Zurich University of Applied Sciences
School of Management and Law

Master Thesis

Master of Science (MSc) in Business Information Technology

Optimizing IT Operations with AIOps: An Investigation into the Opportunities and Challenges for Enterprise Adoption

How can companies benefit from using AIOps on their mission critical applications?

How can AIOps be implemented in established IT departments?

Supervisors: Dr. Anna Wiedemann
Björn Scheppler

Author: Mario Gian Locher
18-675-199

MAWIN21HSb

Submission date: May 31, 2023

Acknowledgements

With these lines I would like to thank all the people who supported me in different ways during the realization of this Master Thesis.

Special thanks go to my supervising lecturer, Dr. Anna Wiedemann, for her very valuable support during the whole period of my work.

In addition, I would like to express my special thanks to my interview partners, who made significant contributions to this work through their very valuable statements and insights.

Finally, my thanks go to my personal environment, my partner, friends, and family for the always stable support during the entire study.

Mario Gian Locher

Management Summary

Digitalization and the accompanying technological change are forcing companies to constantly evolve. This also applies to IT operations, which must transform digitally. As cloud computing and infrastructure virtualization are standard in today's IT environments, IT operations teams are faced with increased complexity. The vast amount of data produced cannot be managed by humans. Hence, they need to leverage advanced technologies to prevent and manage incidents that threaten the IT operation and thus also the company. Artificial Intelligence for IT Operations (AIOps) promises to solve today's challenges in IT operations by incorporating Artificial Intelligence (AI) into widely used solutions in IT operations. AIOps should allow operations to move from a reactive to a proactive approach, identifying and managing incidents before they occur, while also allowing teams to build resilience into the system.

The aim of this Master Thesis was to show how companies can benefit from using AIOps on their mission critical applications. Besides that, it should be shown how AIOps can be implemented in established IT departments. To provide a holistic view on the topic, challenges and limitations were also considered. As applying AI to IT operations does not in itself solve a business problem, a business AIOps alignment model is presented that should provide guidance for companies considering AIOps.

To answer the research questions, a single case study was conducted in addition to a multivocal literature review on AIOps. The case study focused on a provider of AIOps solutions and its implementation partners. The interviews provided insight in the real-world adoption of AIOps.

The findings of this work show that AIOps should rather be seen as a journey than as a specific technology. AIOps allows companies to move from a reactive IT operations approach to a proactive one. By freeing up time from operations teams, companies can focus on building resilience into the system, which is seen as the most successful incident prevention strategy. Although technology is already capable of predicting incidents in advance, this capability has not yet caught on in the market, largely because the data, processes, culture, and tools in organizations are not ready. Successfully adopting AIOps requires alignment to the business strategy which can be achieved using the presented business AIOps alignment model. Although aiming to solve today's IT operations challenges, implementing AIOps holds organizational, cultural, and technological challenges. These challenges must be considered and overcome to successfully deploy AIOps and fully realize its potential. Properly implemented and deployed, AIOps helps organizations reduce the highly negative impact of incidents in their mission-critical applications.

Table of Content

List of Figures

List of Tables

Abbreviation List

AI	Artificial Intelligence
AIAA	Artificial Intelligence and Autonomous Applications
AIOps	Artificial Intelligence for IT Operations
DNN	Deep Neural Network
DRAM	Dynamic Random Access Memory
DSA	Decision Support Analytics
ETL	Extract, Transform and Load
HTTP	Hypertext Transfer Protocol
I/O	Input/Output
IcM	Incident Management
IT	Information Technology
KPI	Key Performance Indicator
LSTM	Long Short-Term Memory
ML	Machine Learning
MLR	Multivocal Literature Review
MTTD	Mean Time to Detect
MTTR	Mean Time to Repair
MSP	Managed Service Provider
MVP	Minimum Viable Product
NLP	Natural Language Processing
OCE	On-call Engineer
PoC	Proof of Concept
QoS	Quality of Service
RAID	Redundant Array of Independent Disks
SLA	Service Level Agreement
SLO	Service Level Objective
SMART	Self-Monitoring Analysis and Reporting Technology
SRE	Site Reliability Engineer
SSH	Secure Shell
SVM	Support Vector Machine
TTM	Time to Mitigate
TTR	Time to Resolve
VM	Virtual Machine

1 Introduction

Advances in cloud computing and infrastructure virtualization revolutionized the development, deployment, and operation of enterprise applications in the last decade (McCreadie et al., 2022, p. 136). Modern enterprises are transitioning form static and fragmented physical systems to dynamic cloud-based environments, which combine on-premises and cloud-based resources (McCreadie et al., 2022, p. 136). As a result of this cloud adoption, multiple cloud architectures, such as hybrid cloud and multi cloud, have emerged (Illsley & Grossner, 2021, p. 3). Cloud migration results in an operating landscape that is becoming increasingly complex (Shen et al., 2020, p. 276). IT operations teams must deal with vast amounts of data and alarms generated by these modern systems, making it impossible for a human to respond rapidly enough to keep the IT systems and networks operational (Shen et al., 2020, p. 276; Illsley & Grossner, 2021, p. 3). Furthermore, downtimes and components that are not functioning correctly can lead to significant economic loss and image damage, especially when mission-critical applications are affected (Gillis, 2018; Gulenko, 2020, p. 1). Mission-critical applications can be defined as "software program or suite of related programs that must function continuously in order for a business or segment of a business to be successful" (Gillis, 2018).

While this cloud adoption leads to more complexity, advancements in technology also open new possibilities to control all levels of the infrastructure stack, not only the server landscape, but also the connected front-end devices and the communication paths (Gulenko et al., 2020, p. 1). Artificial Intelligence for IT Operations (AIOps), first introduced by Gartner in 2016, uses big data, machine learning, and other advanced analytic technologies to improve IT operations directly and indirectly (Shen et al., 2020, p. 276). This optimization potential can be used to increase the reliability and resilience of the overall system (Gulenko et al., 2020, p. 1).

