions deployed on the edge of the network is gaining traction, and some authors are already implementing platforms [123] and architectures [8] to benefit from the latency improvements in comparison with the deployment of machine learning workflows on the cloud. In addition, there exist frameworks [18] specifically tailored for being able to harness the benefits of deploying artificial intelligence workloads on low consumption and limited memory edge devices. In this regard, the Internet of Things is a challenging ecosystem for developers, who often lack the expertise to work in these complex environments [14]. Due to this, there exists research on dedicated architectures [3, 33], frameworks [4, 14], and platforms [123] designed to alleviate this burden and aid in harnessing its full potential. On the other hand, well known frameworks such as TensorFlow Lite [110] and Core ML [58] are specifically tailored to utilize ML capabilities on mobile and edge devices [47]. On a similar note, yet another approach is the use of genetic algorithms for finding the configuration that minimizes model size while maximizing accuracy on edge deployments [111]. In [125], the authors propose an architecture, coined TMLaaS for the execution of ML models on low-power IoT devices.

4.4.3 Lifecycle Management. The distributed training and inference of machine learning workflows originated by the emergence of these new paradigms is gaining traction in the research community [8, 31, 59]. This strengthens the necessity of managing the lifecycle and key components of AI applications through dedicated frameworks [54] and attracts research attention

towards handling its various stages. In this regard, the authors in [32] propose an AI lifecycle approach for tackling the challenges of AI-based solutions from conception to production. Firstly, a semantically enhanced pipeline can automate data preparation [129]. Secondly, the training of ML workloads can be characterized into smaller pieces for distributed execution [62], and frameworks for retraining ML models to dynamically adjust to varying energy and memory constraints can speed up inference [98]. Thirdly, many authors propose frameworks and architectures that tackle the deployment stage of data science projects in different computational environments, such as the cloud [44, 80, 88] and the edge [8, 44, 88], including devices such as microcontrollers [87]. The authors of [36] propose a goal-driven framework for the operationalization of distributed analytical pipelines across the cloud continuum. In addition, the deployment of certain flavors of machine learning algorithms, such as deep learning [78] and deep neural networks [18], are also addressed. Yet another angle is to address the deployment of such workloads by maximizing resource utilization and operator revenue [29]. Finally, some studies focus on the monitoring stage by proposing architectures [3] and frameworks [53, 54] that supervise processes and events in order to react to their deviations during runtime. The authors of [67] propose an AIOps anomaly detection framework that addresses functional and performance failures in software systems. Similarly, [96] suggests an automated pipeline for AIOps monitoring and maintenance. In summary, frameworks that are able to orchestrate ML workflows [4] and their associated resources [29] become essential to navigate through the complexity of this endeavor.

4.4.4 Technologies. From a technological standpoint, techniques like software-defined networks simplify network management and enable the implementation of unified services for optimal AI deployment, optimizing the available resources [123]. In [68], an architecture for the deployment of AI solutions in B5G networks is proposed. ML automation processes such as AutoML are pivotal for the maturity and efficiency of the ML models in production environments, and more research is required [108]. Containerized solutions represent a unique opportunity for the operationalization of data science workloads since they promote the ubiquitous deployment of pipeline stages. Due to this, some architectures [33] and frameworks [4, 16] found in academia devote to this technology for the deployment and orchestration of AI workloads. Yet another technology utilized for the operationalization of predictive models is the use of API resources, as it provides a simplified abstraction on top of sophisticated prediction models, promotes resource sharing, and simplifies the consumption of the resources for the end user [43]. On the other hand, the operationalization of some flavors of algorithms captivates the attention of researchers, and the operationalization of both deep neural networks [18, 87, 107] and deep learning [17, 31, 44, 78] solutions is frequently addressed. A frequent problem is addressing the particularities of the infrastructural devices on which they are deployed.

On top of the inherent particularities of MLOps and AIOps, data science projects frequently need to overcome the challenges of Big Data ecosystems. These platforms are often operated leveraging HPC technologies, which require the involvement of varying backgrounds of stakeholders. In this regard, containerized solutions and serverless technologies alleviate the deployment of ML-based solutions and their dependencies on different platforms and computational layers, such as the cloud and the edge. However, it also raises the manifold challenges of orchestrating distributed pipelines; hence, tools that can handle the ML lifecycle gain relevance. In this regard, the deployment of AI workloads receives the majority of the attention, while other phases like monitoring or training remain understudied. AIOps solutions are still underrepresented in the fields of infrastructure management and in the deployment of ML solutions.

Table 7. A Taxonomy of the Fields where MLOps and AIOps Methodologies are Thriving

		MLOps	AIOps
Research	Space	[40]	[40]
	Physics	[17, 46]	-
Industry	Factories	[3, 118]	-
	Construction	[10, 37]	[37]
	Autonomous Vehicles	[44]	-
	Health	[15, 51]	[15]
	Failure Management	-	[84]
IT	5G	[68, 94]	[29, 66, 94, 123]
	6G	[59]	-
	Networking	[15, 116]	[15, 90, 116, 117]
	Service Management	-	[1]
	Logs	-	[76, 86, 121]

4.5 What are the Current and Future Fields in which MLOps and AIOps are Thriving? (RQ5)

This subsection provides an analysis of the current and future fields of MLOps and AIOps methodologies. Table 7 presents a taxonomy of the various fields in which MLOps and AIOps methodologies are being utilized, and highlights the areas in which each paradigm thrive.

4.5.1 Industry and Research. The widespread adoption of artificial intelligence solutions for increased competitiveness has reached traditional corporations. In this regard, the authors in [84] identify proactive and reactive failure management (prevention, prediction, detection, root cause analysis, remediation), and resource provisioning (consolidation, scheduling, power management, service composition, workload estimation) as the main areas in which AIOps are thriving. Some authors describe the deployment of predictive maintenance systems in stamping machines to minimize the effects and impact of unexpected failures [3]. Similarly, the redeployment of intelligent algorithms in cyber-physical production systems in Industry 4.0 remains a challenge due to the differences in reaction times, communications, and computation power in the infrastructural devices; positive feedback has been reported by experts using a domain-specific language for modeling these industrial use cases [118]. The building and construction industries have also adopted AI solutions [10], but their applications remain a challenge for large-scale projects [37]. On the other hand, innovative industries also require expertise in MLOps and AIOps should they want to incorporate the benefits of artificial intelligence into their solutions. In the wind power industry, Wireless Sensor Networks are pivotal for the monitoring of power generation systems, but the harsh environmental conditions in which wind farms are often located make their optimal deployment troublesome [117]. In the automotive sector, the elastic deployment of training tasks over cloud and edge resources, leveraging stringent network and privacy requirements, facilitates the improvement of autonomous driving applications [44]. In space exploration, AI solutions are already applied for enhanced monitoring and diagnostics, prediction, and image analysis, but bringing AI on board remains a challenge due to the scarce computational and network resources available [40]. Next, recent advances in mobile technologies enable the development and deployment of ML-based patient monitoring right on mobile devices within the healthcare industry, but the associated challenges have not been extensively studied by the research community, and a set of recommendations is required [15]. In addition, LLMs can also be applied to healthcare by transforming data management workflows [51]. Finally, fields more traditionally associated with research are also leveraging the MLOps. In [46], facilitating the implementation of DL solutions in

gravitational wave physics is discussed. On a similar note, high-energy physics requires the analysis of massive amounts of data using ML technologies and resorts to high-performance computing technologies to cope with the data storage, data transfer, and computation requirements [17].