The literature shows that there are several different application areas of AIOps, such as incident management, anomaly detection, incident prediction, and automated resource allocation (Nedelkoski et al., 2019, p. 179; Chen et al., 2021, p. 1; Lyu et al., 2022, pp. 6-7). The goals of applying AIOps in these areas are high service intelligence and quality, high external user satisfaction, and high internal user satisfaction and productivity (Dang et al., 2019, p. 4; Prasad et al., 2022, pp. 13-14). Most literature focuses on different machine learning models used to serve the different application areas and the benefits cloud providers can get when adopting AIOps. What, on the other hand, mostly not has been considered is the adoption of AIOps in real-world scenarios. Therefore, the focus of this Master Thesis lies on the adoption of AIOps in companies and answers the following two research questions: (1) How can companies benefit from

using AIOps on their mission critical applications? (2) How can AIOps be implemented in established IT departments?

The structure of this Master Thesis is as follows. First, a multivocal literature review is conducted to evaluate the state-of-the-art and -practice in the field of AIOps. There, the different application areas and goals of AIOps are presented to then show the needed capabilities of an AIOps solution. After that, the theoretical background of business IT alignment and the challenges in creating value from AI are shown. After that, the case study to be conducted is described. The case study is used to show how AIOps can be implemented, what benefits AIOps brings and what challenges companies face in IT operations. The case will be evaluated against the literature and provides insights of a vendor and two implementation partners of AIOps solutions. Finally, the findings of the case study research are discussed, and a business AIOps alignment model is presented, which should function as a reference for companies considering an AIOps implementation. The model is based on the findings of the multivocal literature review and the case study.

2 AIOps – Multivocal Literature Review

The following paragraphs describe the conducted literature review, which has been done to identify the research gap in the field of AIOps. As Prasad et al. (2022, p. 7) state, “there is no future of IT operations that does not include AIOps”. The following chapter is structured as follows: First, the approach is described and justified. This is followed by a brief overview of IT operations, the history of AI, and the various levels of automation. Then, the findings of the literature review are presented, including different AIOps application areas, the goals of AIOps, and the capabilities of AIOps solutions.

2.1 Planning the Review

AIOps is a new concept, which was first introduced by Gartner in 2016 (Shen et al., 2020, p. 276). Therefore, the academic literature on AIOps is limited. Garousi et al. (2019, p. 104) state that including gray literature in a literature review can have substantial benefits. In contrast to formal literature, which has been peer reviewed, gray literature is defined as literature which has not been formally published in books or journals and thus did not undergo a formal peer review process (Garousi et al, 2017, p. 3). Literature reviews including both formal and gray literature are called Multivocal Literature Reviews (MLR) (Garousi et al., 2017, p. 3). The benefit of MLRs is that they include the state-of-the-art and -practice, including the view of practitioners, which gives implications of the current perspectives and complements the formal literature (Garousi et al., 2017, p. 3; Garousi et al., 2019, p. 108). Therefore, MLR studies target both re-

searchers and practitioners, since they aim to synthesize formal and gray literature (Garousi et al., 2017, p. 15).

Garousi et al. (2019, p. 109) propose seven questions to decide whether gray literature should be included in a literature review. These questions are shown and answered regarding the field of AIOps in Table 1.

#	**Question**	**Answer**
1	Is the subject "complex" and not solvable by considering only the formal literature?	Yes
2	Is there a lack of volume or quality of evidence, or a lack of consensus of outcome measurement in the formal literature?	Yes
3	Is the contextual information important to the subject under study?	Yes
4	Is it the goal to validate or corroborate scientific outcomes with practical experiences?	Yes
5	Is it the goal to challenge assumptions or falsify results from practice using academic research or vice versa?	Yes
6	Would a synthesis of insights and evidence from the industrial and academic community be useful to one or even both communities?	Yes
7	Is there a large volume of practitioner sources indicating high practitioner interest in a topic?	Yes

Table 1: Including Gray Literature in the Literature Review (Garousi et al, 2019, p. 109)

Garousi et al. (2019, p. 109) suggest, that if one of these questions is answered with a "yes" one should include gray literature in the review. Regarding the field of AIOps, all questions are answered with yes. The exclusion of gray literature is suggested in mature fields of research (Garousi et al., 2017, p. 15), which AIOps is not considered a part of. Thus, gray literature should be included in the conducted literature review. The seven questions were answered positively, because the field of AIOps research is only several years old and the found white and gray literature implies that the AIOps topic finds interest in both the academic and the practical community (Shen et al., 2020, pp. 276-280; Wang et al., 2020, pp. 417-422; McKeon-White et al., 2021, pp. 1-9; Prasad et al., 2022, pp. 1-21).

Garousi et al. (2019, p. 103) define three different tiers of gray literature based on the credibility and outlet control of the sources. First tier includes sources with high outlet control and high credibility (e.g., books, magazines, government reports, white papers). Second tier have moder-

ate outlet control and credibility (e.g., annual reports, news articles, presentations, videos, Q&A sites, wiki articles). Third tier are of low control and low credibility (e.g., blogs, e-mails, tweets). For the conducted MLR only first tier gray literature was included to ensure high credibility.

2.2 Conducting the Review

After stating why a MLR approach has been chosen, the next paragraphs focus on the review process. The search process and the data extraction process will be shown to proof the credibility of the review (Vom Brocke et al., 2009, p. 1).

2.2.1 Search Process

The first step of the search process was defining a search string to use in the different databases. To narrow down the search results, the search string was defined to "AIOps". The search for academic literature was conducted in the following in Table 2 presented databases. The search timeframe was limited to 2016 until 2022, since AIOps was first introduced in 2016 (Shen et al., 2020, p. 276). To further limit the number of papers, only articles with the string "AIOps" included in the title and abstract were considered in a first step and appear in the below Table 2 as "Reviewed". In addition, only papers in English were considered.

Database	**Hits**	**Reviewed**
IEEE	62	14
ACM	60	13
Swisscovery	139	27
Google Scholar	1030	Top 50
ScienceDirect	53	2

Table 2: Reviewed Databases

Besides that, the following in Table 3 stated journals on special interest groups for Artificial Intelligence and Autonomous Applications (AIAA) and Decision Support Analytics (DSA) recommended journals have been searched.

Journal	Hits	Reviewed
IEEE Intelligent Systems	0	0
Expert Systems	1	0
Expert Systems with Applications	3	3
Intelligent Systems in Accounting, Finance and Management	0	0
Decision Support Systems	0	0
Decision Sciences	0	0
Information Systems Frontiers	0	0
European Journal of Operational Research	1	0

Table 3: Reviewed Journals

This small number of search results in the stated journals confirms that the number of research papers on AIOps published in journals is limited. To include more relevant sources, forward and backward citation search, also referred to as "snowballing", has been applied (Garousi et al., 2017, p. 20). This led to studies, which did not explicitly include the string "AIOps" in title or abstract but were related to the topic.

The search string "AIOps" also led to 6.55 million Google search results in the timeframe of 2016 until 2022. Like suggested by Garousi et al. (2017, p. 18) the first ten Google result pages have been checked. Garousi et al. (2017, p. 18) refer to this stopping rule as the "effort bounded", meaning that only the top N search engine hits should be considered, to know when to stop searching further.

As Garousi et al. (2017, p. 18) state, the gray literature is more diverse and less controlled than academic literature. Thus, it is needed to perform a source selection process, to define which resources should be considered in the review (Garousi et al., 2017, p. 18). Based on this suggestion, the following source selection process has been used: (a) duplicated records which have already been reviewed have been excluded, (b) webpages from software vendors have been excluded, due to their subjective interpretation of the benefits of AIOps and the assumption, that they are promoting their individual AIOps solutions, which would not add value to the review, (c) focus was placed on relevant magazines, discussing science and business related topics, as well as analyst research from Gartner, Forrester, Omdia, and S&P Global.

The conducted research of both academic and gray literature resulted in a total of 55 sources that were deemed relevant for the review and thus have been analyzed in detail.

2.2.2 Data Extraction Process

After downloading the various sources as PDF, they were inserted into a literature management program. There, the papers were structured based on different topics regarding AIOps. Beginning with IT Operations and Maintenance in general and ending with business use case related information. To be able to define the structure, abstract, introduction and conclusion of each paper were read first. If these sections indicated further exciting information, the entire paper was read, and information was extracted from it accordingly. The extracted information was paraphrased and stored in the hierarchy created. This approach can be defined as thematic analysis (Cruzes & Dybå, 2010, p. 3).

2.3 IT Operations

Every industry and aspect of human life has been transformed by information technology (IT) (Levin et al., 2019, p. 165). As humanity's reliance on computing grows, IT installations become larger and more complex, necessitating an increasing number of resources for setup and operation (Levin et al., 2019, p. 165). In the last decade, advances in cloud computing and infrastructure virtualization have revolutionized enterprise application development, deployment, and operation (McCreadie et al., 2022, p. 136). The introduction of containers and operating system (OS) virtualization enables the packaging of complex applications within isolated environments, raising the abstraction level for application developers while also increasing cost effectiveness and deployment flexibility (McCreadie et al., 2022, p. 136). Similarly, microservice architectures enable application delivery via composite services that can be developed and deployed independently by different IT teams (McCreadie et al., 2022, p. 136). In this context, organizations are shifting away from traditional static and fragmented physical systems and toward more dynamic cloud-based environments that combine resources from various on-premises and cloud environments (McCreadie et al., 2022, p. 136).

This new complexity led to difficulties for IT operation and maintenance teams in managing the further growing IT landscape (Shen et al., 2020, p. 276). A massive amount of data and alerts are generated by the different systems, which makes it impossible to manage them manually (Shen et al., 2020, p. 276). While providing advanced data-driven analytics to the users, the IT operation itself relied on manual work for a long time (Levin et al., 2019, p. 165). To be able to manage this new era of IT systems, IT operations needs to digitally transform itself to be able to oversee and manage these systems holistically (Levin et al., 2019, p. 165; Gulenko et al., 2020, p. 1).

According to Shen et al. (2020, p. 276) the IT operations practices can be divided into five different eras along with the rapid evolvement of the whole IT industry. Starting with the "age of

manual", where all operations were performed by manually logging into the devices (Shen et al., 2020, p. 276). In the "age of scripts", operations management was done by writing scripts, which automatically performed the defined functions (Shen et al., 2020, p. 276). Scripts are still a common part of IT operations due to their convenient use (Shen et al., 2020, p. 276). However, scripts are not user-friendly and thus, in the "age of small systems" client/server and browser/server architectures became popular (Shen et al., 2020, p. 276). Unfortunately, the interoperability and data sharing possibilities were only poorly supported by such systems, which then led to the "age of platforms" (Shen et al., 2020, p. 276). The different modules running on these platforms shared storage and computing resources and were interoperable (Shen et al., 2020, p. 276). These functionalities strongly benefited from the rise of cloud computing and big data technology (Shen et al., 2020, p. 276). An extension of such platforms led to the fifth era, defined as the "age of AIOps", in which artificial intelligence (AI) and big data technology were incorporated in the platforms (Shen et al., 2020, p. 276). According to Shen et al. (2020, p. 276) this era is still in primary stage in terms of technology and applications.

The goal of IT operations and maintenance in all the above-mentioned eras is to provide a high Quality of Service (QoS) with uninterrupted services (Gulenko et al., 2020, p. 1). Downtimes due to malfunctioning infrastructure or system components can have a significant negative financial impact (Gulenko et al., 2020, p. 1). For instance, the average cost for an hour of server downtime of an organization is stated to be between $300.000 and $400.000 (Lyu et al., 2022, p. 2). Such significant losses highlight the need to address the main reasons for system failures (Farshchi et al., 2018, p. 531). Operational and configuration problems are reported to be a major cause of system failures overall (Farshchi et al., 2018, p. 531). A reason for that is the complexity of modern, large-scale applications, especially in cloud environments (Farshchi et al., 2018, p. 531).

Reducing the impact of incidents in the system, and thus ensuring a high QoS, is a task of so called On-Call Engineers (OCEs) (Jiang et al., 2020, p. 1411). When an incident occurs, it is reported automatically by the system, in case a monitoring system is used (Jiang et al., 2020, p. 1411). Such an incident management system continuously monitors the services and detects incidents (Jiang et al., 2020, p. 1411). If an incident occurs and is detected, the OCEs will be informed automatically and the investigation process of the incident starts (Jiang et al., 2020, p. 1411). To be able to solve the issue and minimize the impact, the OCEs need to understand the reasons of the incident and identify the root cause (Jiang et al., 2020, p. 1411). However, root causes of incidents are diverse and thus it usually needs a lot of time to locate them (Jiang et al., 2020, p. 1411). Therefore, OCEs need to mitigate the incident and bring the service back to its normal state first and then resolve the root cause in a second step (Jiang et al., 2020, p. 1411).