4.5.2 Information Technology. One of the fields in which both methodologies are more popular is communications and networking. In particular, 5G technologies are where most of the academic attention is focused. Some authors are applying ML technologies for the rapid deployment of Quality of Transmission predictors in complex 5G network operation scenarios [94]. Next, **multi-access edge computing (MEC)** is a promising technology aiming to improve **Quality of user Experience (QoE)** of AI applications in IoT infrastructure. The authors in [123] deploy services in MEC 5G edge infrastructural devices to save costs and meet QoE requirements in massive edge data centers. Future trends for IoT involve shifting from a static to a dynamically evolving and self-organized architecture, which fits with the capabilities of 5G networks to continuously adapt and reorganize based on changing requirements. The deployment of IoT applications on these 5G edge architectures for task offloading is discussed in [66]. Supporting the specific requirements and priorities of 5G networks is cumbersome, and AI technologies can support the 5G slice deployment and orchestration for enhanced resource utilization and reduced slice request dropping probabilities [29]. In [68], the authors pursue the integration of ML techniques to optimize 5G systems. The next generation of wireless communication technologies, coined 6G, promotes the ubiquity of AI services, and specialized frameworks for distributed AI provisioning are required [59]. Next, the appropriateness of ML technologies is discussed as a technique to perform traffic analysis and classification to identify the correct procedures and achieve the desired outcomes [90]. In this regard, AIOps is leveraged by the authors in [76, 86] for the log analysis in incident remediation, whereas [121] focuses on automated log labeling. In [1], the authors evaluate the application of AIOps technologies in **IT Service Management (ITSM)**, highlighting the potential for predicting and resolving IT incidents in the shortest time. Finally, the many challenges associated with the heterogeneity of mobile devices have raised interest in counteracting the problems associated with the inference of DL apps [116].

Traditional corporations such as the building and construction industry and the automotive sector have already adopted AI-based solutions, but more effort is required for their implementation in larger endeavors. Similarly, more innovative sectors, such as the wind power industry or even space exploration, are investing in AI solutions to address the challenges associated with the harsh environmental conditions in which they operate. Academic disciplines, such as physics, are devoted to MLOps for coping with stringent data requirements. However, the area in which MLOps and AIOps are more prevalent is communications and networking. The rise of 5G and 6G architectures has yielded some challenges and opportunities, such as slice deployment and traffic analysis, for which AI-based solutions are ideal. It is anticipated that much of the current 5G effort will gradually shift toward 6G technologies. Service management and log analysis are highly attractive areas for AIOps solutions, whereas physics, autonomous vehicles, and traditional factories are more akin to MLOps.

5 RELATED WORK

We have compiled all pertinent studies and reviews in the fields of MLOps and AIOps to provide the rationale for this work and situate it within the body of existing research. There is one study offering a general perspective on AIOps, three studies on specific fields of AIOps, and one on MLOps. To the best of our knowledge, there has not been a study that conveys both MLOps and AIOps and provides as detailed a broad view of both fields as this one.

Table 8. Related Works Conducted in Areas Covered by this Research

Paper	Type	Time frames	Topic	Focus
This study	SLR ¹	2018 – 2023	MLOps, AIOps	Challenges, Architectures, Future Fields
Rijal et al. [99]	MLR ²	2017 – 2021	AIOps	Benefits, challenges, and future.
Notaro et al. [85]	SMS ³	1990 – 2020	AIOps	Failure Management
Zhaoxue et al. [128]	Survey	up to 2020 ⁴	AIOps	Logs
Lima et al. [73]	SLR ¹	up to 2021	MLOps	Maturity Models, tools and challenges.
Ahmed et al. [1]	SLR ¹	2000 – 2022	AIOps	Risk Prediction
Kreuzberger et al. [69]	MM ⁵	up to 2021	MLOps	Definition and Architecture
Cheng et al. [25]	Survey	up to 2023	AIOps	Opportunities and challenges

¹Systematic Literature Review.²Multivocal Literature Review.³Systematic Mapping Study.⁴Time frame

not explicitly stated.

⁵Mixed Method.

In [99], the authors perform a **multivocal literature review (MLR)** in which they identify that the adoption of AIOps helps in monitoring IT work and improves human-AI collaboration. There are, however, concerns about the effectiveness of AI and ML and the quality of the quality of the data utilized to obtain the results. Notaro et al. [85] provide five different categories and fourteen subcategories for the categorization of failure management IT solutions using AIOps. In [129], the authors review and categorize existing works around three key processes in log processing, such as log enhancement, log parsing, and log analysis. They finalize future directions and development trends in the field of log research. Lima et al. [73] provide a systematic literature review on practices, maturity models, roles, tools, and challenges for MLOps and establish that this methodology is still in its infancy, leaving room for future academic studies that will guide organizations. A systematic literature review of existing work and challenges in the field of IT risk prediction is offered in [1], which reveals ML classifiers as the preferred method for implementing IT service management and highlights the importance of adopting more advanced state-of-the-art methods, such as DL and Transformers. In [69] the authors focus on the definition and architecture of MLOps and perform a literature review (up to May 2021), a tool review, and interviews with experts in the field. They conclude that model development and benchmarking have received more attention in the academic community than operating ML systems, which is still challenging today. The authors of [25] conducted a survey regarding the opportunities and challenges of AIOps. They suggest that, with the growth of IT infrastructure, AIOps is the only promising solution to cope with it. Numerous techniques, such as anomaly detection, root-cause analysis, failure predictions, automated actions, and resource management, are utilized in AIOps. They have found that many of the solutions focus on detection and root cause analysis, but automation is still limited.