Incident management of traditional on-premises software and online services is different (Lou et al., 2017, p. 907). Lou et al. (2017, p. 907) describe the following three main differences: First, for an online service system the cost of each hour of service downtime is higher, thus the incident needs to be resolved quicker. Second, OCEs usually use temporary workaround solutions rather than root cause resolution to restore service as soon as possible. Third, unlike on-premises software, when an incident occurs in an online service, it is usually impractical to attach a debugger to the service to diagnose the incident. As a result, the only way for OCEs to diagnose the incident and manage it, is to analyze collected monitoring data. This data is usually collected at runtime in form of service logs, performance counters, and machine-/process- and service-level events (Lou et al., 2017, p. 907). Analyzing such data manually is an impossible task (Gulenko et al., 2020, p. 1). Thus, technology to analyze this amount of data is needed to successfully provide a high QoS.

Technology opens new possibilities to control all levels of the infrastructure stack, not only the server landscape, but also the connected front-end devices and the communication paths (Gulenko et al., 2020, p. 1). This optimization potential can be used to increase the reliability and resilience of IT systems (Gulenko et al., 2020, p. 1). Response time can be shortened in case an urgent action is required – e.g., in case of performance problems, failures or security incidents (Gulenko et al., 2020, p. 1). In such a situation, the system operates outside the expected and defined parameters (Gulenko et al., 2020, p. 1). Thus, an anomaly occurs that must be detected and corrected before it causes a system component to fail (Gulenko et al., 2020, p. 1). The search for the cause takes valuable time, which in the worst case leads to a service failure (Gulenko et al., 2020, p. 1). To speed up this process, mechanisms and tools have been developed to detect and fix these anomalies (Gulenko et al., 2020, p. 1). These are referred to as Artificial Intelligence for IT Operations (AIOps) (Gulenko et al., 2020, p. 2).

2.4 Artificial Intelligence for IT Operations (AIOps)

The following sections aim to link IT operations with Artificial Intelligence (AI). For this purpose, first AI in general and the levels of automation in operations will be discussed, and then the current state-of-the-art and -practice in the field of AIOps is presented. The concept of AIOps was first introduced by Gartner in 2016 as “Algorithmic IT Operations”, to then be redefined in 2017 to “Artificial Intelligence for IT Operations” (Shen et al., 2020, p. 276).

2.4.1 Artificial Intelligence and Automation

The term “Artificial Intelligence” was first used 1956 in a research project at Dartmouth College (Dick, 2019, p. 1). The initial AI research attempted to uncover formal processes that incorporated intelligent human behavior in medical diagnosis, chess, mathematics, language processing, etc., with the goal of automating such behavior (Dick, 2019, p. 1). Human intelligence served as the central model for early attempts at automation (Dick, 2019, p. 1). However, today’s research aims to design automated systems that perform well in complex problem domains in any way other than human-like methods (Dick, 2019, p. 1). Nowadays, AI-related applications perform a wide range of tasks that would be impossible for unaided human intelligence to complete (Floridi, 2016, p. 140). In an increasing number of contexts, reproductive AI outperforms and replaces human intelligence (Floridi, 2016, p. 140).

The degree of automation can be divided into five levels proposed by Ganek and Corbi (2003, p. 9), namely “basic”, “managed”, “predictive”, “adaptive”, and “autonomic”. In Level 1, a human manually controls the operation and is not assisted by a system. In Level 2, a system supports the human by consolidating the data, based on which the IT staff can take actions. Thus, the system awareness and productivity can be improved. In Level 3, the system recommends actions, based on the correlated monitoring data, which can then be approved and initiated by a human. This reduces the dependency on deep skills of the IT personnel and enables faster and better decision making. In Level 4, the system automatically takes actions, and the IT staff only manages the performance against predefined Service Level Agreements (SLAs). This increases the IT agility and resilience and requires minimal human intervention. In Level 5, the system components are automatically managed based on business rules and policies. The focus of the IT personnel lies on the enablement of business needs. Hence, the business policies drive the management of IT and business agility, and resilience is increased. The defined levels do not state a specific technology, and thus can be applied to the field of AIOps.

From Level 3, the system needs intelligence to be able to predict failures and automatically take actions. E.g., to automatically detect an anomaly in the system, which could lead to a service failure, machine learning techniques are needed (Wang et al., 2019, pp. 94-95). The two main approaches are unsupervised and supervised learning (Wang et al., 2019, pp. 94-95). In unsupervised learning, the model can be trained without labelled data (Wang et al., 2019, p. 95). Contrarily, supervised learning uses manually labelled data to train the model (Wang et al., 2019, p. 95). While unsupervised learning focuses on normal data, and self-detects when an anomaly in the data occurs, supervised learning needs to have the anomalous data manually labeled by an operator (Wang et al., 2019, p. 95). Because supervised learning focuses on both normal and anomalous data, it is easier to get accurate results with supervised learning models

(Wang et al., 2019, p. 95). However, the accuracy of the model heavily relies on the accuracy of manual data labelling, since it takes the given labels as ground truth (Wang et al., 2019, p. 95). Hence, inappropriate labelling could impact the model performance (Wang et al., 2019, p. 95).
An example of supervised learning are random forest models (Breiman, 2001, p. 5). A random forest consists of several combined tree predictors, such as classification and regression trees (Bansal et al., 2020, p. 205). Both classification and regression trees are comprised of split and leaf nodes, hierarchically shaped like a tree, with each node being described by a predicate (Bansal et al., 2020, p. 205). Classification trees are used if the target variable is categorical (e.g., true/false); regression trees are used if the target variable is continuous (e.g., temperature or age) (Bansal et al., 2020, p. 205).