Table 8 provides a comparison framework for this SLR with each of the studies described above. For the purpose of evaluating our work, we have established four different criteria. Firstly, we have followed the PRISMA methodology for systematic literature reviews to transparently report the reasoning for doing the review, what has been accomplished, and the discoveries of the review. Secondly, this manuscript is the most contemporary, as it contains studies up to 2023. Thirdly, it is the only one that offers a joint perspective on both MLOps and AIOps methodologies, as we think they have a strong influence on one another. Finally, the focus of this study is the widest, as it comprises not only a general overview of the challenges and benefits of these methodologies but also a deep dive into their current and future use in both industry and academia.

6 CONCLUSIONS AND FUTURE WORK

The overarching goal of this SLR is to offer insights into the adoption of MLOps and AIOps methodologies in both industry and academia. We have conducted an in-depth search of the scientific literature based on the search terms described in Figure 1, and we elaborate the discussion on the research questions outlined in Table 2. The necessity for this SLR stems from the fact that the implementation of data science projects in production environments is a challenging endeavor in which MLOps and AIOps can play a facilitating role. However, organizations and stakeholders need to have both a collaborative culture and a unique cross-domain skillset of software engineering, data science, and IT operations. Data management plays a significant role in these projects and must be planned accordingly. In this regard, the complexity is amplified by big data ecosystems and their unique requirements and characteristics. In addition, recent computing paradigms such as cloud and edge computing require innovative solutions due to the distributed and heterogeneous nature of the infrastructural devices they comprise. To circumvent these challenges, more focus is required on applying software engineering principles to the ML lifecycle. In this regard, technologies such as containerization, data, and model versioning, FaaS, and serverless are the cornerstones for supporting this lifecycle. The monitoring stage, which allows the retraining and redeployment of the relevant components of the architecture subject to the various drifts of the production environments, can benefit from AIOps solutions for relevant KPI inference. On the other hand, AIOps solutions are often focused on solving the myriad networking requirements, and hardware configurations of modern ecosystems. Due to this, the use of data science orchestration frameworks can be beneficial since they are specifically tailored to address various stages of the ML lifecycle (e.g., re-training, re-deployment, monitoring, versioning). The reviewed manuscripts showcase that AI-based solutions are no longer restricted to academia but have reached not only innovative industries such as space exploration and the wind power industry, but also traditional corporations such as the construction and automotive sectors. Finally, the rise of 5G and 6G technologies and architectures leveraging MLOps and AIOps methodologies pose an opportunity for AI-based solutions. As for future work, since AIOps is such a new trend, this manuscript provides a shared perspective on the challenges associated with the adoption of both MLOps and AIOps. We reckon sufficient AIOps-related manuscripts will emerge in the near future to be able to distinguish the challenges from those of MLOps. In addition, the recent upsurge in LLMs will spread to MLOps and AIOps, and further analysis on this subject will be necessary.

A APPENDIX

Table 9. Article Assessment based on the Quality Metrics

Paper	M1	M2	M3	M4	M5	M6	M7	Total
[1]	1	0	1.2	1	0	0	0	3.2
[2]	1	0	0.4	0	0	0	2	3.4
[4]	1	0	1.2	1	1	0	1.3	5.5
[3]	0	0	0.8	1	0	0	0.7	2.5
[5]	0	1	0.8	0	0	0	2	3.8
[8]	1	1	0.8	0	0	0	0	2.8
[10]	1	0	0.4	1	0	0	0.7	3.1
[11]	1	0	0.4	1	0	0	0	2.4
[12]	1	0	0.4	0	0	0	0	1.4

(Continued)

Table 9. Continued

Paper	M1	M2	M3	M4	M5	M6	M7	Total
[14]	1	0	0.8	0	0	0	0.7	2.5
[15]	1	0	1.2	1	1	0	1.3	5.5
[17]	0	0	1.6	1	0	0	2	4.6
[16]	0	1	1.2	0	0	0	0.7	2.9
[18]	1	1	1.2	1	1	0	2	7.2
[20]	0	0	0.4	1	0	0	0	1.4
[21]	0	0	0.8	1	1	0	1.3	4.1
[22]	1	0	0.8	0	0	0	2	3.8
[23]	0	0	0.4	0	0	0	1.3	1.7
[24]	0	0	0.4	0	1	0	0	1.4
[29]	0	1	0.8	1	1	1	1.3	6.1
[30]	0	0	0.8	0	0	0	2	2.8
[31]	0	0	0.4	1	1	0	2	4.4
[32]	0	0	0.4	1	1	0	0.7	3.1
[33]	1	0	0.4	1	0	0	2	4.4
[35]	1	1	0.8	0	0	0	1.3	4.1
[36]	1	1	0.8	0	1	1	0.7	5.5
[37]	1	0	1.2	0	0	0	0	2.2
[39]	0	0	0.4	1	1	0	0.7	3.1
[40]	0	0	1.2	0	0	0	0.7	1.9
[42]	1	0	0.4	0	0	0	0	1.4
[43]	0	0	0.8	1	1	0	2	4.8
[44]	0	1	1.2	1	1	0	0.7	4.9
[45]	0	0	0.8	0	0	0	2	2.8
[46]	0	0	0.4	1	0	0	0	1.4
[47]	1	0	0.8	1	0	0	2	4.8
[48]	1	0	0.4	0	0	0	0	1.4
[49]	1	0	0.8	1	1	1	0.7	5.5
[51]	0	0	1.2	1	0	0	0	2.2
[52]	0	0	0.8	0	1	1	0.7	3.5
[53]	1	0	1.2	0	0	0	0	2.2
[54]	1	1	0.8	0	0	0	2	4.8
[56]	1	0	0.8	1	0	0	0.7	3.5
[59]	1	1	0.8	1	1	0	2	6.8
[60]	0	0	0.8	0	1	1	0	2.8
[61]	0	0	0.4	0	0	0	0	0.4
[62]	0	1	1.2	0	0	0	0.7	2.9
[64]	1	1	0.8	1	1	0	0	4.8
[66]	1	0	1.2	0	0	0	2	4.2
[67]	0	1	0.8	1	1	0	0	3.8
[68]	0	0	0.8	1	1	0	0	2.8
[70]	0	0	0.8	0	0	0	0.7	1.5
[71]	0	0	0.8	1	0	0	0	1.8
[74]	1	0	0.4	0	0	0	0.7	2.1

(Continued)

Table 9. Continued

Paper	M1	M2	M3	M4	M5	M6	M7	Total
[75]	1	0	1.2	0	1	0	2	5.2
[76]	0	1	1.2	0	0	0	1.3	3.5
[77]	1	0	0.4	1	0	0	0.7	3.1
[78]	0	0	1.2	1	0	0	1.3	3.5
[80]	1	0	0.4	1	1	0	0	3.4
[81]	1	0	0.8	0	1	1	1.3	5.1
[82]	0	0	0.4	1	0	0	0	1.4
[84]	1	0	1.2	1	0	0	0.7	3.9
[86]	0	0	0.8	0	1	0	0	1.8
[87]	1	0	1.2	1	1	0	1.3	5.5
[88]	0	0	0.8	1	1	0	1.3	4.1
[89]	1	0	0.8	1	0	0	0.7	3.5
[90]	1	0	0.8	1	1	1	2	6.8
[93]	0	0	0.4	1	0	0	0	1.4
[94]	0	0	1.2	1	1	0	0.7	3.9
[95]	0	1	0.4	0	1	1	0	3.4
[96]	0	0	1.2	0	0	0	0	1.2
[97]	1	0	0.4	0	1	0	2	4.4
[98]	1	0	0.4	0	0	0	0	1.4
[99]	1	1	0.8	0	0	0	0	2.8
[100]	1	0	0.4	1	1	0	0	3.4
[101]	1	0	0.4	1	1	1	0	0.4
[102]	0	0	0.4	1	0	0	0	1.4
[103]	1	0	0.4	0	0	0	0	1.4
[104]	0	0	0.4	0	0	0	0	0.4
[107]	0	0	0.8	0	0	0	0.7	1.5
[108]	0	0	0.8	1	0	0	0	1.8
[109]	1	0	0.8	0	0	0	1.3	3.1
[111]	0	1	1.2	1	1	0	0	4.2
[116]	0	0	1.2	1	0	0	0.7	2.9
[115]	0	0	0.8	1	0	0	0.7	2.5
[117]	1	0	1.2	0	1	0	1.3	4.5
[118]	1	0	1.2	1	1	1	1.3	6.5
[120]	1	0	0.8	0	0	0	0.7	2.5
[121]	0	0	1.2	1	0	0	0.7	2.9
[123]	1	0	1.6	0	1	1	0.7	5.3
[124]	0	0	0.4	0	0	0	2	2.4
[125]	0	0	0.8	1	1	0	0	2.8
[126]	1	0	0.4	0	0	0	0	1.4
[129]	1	1	0.8	0	0	0	2	4.8

REFERENCES

- [1] Salman Ahmed, Muskaan Singh, Brendan Doherty, Effirul Ramlan, Kathryn Harkin, and Damien Coyle. 2022. AI for information technology operation (AIOps): A review of IT incident risk prediction. In *2022 9th International Conference on Soft Computing & Machine Intelligence (ISCM)*. IEEE, 253–257.

- [2] Moayad Alshangiti, Hitesh Sapkota, Pradeep K. Murukannaiah, Xumin Liu, and Qi Yu. 2019. Why is developing machine learning applications challenging? A study on stack overflow posts. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 1–11.
- [3] Filipe Alves, Hasmik Badikyan, H. J. Antonio Moreira, João Azevedo, Pedro Miguel Moreira, Luís Romero, and Paulo Leitão. 2020. Deployment of a smart and predictive maintenance system in an industrial case study. In *2020 IEEE 29th International Symposium on Industrial Electronics (ISIE)*. IEEE, 493–498.
- [4] José M. Alves, Leonardo M. Honório, and Miriam A. M. Capretz. 2019. ML4IoT: A framework to orchestrate machine learning workflows on internet of things data. *IEEE Access* 7 (2019), 152953–152967.
- [5] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. 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, 291–300.
- [6] Edoardo Aromataris and Alan Pearson. 2014. The systematic review: An overview. *AJN The American Journal of Nursing* 114, 3 (2014), 53–58.
- [7] Filipe Assunção, Nuno Lourenço, Bernardete Ribeiro, and Penousal Machado. 2020. Evolution of scikit-learn pipelines with dynamic structured grammatical evolution. In *International Conference on the Applications of Evolutionary Computation (Part of EvoStar)*. Springer, 530–545.
- [8] Ta Phuong Bac, Minh Ngoc Tran, and YoungHan Kim. 2022. Serverless computing approach for deploying machine learning applications in edge layer. In *2022 International Conference on Information Networking (ICOIN)*. IEEE, 396–401.
- [9] Aritz Badiola-Bengoa and Amaia Mendez-Zorrilla. 2021. A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise. *Sensors* 21, 18 (2021), 5996.
- [10] Shanaka Kristombu Baduge, Sadeep Thilakarathna, Jude Shalitha Perera, Mehrdad Arashpour, Pejman Sharafi, Bertrand Teodosio, Amkit Shringi, and Priyan Mendis. 2022. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction* 141 (2022), 104440.
- [11] Amine Barrak, Fabio Petrillo, and Fehmi Jaafar. 2022. Serverless on machine learning: A systematic mapping study. *IEEE Access* (2022).
- [12] Mariam Barry, Saad El Jaouhari, Albert Bifet, Jacob Montiel, Eric Guerizec, and Raja Chiky. 2022. StreamFlow: A system for summarizing and learning over industrial big data streams. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2198–2205.
- [13] Hamsa Bastani, Dennis J. Zhang, and Heng Zhang. 2022. Applied machine learning in operations management. In *Innovative Technology at the Interface of Finance and Operations*. Springer, 189–222.
- [14] Anirban Bhattacharjee, Yogesh Barve, Shweta Khare, Shunxing Bao, Zhuangwei Kang, Aniruddha Gokhale, and Thomas Damiano. 2019. STRATUM: A bigdata-as-a-service for lifecycle management of IoT analytics applications. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 1607–1612.
- [15] Omar Boursalie, Reza Samavi, and Thomas E. Doyle. 2018. Machine learning and mobile health monitoring platforms: A case study on research and implementation challenges. *Journal of Healthcare Informatics Research* 2, 1 (2018), 179–203.
- [16] David Brayford and Sofia Vallecorsa. 2020. Deploying scientific AI networks at petaflop scale on secure large scale HPC production systems with containers. In *Proceedings of the Platform for Advanced Scientific Computing Conference*. 1–8.
- [17] David Brayford, Sofia Vallecorsa, Atanas Atanasov, Fabio Baruffa, and Walter Riviera. 2019. Deploying AI frameworks on secure HPC systems with containers. In *2019 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 1–6.
- [18] Alessio Burrello, Angelo Garofalo, Nazareno Bruschi, Giuseppe Tagliavini, Davide Rossi, and Francesco Conti. 2021. DORY: Automatic end-to-end deployment of real-world DNNs on low-cost IoT MCUs. *IEEE Trans. Comput.* 70, 8 (2021), 1253–1268.
- [19] Elsevier B.V. 2022. Elsevier Developer Portal. (2022). <https://dev.elsevier.com/>. Last accessed 19 June 2022.
- [20] Fabio Calefato, Filippo Lanubile, and Luigi Quaranta. 2022. A preliminary investigation of MLOps practices in GitHub. In *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. 283–288.
- [21] Averill Champion, Mila-Gasco Hernandez, Slava Mikhaylov Jankin, and Marc Esteve. 2020. Managing artificial intelligence deployment in the public sector. *Computer* 53, 10 (2020), 28–37.
- [22] Joao Carreira, Pedro Fonseca, Alexey Tumanov, Andrew Zhang, and Randy Katz. 2019. Cirrus: A serverless framework for end-to-end ML workflows. In *Proceedings of the ACM Symposium on Cloud Computing*. 13–24.