2.4.2 AIOps Application Areas

For many years, the failure of systems and software has been a topic for research and investigation (Farshchi et al., 2018, p. 531). Different industry surveys show that several types of system failures result in significant losses of money, market share, and reputation (Farshchi et al., 2018, p. 531). Problems such as job termination, hard disk failures, and performance anomalies in computer systems are unavoidable (Lyu et al., 2022, p. 6). To reduce or eliminate the associated financial losses, service reliability must be ensured (Lyu et al., 2022, p. 6). The reduction of the influence an incident has on the system can be done in two ways. Either the occurrence of an incident is predicted in advance, which allows engineers to take proactive actions to prevent it, or the already happened incident is mitigated in the shortest possible time span (Zhao et al., 2020, p. 315). AIOps solutions are designed to contribute to this influence reduction an incident has on the system (Lyu et al., 2022, p. 6). Several studies have been conducted on different applications of AIOps. The following aims to discuss the outcomes of these studies and show the possible AIOps application areas.

2.4.2.1 Incident Management

Cloud providers such as Microsoft, Amazon, Google, and IBM strive to deliver computing resources as quickly as possible in a dynamically scalable and virtualized environment (Chen et al., 2019b, p. 2659). A typical cloud system includes numerous subsystems (i.e., services), each of which is made up of many interconnected components (Chen et al., 2019b, p. 2659). These services run 24x7 and must be available without interruptions (Lou et al., 2014, p. 1583). Each component has its own monitors that verify the component's runtime status on a frequent basis (Chen et al., 2019b, p. 2659). However, disruptions during ongoing operations (unplanned interruptions or service outages) often cannot be prevented (Lou et al., 2014, p. 1583). A failure can dramatically degrade the availability of the system which can lead to bad user experience (Chen et al., 2019, p. 2659). Such incidents can lead to significant economic losses or other profound

consequences (Lou et al., 2014, p. 1583). Several studies have been conducted to simplify incident management through technology, covering topics such as incident diagnosing (Lim et al., 2014, pp. 320-329; Lou et al., 2014, pp. 1583-1592; Aggarwal et al., 2021, pp. 124-135; Arya et al., 2021, pp. 188-192; Shi et al., 2021, pp. 1-11), linking (Chen et al., 2020b, pp. 304-314), prioritizing (Chen et al., 2020a, pp. 373-384), and triaging (Chen et al., 2019a, pp. 364-375), as well as a combination of them (Chen et al., 2019b, pp. 2659-2665). The following paragraphs explain each of these topics and show the results of the conducted studies.

Aggarwal et al. (2021, p. 124) found, that if an incident is correctly diagnosed, Site Reliability Engineers (SREs) would quickly be able to derive the actions needed to solve the problem. Thus, to minimize downtimes and financial losses, a lot of work has been invested in improving the efficiency of service incident diagnosis (Lou et al., 2014, p. 1583).

System failures keep occurring due to frequent updates of system components, changes in the operation environment, mobility of devices etc. (Chen et al., 2020b, p. 304). In practice, an incident management (IcM) system is used to manage incidents and to ensure a high quality of service (QoS) (Chen et al., 2020b, p. 304). Once an incident occurs, technicians review system logs and perform troubleshooting actions (Chen et al., 2020b, p. 304). To detect an incident, data analysis of runtime telemetry data is required (Lou et al., 2014, p. 1583). Telemetry data can be divided into continuous time series data and temporal event data (Lou et al., 2014, p. 1583). A time series is a sequence of real-valued data points measured at different points in time in a uniform time interval (Lou et al., 2014, p. 1583). An example for time series data in an online service is the CPU utilization performance counter (Lou et al., 2014, p. 1583). On the other hand, in temporal event data, temporal event sequences record the occurrence of a particular software message, indicating that something has happened in the system (Lou et al., 2014, p. 1583). For instance, a “memory shortage” event sequence contains events where the system ran out of memory, which means that there was not enough memory in the system (Lou et al., 2014, p. 1583).

A key factor of data-driven incident diagnosis is correlation analysis (Lou et al., 2014, p. 1583). Although correlation relationships do not necessarily reveal the root cause of an incident, they often provide insights for causality analysis and provide information that points to the root cause (Lou et al., 2014, p. 1583). However, due to the heterogeneity of the data types mentioned in the previous paragraph (time series and event data), conventional correlation analysis often cannot provide satisfactory results (Lou et al., 2014, p. 1583). Luo et al. (2014, pp. 1586-1592) propose an approach to evaluate the correlation between a time series and an event sequence, which shows to be an effective method to diagnose incidents. They were able to show that by

transforming the correlation problem to a two-sample problem, and thus solving the problems independently before evaluating the correlation, can be an effective approach to address the challenge of different data types (Luo et al., 2014, pp. 1591f.). However, their work does not show how this correlation is then used to diagnose incidents. Arya et al. (2021, pp. 189-191), on the other hand, use a combination of both time series and temporal event sequences for root cause analysis, and state how this combination is used in incident diagnosis. By modeling log data as time series and event sequence, they create a causal graph between different microservices, to evaluate if errors in one microservice are caused by errors in other microservices (Arya et al., 2021, p. 189). With their approach, applied to microservices of a train ticketing system, they were able to automatically identify root causes of occurring incidents (Arya et al., 2021, p. 191). Both studies show that correlating time series and event data is important in incident diagnosis.

Besides system failures, performance issues are another factor that impacts a service negatively, and the same issues often occur multiple times (Lim et al., 2014, p. 320). System performance is critical for user satisfaction and project success, as users may switch to competing providers, if their performance need is not fulfilled (Lim et al., 2014, p. 320). Measuring system performance is usually done by using Key Performance Indicators (KPIs), which reflect the end-user experience (Lim et al., 2014, p. 321). These KPIs can be calculated, for instance, through user request tracking at the server side or by measuring the response time at the client side (Lim et al., 2014, p. 321). To check whether the system is in a healthy state, each KPI has a Service Level Objective (SLO) threshold (Lim et al., 2014, p. 321). If a KPI exceeds its threshold, the SLO is violated, and the system experiences a performance issue (Lim et al., 2014, p. 321). To be able to diagnose performance issues, a large amount of metric data is collected while the system is executing (Lim et al., 2014, p. 321). These metrics reflect system events and resource usage, such as CPU utilization, memory usage, disk queue lengths, Input/Output operation rate, and kernel events (Lim et al., 2014, p. 321).