- [23] Dheeraj Chahal, Ravi Ojha, Manju Ramesh, and Rekha Singhal. 2020. Migrating large deep learning models to server-less architecture. In *2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*. IEEE, 111–116.
- [24] Ruibo Chen, Yanjun Pu, Bowen Shi, and Wenjun Wu. 2023. An automatic model management system and its implementation for AIOps on microservice platforms. *The Journal of Supercomputing* 79, 10 (2023), 11410–11426.
- [25] Qian Cheng, Doyen Sahoo, Amrita Saha, Wenzhuo Yang, Chenghao Liu, Gerald Woo, Manpreet Singh, Silvio Saverese, and Steven C. H. Hoi. 2023. AI for IT operations (AIOps) on cloud platforms: Reviews, opportunities and challenges. *arXiv preprint arXiv:2304.04661* (2023).
- [26] Venkatarami Reddy Chintapalli, Koteswararao Kondepudi, Andrea Sgambelluri, Bheemarjuna Reddy Tamma, Piero Castoldi, and Luca Valcarenghi. 2020. Orchestrating edge-and cloud-based predictive analytics services. In *2020 European Conference on Networks and Communications (EuCNC)*. IEEE, 214–218.
- [27] Forrester commissioned by Google. 2023. Modernize with AIOps to Maximize Your Impact. (2023). https://services.google.com/fh/files/misc/google_active_assist_forrester_opp_snap_study_final.pdf. Last accessed 29 August 2023.
- [28] Crossref. 2022. Crossref API Documentation. (2022). <https://www.crossref.org/documentation/retrieve-metadata/rest-api/>. Last accessed 19 June 2022.
- [29] Ghina Dandachi, Antonio De Domenico, Dinh Thai Hoang, and Dusit Niyato. 2019. An artificial intelligence framework for slice deployment and orchestration in 5G networks. *IEEE Transactions on Cognitive Communications and Networking* 6, 2 (2019), 858–871.
- [30] Yingnong Dang, Qingwei Lin, and Peng Huang. 2019. AIOps: Real-world challenges and research innovations. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)*. IEEE, 4–5.
- [31] Elias De Coninck, Steven Bohez, Sam Leroux, Tim Verbelen, Bert Vankeirsbilck, Pieter Simoens, and Bart Dhoedt. 2018. DIANNE: A modular framework for designing, training and deploying deep neural networks on heterogeneous distributed infrastructure. *Journal of Systems and Software* 141 (2018), 52–65.
- [32] Daswin De Silva and Daminda Alahakoon. 2022. An artificial intelligence life cycle: From conception to production. *Patterns* (2022), 100489.
- [33] Olivier Debauche, Saïd Mahmoudi, Sidi Ahmed Mahmoudi, Pierre Manneback, and Frédéric Lebeau. 2020. A new edge architecture for AO-IoT services deployment. *Procedia Computer Science* 175 (2020), 10–19.
- [34] Josu Diaz-de Arcaya, Raul Miñón, Ana Isabel Torre-Bastida, Aitor Almeida, and Gorka Zarate. 2022. Official SLR repository. (2022). <https://github.com/josu-arcaya/slr>. Last accessed 15 December 2022.
- [35] Josu Díaz-de Arcaya, Raúl Miñón, Ana I. Torre-Bastida, Javier Del Ser, and Aitor Almeida. 2020. PADL: A language for the operationalization of distributed analytical pipelines over edge/fog computing environments. In *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, 1–6.
- [36] Josu Díaz-de Arcaya, Ana I. Torre-Bastida, Raúl Miñón, and Aitor Almeida. 2023. Orfeon: An AIOps framework for the goal-driven operationalization of distributed analytical pipelines. *Future Generation Computer Systems* 140 (2023), 18–35.
- [37] Rimma Dzhusupova, Jan Bosch, and Helena Holmström Olsson. 2022. Challenges in developing and deploying AI in the engineering, procurement and construction industry. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 1070–1075.
- [38] Christof Ebert, Gorka Gallardo, Josune Hernantes, and Nicolas Serrano. 2016. DevOps. *IEEE Software* 33, 3 (2016), 94–100.
- [39] Georgios Flamis, Stavros Kalapothas, and Paris Kitsos. 2021. Best practices for the deployment of edge inference: The conclusions to start designing. *Electronics* 10, 16 (2021), 1912.
- [40] Gianluca Furano, Antonis Tavoularis, and Marco Rovatti. 2020. AI in space: Applications examples and challenges. In *2020 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*. IEEE, 1–6.
- [41] Vahid Garousi, Michael Felderer, and Mika V. Mäntylä. 2016. The need for multivocal literature reviews in software engineering: Complementing systematic literature reviews with grey literature. In *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*. 1–6.
- [42] Nikhil Krishna Gopalakrishna, Dharun Anandayavaraj, Annan Detti, Forrest Lee Bland, Sazzadur Rahaman, and James C. Davis. 2022. "If security is required": Engineering and Security Practices for Machine Learning-based IoT Devices. (2022).
- [43] Eric Gossett, Cormac Toher, Corey Oses, Olexandr Isayev, Fleur Legrain, Frisco Rose, Eva Zurek, Jesús Carrete, Natalio Mingo, Alexander Tropsha, and Stefano Curtarolo. 2018. AFLOW-ML: A RESTful API for machine-learning predictions of materials properties. *Computational Materials Science* 152 (2018), 134–145.
- [44] Sorin Grigorescu, Tiberiu Cocias, Bogdan Trasnea, Andrea Margheri, Federico Lombardi, and Leonardo Aniello. 2020. Cloud2edge elastic AI framework for prototyping and deployment of AI inference engines in autonomous vehicles. *Sensors* 20, 19 (2020), 5450.

- [45] Johannes Grohmann, Patrick K. Nicholson, Jesus Omana Iglesias, Samuel Kounev, and Diego Lugones. 2019. Monitorless: Predicting performance degradation in cloud applications with machine learning. In *Proceedings of the 20th International Middleware Conference*. 149–162.
- [46] Alec Gunny, Dylan Rankin, Philip Harris, Erik Katsavounidis, Ethan Marx, Muhammed Saleem, Michael Coughlin, and William Benoit. 2022. A software ecosystem for deploying deep learning in gravitational wave physics. In *Proceedings of the 12th Workshop on AI and Scientific Computing at Scale using Flexible Computing Infrastructures*. 9–17.
- [47] Qianyu Guo, Sen Chen, Xiaofei Xie, Lei Ma, Qiang Hu, Hongtao Liu, Yang Liu, Jianjun Zhao, and Xiaohong Li. 2019. An empirical study towards characterizing deep learning development and deployment across different frameworks and platforms. In *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 810–822.
- [48] Gifty Gupta and Neeraj Mangla. 2022. Trust aware multi-objective metaheuristics for workflow scheduling in cloud computing. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, Vol. 1. IEEE, 602–609.
- [49] Mark Haakman, Luis Cruz, Hennie Huijgens, and Arie van Deursen. 2021. AI lifecycle models need to be revised. *Empirical Software Engineering* 26, 5 (2021), 1–29.
- [50] Neal R. Haddaway, Matthew J. Page, Chris C. Pritchard, and Luke A. McGuinness. 2022. PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews* 18, 2 (2022), e1230.
- [51] Stefan Harrer. 2023. Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine* 90 (2023).
- [52] Abhishek Hazra, Mainak Adhikari, Tarachand Amgoth, and Satish Narayana Srirama. 2021. Intelligent service deployment policy for next-generation industrial edge networks. *IEEE Transactions on Network Science and Engineering* (2021).
- [53] Maximilian Hoffmann, Lukas Malburg, and Ralph Bergmann. 2021. ProGAN: Toward a framework for process monitoring and flexibility by change via generative adversarial networks. In *International Conference on Business Process Management*. Springer, 43–55.
- [54] Waldemar Hummer, Vinod Muthusamy, Thomas Rausch, Parijat Dube, Kaoutar El Maghraoui, Anupama Murthi, and Punleuk Oum. 2019. ModelOps: Cloud-based lifecycle management for reliable and trusted AI. In *2019 IEEE International Conference on Cloud Engineering (IC2E)*. IEEE, 113–120.
- [55] IBM. 2023. AIOps Solutions. (2023). <https://www.ibm.com/aiops>. Last accessed 29 August 2023.
- [56] Samuel Idowu, Daniel Strüber, and Thorsten Berger. 2021. Asset management in machine learning: A survey. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 51–60.
- [57] IEEE. 2022. IEEE Xplore API Portal. (2022). <https://developer.ieee.org/>. Last accessed 19 June 2022.
- [58] Apple Inc. 2022. Core ML. (2022). <https://developer.apple.com/documentation/coreml>. Last accessed 2 October 2022.
- [59] Nourah Janbi, Iyad Katib, Aiiad Albesbri, and Rashid Mehmood. 2020. Distributed artificial intelligence-as-a-service (DAIaaS) for smarter IoT and 6G environments. *Sensors* 20, 20 (2020), 5796.
- [60] Lin Jia, Zhi Zhou, Fei Xu, and Hai Jin. 2021. Cost-efficient continuous edge learning for artificial intelligence of things. *IEEE Internet of Things Journal* 9, 10 (2021), 7325–7337.
- [61] Runyu Jin, Paul Muench, Veera Deenadhayalan, and Brian Hatfield. 2022. AIOps Essential to unified resiliency management in data lakehouses. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 4777–4781.
- [62] Alekh Jindal, Shi Qiao, Rathijit Sen, and Hiren Patel. 2021. Microlearner: A fine-grained learning optimizer for big data workloads at Microsoft. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2423–2434.
- [63] Meenu Mary John, Helena Holmström Olsson, and Jan Bosch. 2021. Towards MLOps: A framework and maturity model. In *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, 1–8.
- [64] Meenu Mary John, Helena Holmström Olsson, and Jan Bosch. 2022. Towards an AI-driven business development framework: A multi-case study. *Journal of Software: Evolution and Process* (2022), e2432.
- [65] Staffs Keele. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Technical Report. Technical report, ver. 2.3 EBSE.
- [66] Péter Kiss, Anna Reale, Charles Jose Ferrari, and Zoltán Istenes. 2018. Deployment of IoT applications on 5G edge. In *2018 IEEE International Conference on Future IoT Technologies (Future IoT)*. IEEE, 1–9.
- [67] Shiyi Kong, Jun Ai, and Minyan Lu. 2023. CL-MMAD: A contrastive learning based multimodal software runtime anomaly detection method. *Applied Sciences* 13, 6 (2023), 3596.
- [68] Georgios P. Koudouridis, Qing He, and György Dán. 2022. An architecture and performance evaluation framework for artificial intelligence solutions in beyond 5G radio access networks. *EURASIP Journal on Wireless Communications and Networking* 2022, 1 (2022), 1–32.