Once a performance issue is identified, engineers must resolve it and restore the system as soon as possible (Lim et al., 2014, p. 321). However, manually detecting and fixing such performance issues can cause long maintenance time, due to the system's complexity (Lim et al., 2014, p. 321). It is therefore needed to automatically discover issues to help reducing the time wasted for repeated diagnosing and troubleshooting (Lim et al., 2014, p. 321). Lim et al. (2014, p. 320-329) proposed an approach, where they automatically perform performance issue diagnosis, by using a clustering technique. This enables them to effectively identify and detect reoccurring performance issues as well as unknown issues (Lim et al., 2014, p. 329).

Incidents, performance issues and in the worst-case downtimes are often caused by failing IT equipment (Shi et al., 2021, p. 1). Hard disks are among the most frequently failing components in today's IT environments, especially for cloud-based applications (Shi et al., 2021, p. 1). Diagnosing such hard disk failures is challenging (Shi et al., 2021, p. 1). Shi et al. (2021, p. 1) state, that several researched and developed models for diagnosing hard disk failures assume that the Self-Monitoring Analysis and Reporting Technology (SMART) data from different disks are subject to the same distribution. They mention however, that in a data center, there are diverse types of hard disks from several manufacturers in place, and that their SMART encoding varies widely (Shi et al., 2021, p. 1). This influences the generalization of machine learning methods used for hard disk diagnosing (Shi et al., 2021, p. 1). Another factor mentioned by Shi et al. (2021, p. 2) is that many studies assume that faulty hard disks are easy to find in the target area. However, they note that in a real IT operating scenario, hard disks usually operate in a healthy state, which means that faulty events are exceedingly rare and almost never occur in new hard disks (Shi et al., 2021, p. 1). Hence, it is challenging to diagnose faults on new hard disks that are different from old hard disks when there are no faulty samples on them.

The approach shown by Shi et al. (2021, pp. 3-6) overcomes the just mentioned challenges by using a deep network that generates faulty examples data and combines it with a transfer learning model. This solves the distribution difference issue between the different disk types (Shi et al., 2021, p. 5). They were able to show that their model can then be used to diagnose faults of new hard disks and outperforms other methods for recognizing new hard disk failures (Shi et al., 2021, pp. 5-10). Moreover, they state that it can be easily brought into operation due to the reduction of convergence time and speeding up of the model training (Shi et al., 2021, p. 10).

Another part of incident diagnosing is the linking of incidents. As Chen et al. (2020b, p. 304) state, incidents are often linked together, meaning that resolving one incident may as well resolve others. Thus, they presented a framework for predicting linked incidents (Chen et al., 2020b, pp. 304-313). Their framework can identify links among different incidents using semantic information in incident description and the dependency structure of the online service system (Chen et al., 2020b, pp. 309-312). To analyze the semantic information, they used a deep learning textual embedding module, which can encode the symptoms reported from automated monitors (expressed in a structured pattern) as well as human engineers (expressed in natural language) (Chen et al., 2020b, p. 310). They state that identifying these links correctly can not only help mitigating the incidents but also analyzing the root causes to prevent the occurrence of similar incidents in the future (Chen et al., 2020b, p. 313).

Due to the complexity of an online service system, built with different components such as hardware, virtual machines, network, database etc., incidents can occur frequently (Chen et al., 2020a, p. 374). Incidents should be mitigated timely since a long Time to Mitigate (TTM) could lead to poor service availability and cause huge economic loss (Chen et al., 2020a, p. 374). However, the number of engineers as well as the computing resources are limited (Chen et al., 2020a, p. 374). Hence, it is impossible to mitigate every incident timely (Chen et al., 2020a, p. 374). To reduce the impact of incidents on the service, one of the most cost-effective solutions is to deal with more serious and urgent incidents sooner (Chen et al., 2020a, p. 374). That is, engineers must prioritize incidents to optimize the incident management process (Chen et al., 2020a, p. 374).

In practice, after an incident is diagnosed, engineers often manually record whether the incident needs to be fixed with a high priority and give a simple explanation for the incident in the Incident Management System (IcM) (Chen et al., 2020a, p. 375). If an incident needs to be fixed, they include the needed fixing steps in this record (Chen et al., 2020a, p. 375). Thus, even if an incident is low priority, engineers must still spend time and resources diagnosing why the incident occurs, only to discover that it is indeed low priority (Chen et al., 2020a, p. 375). Chen et al. (2020a, p. 376) investigated eighteen real-world online service systems and found that more than half of incidents that occurred are low-priority incidents which should not be managed by engineers. If these are not correctly prioritized, a lot of engineer effort is wasted by spending time to solve these low-priority incidents (Chen et al., 2020a, p. 376). They state that the time spent on resolving this high percentage of incidental incidents takes on average more than half of the Time to Resolve (TTR) spent (Chen et al., 2020a, p. 376). Meaning that the cost spent on incidental incidents is almost the same as that spent on high priority incidents in terms of TTR (Chen et al., 2020a, p. 376). Hence, the resolution of high priority incidents may be delayed and thus results in greater economic loss (Chen et al., 2020a, p. 376).

A large amount of labeled incident data is accumulated during the incident management process (Chen et al., 2020a, p. 376). Each incident is recorded, along with details such as the symptom description and the environment in which it occurred (Chen et al., 2020a, p. 376). This huge amount of data allows for the automatic determination of whether an incident is low or high priority (Chen et al., 2020a, p. 378). Chen et al. (2020a, p. 378) used this factor to build an approach which automatically detects low priority incidents by analyzing the incident data and classifying it (Chen et al., 2020a, p. 378). The probability of an incident being low priority then shows how the incidents should be prioritized (Chen et al., 2020a, p. 378). Hence, engineers can manage incidents based on priorities, and the incident management process can be approved (Chen et al., 2020a, p. 378).

Another key step in incident management is incident triaging, which is the assignment of a new incident to the responsible team (Chen et al., 2019a, p. 364). This is a critical process, because if an incident is assigned to the wrong team, the incident mitigation could take longer, and more economic loss could be incurred (Chen et al., 2019a, p. 364). However, having an accurate and efficient triaging process is challenging. Especially in large-scale online service systems, since most incidents are automatically reported by monitors rather than users (Chen et al., 2019a, p. 364). This incident triaging is a continuous process and not done only once (Chen et al., 2019a, p. 365). Chen et al. (2019a, p. 365) found that in a setting of eight online services from Microsoft the percentage of incidents that are assigned at least twice ranges from 5.43% up to 68.26%. Furthermore, they found that on average up to 11.32 discussions are held before an incident is assigned correctly (Chen et al., 2019a, p. 365). This shows, that in practice, many incidents are assigned incorrectly, and significant effort is needed for discussion and then assigning the incident to the right team (Chen et al., 2019a, p. 365).

To address this problem, Chen et al. (2019a, pp. 365-374) proposed an automated approach for incident triaging, based on a Deep Learning model, which learns from human discussion. They were able to show that their approach enables a more accurate and efficient incident triaging process, which leads to less discussions needed to assign the incident to the right team (Chen et al., 2019a, p. 374).

Combining the process of diagnosing, triaging as well as forecasting Chen et al. (2019b, pp. 2659-2664) propose an intelligent outage management tool, which functions as a global watcher of the entire system. In their work, they collect alerting signals across the whole cloud system and use them to diagnose and predict system failures (Chen et al., 2019b, pp. 2659-2664). These outages come from two distinct levels: component- and service-level (Chen et al., 2019b, pp. 2659f.). Service-level outages consist of component-level outages; thus, these two levels are hierarchical (Chen et al., 2019b, p. 2660). The service-level outage prediction can assist in finding suspicious behavior in the overall system (Chen et al., 2019b, p. 2660). Additionally, determining which component is responsible for the outage can help reduce the cost of diagnosing and debugging (Chen et al., 2019b, p. 2660). Their approach identifies outages and examines dependence links between signals and outages (Chen et al., 2019b, p. 2660). Furthermore, they use predictive models to provide accurate outage forecasting and can correctly assign the predicted outages to the right team (Chen et al., 2019, p. 2663).

As shown, effective fault diagnosis is essential for minimizing downtime and financial loss. System incidents can be caused by several factors, including frequent upgrades, changes in the operating environment, and portability of equipment. Incident management systems and data-

driven failure diagnostics are key to managing these incidents. The above discussed studies have shown the importance of correlating time series and event data for fault diagnosis. Especially performance issues and disruptions caused by the failure of IT devices, such as hard drives, negatively impact a service. Several methods have been proposed to automatically diagnose performance problems and hard disk failures. It was shown that identifying correlations between incidents can help mitigate them and prevent future incidents. Furthermore, prioritizing incidents is critical to optimizing the incident management process, and approaches that use large amounts of labeled incident data to automatically detect and prioritize incidents have been proposed. Moreover, in incident management, triaging, i.e., assigning incidents to the appropriate team, is a critical process that saves valuable time if done correctly. The proposed approaches for incident triage based on deep learning models lead to a more accurate and effective incident triaging process.

2.4.2.2 Anomaly Detection

Another key aspect of IT operations, which is valued in AIOps, is anomaly detection (Wang et al., 2019, p. 94). Several studies were conducted in this field, showing its importance (Farshchi et al., 2018, pp. 531-547; Wang et al., 2019, pp. 94-103; Bagatinovski & Nedelkoski, 2021, pp. 1-12; Wu et al., 2021, pp. 1-11). The following paragraphs present these studies and their results.

The timely detection of anomalies allows preventing potential system failures and increases the time window for operators to react and solve the issue (Bagatinovski & Nedelkoski, 2021, p. 1). Thus, the system reliability can be improved by monitoring the ongoing operations in real time (Farshchi et al., 2018, p. 532). The metrics used to check the status of a service are time series data which measure the CPU and memory usage, the disk utilization as well as data and network throughput and service call latency (Wang et al., 2019, p. 94; Bagatinovski & Nedelkoski, 2021, p. 2). When an anomaly occurs, the corresponding KPI is likely to deviate from the expected pattern (Wang et al., 2019, p. 94).

A challenge in the monitoring of the system and the detection of anomalies is that system operators are faced with tracking multiple monitoring metrics and receiving too much monitoring information, which often includes false warnings and alarms (Farshchi et al., 2018, p. 532). Another difficulty is that system monitoring approaches often focus exclusively on point data for model training and detection; they monitor the state of hardware and software metrics, without monitoring the contextual behavior of operations or examining the impact of operation steps on system resources (Farshchi et al., 2018, p. 532; Wang et al., 2019, p. 95).

Farshchi et al. (2018, p. 532-547) developed a statistical approach, addressing the before named difficulties, through using the correlation between resources and operations' activities. They focused their work on monitoring and assuring dependability of DevOps application operations, also named "sporadic operations" in public cloud environments (Farshchi et al., 2018, p. 532). Sporadic operations are e.g., backup, upgrade, cloud migration, reconfiguration, on-demand scaling, rollback/undo, and deployment (Farshchi et al., 2018, p. 532). These are called sporadic operations because they mostly do not have a scheduled routine (Farshchi et al., 2018, p. 532). These sporadic operations commonly impact the whole system, and technological complexities make it difficult to ensure their successful execution (Farshchi et al., 2018, p. 532). The core of their approach is a domain-agnostic regression-based correlation analysis technique, which correlates event logs and resource metrics from operations (Farshchi et al., 2018, p. 547). Based on this, they can determine which monitoring metrics are affected by operational activities and in what way (Farshchi et al., 2018, p. 547).

The emphasis on single time points in anomaly detection has another issue; anomaly intervals are not detected (Wang et al., 2019, p. 95). One of these problematic approaches is change-point detection, which aims at the starting and ending point of an anomaly but does not consider the points in between (Wang et al., 2019, p. 95). The other approach is point-oriented and considers every point inside the anomaly interval as equal (Wang et al., 2019, p. 95). The issue with these two approaches is, that when an interval is labelled as anomalous, all points in this interval are considered anomalous (Wang et al., 2019, p. 95). Hence, the importance of each individual point and their difference is not being considered and therefore limits the generalization quality (Wang et al., 2019, pp. 95f.).