- [69] Dominik Kreuzberger, Niklas Kühl, and Sebastian Hirschl. 2023. Machine learning operations (MLOps): Overview, definition, and architecture. *IEEE Access* (2023).
- [70] Minah Lee, Xueyuan She, Biswadeep Chakraborty, Saurabh Dash, Burhan Mudassar, and Saibal Mukhopadhyay. 2021. Reliable edge intelligence in unreliable environment. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 896–901.
- [71] Sam Leroux, Pieter Simoens, Meelis Lootus, Kartik Thakore, and Akshay Sharma. 2022. TinyMLOps: Operational challenges for widespread edge AI adoption. In *2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE, 1003–1010.
- [72] Peizheng Li, Jonathan Thomas, Xiaoyang Wang, Ahmed Khalil, Abdelrahim Ahmad, Rui Inacio, Shipra Kapoor, Arjun Parekh, Angela Doufexi, Arman Shojaeifard, Arman Shojaeifard, and Robert J. Piechocki. 2022. RLOps: Development life-cycle of reinforcement learning aided open RAN. *IEEE Access* (2022).
- [73] Anderson Lima, Luciano Monteiro, and Ana Paula Furtado. 2022. MLOps: Practices, maturity models, roles, tools, and challenges-a systematic literature review. *ICEIS (1)* (2022), 308–320.
- [74] Lucy Ellen Lwakatare, Ivica Crnkovic, and Jan Bosch. 2020. DevOps for AI—challenges in development of AI-enabled applications. In *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 1–6.
- [75] Lucy Ellen Lwakatare, Aiswarya Raj, Ivica Crnkovic, Jan Bosch, and Helena Holmström Olsson. 2020. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and Software Technology* 127 (2020), 106368.
- [76] Ruchi Mahindru, Harshit Kumar, and Sahil Bansal. 2021. Log anomaly to resolution: AI based proactive incident remediation. In *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. IEEE, 1353–1357.
- [77] Sasu Mäkinen, Henrik Skogström, Eero Laaksonen, and Tommi Mikkonen. 2021. 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, 109–112.
- [78] Paolo Meloni, Daniela Loi, Paola Busia, Gianfranco Deriu, Andy D. Pimentel, Dolly Sapra, Todor Stefanov, Svetlana Minakova, Francesco Conti, Luca Benini, Maura Pintor, Battista Biggio, Bernhard Moser, Natalia Shepeleva, Nikos Fragoulis, Ilias Theodorakopoulos, Michael Masin, and Francesca Palumbo. 2019. Optimization and deployment of CNNs at the edge: The ALOHA experience. In *Proceedings of the 16th ACM International Conference on Computing Frontiers*. 326–332.
- [79] Microsoft. 2023. AIOps. (2023). <https://www.microsoft.com/en-us/research/project/aiops/>. Last accessed 29 August 2023.
- [80] Raúl Miñón, Josu Diaz-de Arcaya, Ana I. Torre-Bastida, and Philipp Hartlieb. 2022. Pangea: An MLOps tool for automatically generating infrastructure and deploying analytic pipelines in edge, fog and cloud layers. *Sensors* 22, 12 (2022), 4425.
- [81] Rui Manuel Morais. 2021. On the suitability, requisites, and challenges of machine learning. *Journal of Optical Communications and Networking* 13, 1 (2021), A1–A12.
- [82] Sabrina M. Neuman, Brian Plancher, Bardienu P. Duisterhof, Srivatsan Krishnan, Colby Banbury, Mark Mazumder, Shvetank Prakash, Jason Jabbour, Aleksandra Faust, Guido C. H. E. de Croon, et al. 2022. Tiny robot learning: Challenges and directions for machine learning in resource-constrained robots. *arXiv preprint arXiv:2205.05748* (2022).
- [83] Pia Niemelä, Bilhanan Silverajan, Mikko Nurminen, Jenni Hukkanen, and Hannu-Matti Järvinen. 2022. LAOps: Learning analytics with privacy-aware MLOps. In *CSEDU (2)*. 213–220.
- [84] Paolo Notaro, Jorge Cardoso, and Michael Gerndt. 2020. A systematic mapping study in AIOps. In *International Conference on Service-Oriented Computing*. Springer, 110–123.
- [85] Paolo Notaro, Jorge Cardoso, and Michael Gerndt. 2021. A survey of AIOps methods for failure management. *ACM Transactions on Intelligent Systems and Technology (TIST)* 12, 6 (2021), 1–45.
- [86] Paolo Notaro, Soroush Haeri, Jorge Cardoso, and Michael Gerndt. 2023. LogRule: Efficient structured log mining for root cause analysis. *IEEE Transactions on Network and Service Management* (2023).
- [87] Pierre-Emmanuel Novac, Ghouthi Boukli Hacene, Alain Pegatoquet, Benoît Miramond, and Vincent Gripon. 2021. Quantization and deployment of deep neural networks on microcontrollers. *Sensors* 21, 9 (2021), 2984.
- [88] Laud Charles Ochei, Andrei Petrovski, and Julian M. Bass. 2019. Optimal deployment of components of cloud-hosted application for guaranteeing multitenancy isolation. *Journal of Cloud Computing* 8, 1 (2019), 1–38.
- [89] Moses Openja, Forough Majidi, Foutse Khomh, Bhagya Chembakottu, and Heng Li. 2022. Studying the practices of deploying machine learning projects on docker. *arXiv preprint arXiv:2206.00699* (2022).
- [90] Fannia Pacheco, Ernesto Exposito, Mathieu Gineste, Cedric Baudoin, and Jose Aguilar. 2018. Towards the deployment of machine learning solutions in network traffic classification: A systematic survey. *IEEE Communications Surveys & Tutorials* 21, 2 (2018), 1988–2014.

- [91] Matthew J. Page, Joanne E McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery* 88 (2021), 105906.
- [92] Matthew J. Page, David Moher, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and Joanne E. McKenzie. 2021. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* 372 (2021).
- [93] Andrei Paleyes, Christian Cabrera, and Neil D. Lawrence. 2022. An empirical evaluation of flow based programming in the machine learning deployment context. *arXiv preprint arXiv:2204.12781* (2022).
- [94] Paurakh Paudyal, Sen Shen, Shuangyi Yan, and Dimitra Simeonidou. 2021. Toward deployments of ML applications in optical networks. *IEEE Photonics Technology Letters* 33, 11 (2021), 537–540.
- [95] Yabin Peng, Caixia Liu, Shuxin Liu, Yuchen Liu, and Yiteng Wu. 2022. SmartTRO: Optimizing topology robustness for Internet of Things via deep reinforcement learning with graph convolutional networks. *Computer Networks* 218 (2022), 109385.
- [96] Lorena Poenaru-Olaru, Luis Cruz, Jan S. Rellermeyer, and Arie Van Deursen. 2023. Maintaining and monitoring AIOps models against concept drift. In *2023 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN)*. IEEE, 98–99.
- [97] Neoklis Polyzotis, Sudip Roy, Steven Euijong Whang, and Martin Zinkevich. 2018. Data lifecycle challenges in production machine learning: A survey. *ACM SIGMOD Record* 47, 2 (2018), 17–28.
- [98] Christos Profentzas, Magnus Almgren, and Olaf Landsiedel. 2022. MiniLearn: On-device learning for low-power IoT devices. In *Proceedings of the 2022 International Conference on Embedded Wireless Systems and Networks (Linz, Austria) (EWSN’22)*. Junction Publishing, USA.
- [99] Laxmi Rijal, Ricardo Colomo-Palacios, and Mary Sánchez-Gordón. 2022. AIOps: A multivocal literature review. *Artificial Intelligence for Cloud and Edge Computing* (2022), 31–50.
- [100] Tanzila Saba, Amjad Rehman, Khalid Haseeb, Saeed Ali Bahaj, and Gwanggil Jeon. 2022. Energy-efficient edge optimization embedded system using graph theory with 2-tiered security. *Electronics* 11, 18 (2022), 2942.
- [101] Sofia Schöbel, Anuschka Schmitt, Dennis Benner, Mohammed Saqr, Andreas Janson, and Jan Marco Leimeister. 2023. Charting the evolution and future of conversational agents: A research agenda along five waves and new frontiers. *Information Systems Frontiers* (2023), 1–26.
- [102] Raj Mani Shukla and John Cartlidge. 2022. Challenges faced by industries and their potential solutions in deploying machine learning applications. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 0119–0124.
- [103] Haitao Song, Hongyu Ji, Ye Yu, and Bing Xie. 2022. A review of observability issues in hospital information system. In *2022 IEEE International Conference on Networking, Sensing and Control (ICNSC)*. IEEE, 1–7.
- [104] Rohith Sothilingam, Vik Pant, and Eric Yu. 2022. Using i* to analyze collaboration challenges in MLOps project teams. (2022).
- [105] Springer. 2022. Springer Nature API Portal. (2022). <https://dev.springernature.com/restfuloperations>. Last accessed 18 June 2022.
- [106] Rakshith Subramanya, Seppo Sierla, and Valeriy Vyatkin. 2022. From DevOps to MLOps: Overview and application to electricity market forecasting. *Applied Sciences* 12, 19 (2022), 9851.
- [107] Qi Sun, Chen Bai, Hao Geng, and Bei Yu. 2021. Deep neural network hardware deployment optimization via advanced active learning. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 1510–1515.
- [108] Georgios Symeonidis, Evangelos Nerantzis, Apostolos Kazakis, and George A. Papakostas. 2022. MLOps-definitions, tools and challenges. In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 0453–0460.
- [109] Damian A. Tamburri. 2020. Sustainable MLOps: Trends and challenges. In *2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNAS)*. IEEE, 17–23.
- [110] TensorFlow. 2022. TensorFlow Lite. (2022). <https://www.tensorflow.org/lite>. Last accessed 2 October 2022.
- [111] Kundjanasith Thonglek, Keichi Takahashi, Kohei Ichikawa, Chawanat Nakasan, Hidemoto Nakada, Ryousei Takano, Pattara Leelaprute, and Hajimu Iida. 2022. Automated quantization and retraining for neural network models without labeled data. *IEEE Access* 10 (2022), 73818–73834.
- [112] Cornell University. 2021. arXiv API Access. (2021). <https://arxiv.org/help/api/>. Last accessed 18 June 2022.

- [113] María Cora Urdaneta-Ponte, Amaia Mendez-Zorrilla, and Ibon Oleagordia-Ruiz. 2021. Recommendation systems for education: Systematic review. *Electronics* 10, 14 (2021), 1611.
- [114] Muhammad Usman, Simone Ferlin, Anna Brunstrom, and Javid Taheri. 2022. A survey on observability of distributed edge & container-based microservices. *IEEE Access* (2022).
- [115] Stylianos I. Venieris, Ioannis Panopoulos, Ilias Leontiadis, and Iakovos S. Venieris. 2021. How to reach real-time AI on consumer devices? Solutions for programmable and custom architectures. In *2021 IEEE 32nd International Conference on Application-specific Systems, Architectures and Processors (ASAP)*. IEEE, 93–100.
- [116] Stylianos I. Venieris, Ioannis Panopoulos, and Iakovos S. Venieris. 2021. OODIn: An optimised on-device inference framework for heterogeneous mobile devices. In *2021 IEEE International Conference on Smart Computing (SMART-COMP)*. IEEE, 1–8.
- [117] M. Vergin Raja Sarobin. 2020. Optimized node deployment in wireless sensor network for smart grid application. *Wireless Personal Communications* 111, 3 (2020), 1431–1451.
- [118] Birgit Vogel-Heuser, Emanuel Trunzer, Dominik Hujo, and Michael Sollfrank. 2021. (Re) deployment of smart algorithms in cyber–physical production systems using DSL4hDNCS. *Proc. IEEE* 109, 4 (2021), 542–555.
- [119] Xiaofei Wang, Yiwen Han, Victor C. M. Leung, Dusit Niyato, Xueqiang Yan, and Xu Chen. 2020. Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 22, 2 (2020), 869–904.
- [120] Jinfeng Wen, Zhenpeng Chen, Yi Liu, Yiling Lou, Yun Ma, Gang Huang, Xin Jin, and Xuanzhe Liu. 2021. An empirical study on challenges of application development in serverless computing. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 416–428.
- [121] Thorsten Wittkopp, Philipp Wiesner, Dominik Scheinert, and Alexander Acker. 2021. LogLAB: Attention-based labeling of log data anomalies via weak supervision. In *International Conference on Service-Oriented Computing*. Springer, 700–707.
- [122] Chunwei Xia, Jiacheng Zhao, Huimin Cui, Xiaobing Feng, and Jingling Xue. 2019. DNN Tune: Automatic benchmarking DNN models for mobile-cloud computing. *ACM Transactions on Architecture and Code Optimization (TACO)* 16, 4 (2019), 1–26.
- [123] Shu Yang, Kunkun Xu, Laizhong Cui, Zhongxing Ming, Ziteng Chen, and Zhong Ming. 2020. EBI-PAI: Toward an efficient edge-based IoT platform for artificial intelligence. *IEEE Internet of Things Journal* 8, 12 (2020), 9580–9593.
- [124] Junqi Yin, Shubhankar Gahlot, Nouamane Laanait, Ketan Maheshwari, Jack Morrison, Sajal Dash, and Mallikarjun Shankar. 2019. Strategies to deploy and scale deep learning on the summit supercomputer. In *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS)*. IEEE, 84–94.
- [125] Syed Ali Raza Zaidi, Ali M. Hayajneh, Maryam Hafeez, and Q. Z. Ahmed. 2022. Unlocking edge intelligence through tiny machine learning (TinyML). *IEEE Access* 10 (2022), 100867–100877.
- [126] Gorka Zárate, Raúl Miñón, Josu Díaz-de Arcaya, and Ana I. Torre-Bastida. 2022. K2E: Building MLOps environments for governing data and models catalogues while tracking versions. In *2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C)*. IEEE, 206–209.
- [127] Jie M. Zhang, Mark Harman, Lei Ma, and Yang Liu. 2020. Machine learning testing: Survey, landscapes and horizons. *IEEE Transactions on Software Engineering* (2020).
- [128] Jiang Zhaoxue, Li Tong, Zhang Zhenguao, Ge Jingguo, You Junling, and Li Liangxiong. 2021. A survey on log research of AIOps: Methods and trends. *Mobile Networks and Applications* 26, 6 (2021), 2353–2364.
- [129] Baifan Zhou, Yulia Svetashova, Tim Pychynski, Ildar Baimuratov, Ahmet Soylu, and Evgeny Kharlamov. 2020. SemFE: Facilitating ML pipeline development with semantics. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 3489–3492.

Received 19 December 2022; revised 5 September 2023; accepted 18 September 2023