Wang et al. (2019, pp. 95-103) propose a supervised learning approach, which, in contrast to previous approaches, uses interval data for anomaly detection and analyzes the differences among detected anomalies. They were able to show that the accuracy of interval-oriented anomaly detection is higher than that of point-oriented anomaly detection (Wang et al., 2019, pp. 100-101). The approach of Wang et al. (2019, pp. 95-103) uses a Deep Neural Network (DNN), which is not limited to a specific KPI type. They were able to show that it outperforms other state of the art machine learning approaches for anomaly detection (Wang et al., 2019, pp. 100-101). They generated an adaptive an automated way of anomaly detection, which helps operators correctly identify anomaly intervals while reducing the false positive rate at the same time (Wang et al., 2019, p. 103).

Besides detecting anomalies in log data there is also the possibility of anomaly detection in distributed traces (Bagatinovski & Nedelkoski, 2021, p. 5). Distributed traces are a request-

centered method of describing behavior in a distributed system (Bagatinovski & Nedelkoski, 2021, p. 5). They follow the execution of the user's request through the distributed system in a record known as spans (Bagatinovski & Nedelkoski, 2021, p. 5). The spans represent information about the operations performed when managing an external service request (e.g., start time, end time, service name, HTTP path) (Bagatinovski & Nedelkoski, 2021, p. 5). Bagatinovski and Nedelkoski (2021, pp. 3-11) propose a multimodal anomaly detection approach, considering both logs and traces. They were able to show that their multimodal method delivers accurate results for the prediction of anomalous traces (Bagatinovski & Nedelkoski, 2021, p. 10).

Another vital task in AIOps is the prediction of Dynamic Random Access Memory (DRAM) failures (Wu et al., 2021, p. 1). DRAM is used as the main memory store in computer systems (Sridharan & Liberty, 2012, p. 1). Such DRAM failures (failures in memory) are one of the main causes for hardware failures in data centers (Wu et al., 2021, p. 1). These failures can cause severe outages, which lead to high economic costs and violate the service level agreements with the users (Wu et al., 2021, p. 1). However, reducing costs associated with hardware replacement and service disruption by accurately predicting DRAM failures is a challenging task (Wu et al., 2021, p. 2). In a data center, one single job could run on thousands of different nodes (Wu et al., 2021, p. 2). If a DRAM failure occurs at any of these nodes, and is not detected and resolved, the CPU can be wasted for a long time (Wu et al., 2021, p. 2). Although its importance, only few studies have tried to solve the problem of predicting DRAM failures (Wu et al., 2021, p. 2).

Wu et al. (2021, p. 6) state that for DRAM failure prediction, unsupervised anomaly detection models are more suitable than supervised ones. Because large-scale data centers may use DRAMs from various vendors, it is hard to guarantee that all installed DRAMs have labeled training data for supervised learning (Wu et al., 2021, p. 6). Furthermore, there can be unknown types of failures, which would not be detected by models trained with a supervised approach (Wu et al., 2021, p. 6).

Summarizing the previous paragraphs regarding anomaly detection it can be stated that early detection of anomalies can prevent system failures and give operators more time to fix problems, which increases system reliability. Traditional monitoring methods, however, ignore context and the impact of operations on system resources. Anomalies can be identified in different operating data. However, correctly identifying anomalies is challenging due to the vast amount of data that needs to be analyzed and the difficulty of using the right approach considering not only point data.

2.4.2.3 Incident Prediction

Several studies show that an important factor to ensure the reliability of a system is the prediction of incidents (Botezatu et al., 2016, pp. 40-48; Li et al., 2017, pp. 55-64; Mahidisoltani et al., 2017, pp. 391-402; Lin et al., 2018, pp. 481-489; Xu et al., 2018, pp. 482-491; Wang & Zhang, 2020, pp. 417-423; Zhao et al., 2020, pp. 315-326). The findings of these studies are presented in the upcoming paragraphs.

In an IT environment, hard disks are among the most frequently failing components (Botezatu et al., 2016, p. 39). While a single hard disk failure is uncommon, a system with thousands of hard disks is prone to failures and even simultaneous failures, resulting in service outages and potentially permanent data loss (Li et al., 2017, p. 55). As a result, one of the primary considerations of storage systems is reliability (Li et al., 2017, p. 55). In a storage system, the core components are its hard disks (Wang & Zhang, 2020, p. 417). The reliability of the entire storage system relies on the stable and reliable data access capability (Wang & Zhang, 2020, p. 417). Nowadays, a large-scale storage system consists of tens of thousands of disks, where temperature, duty cycles, and workloads can all have a significant impact on hard disk reliability and performance (Botezatu et al., 2016, p. 39; Wang & Zhang, 2020, p. 417). Thus, the chance of disk failure is apparent and reliability issues are the most severe, as disk failures necessitate replacements (Botezatu et al., 2016, p. 39; Wang & Zhang, 2020, p. 417).

With more disk failures the risk for the users also increases (Wang & Zhang, 2020, p. 417). First, there is the risk of losing data (Wang & Zhang, 2020, p. 417). If consecutive multiple failures occur, data will be lost. Second, the performance of the business deteriorates (Wang & Zhang, 2020, p. 417). Even if a storage system could reconstruct the lost data, the performance during this reconstruction process is lower and the duration of such a process usually exceeds one day (Wang & Zhang, 2020, p. 417).

Besides whole-disk failure, in which a disk stops functioning in a way that it needs to be replaced, another major threat to storage reliability are partial disk failures, in which individual sectors on a disk cannot be read (Mahidisoltani et al., 2017, p. 391). These partial failures can occur due to corrupted data, which cannot be corrected by disk-internal error correcting codes, or it can also happen due to mechanical damage on the disk surface (Mahidisoltani et al., 2017, p. 391). Both result in the problem that the disk cannot recover the data stored in the affected sector (Mahidisoltani et al., 2017, p. 391). Thus, disk errors can be defined as “gray faults” (Wang & Zhang, 2020, p. 418). That is, they are still hidden and difficult to detect when they have already seriously impacted a system, because they defy quickly from conventional system failure detection (Xu et al., 2018, p. 481; Wang & Zhang, 2020, p. 418).

According to Wang and Zhang (2020, p. 418), there are the following three approaches to solve the disk failure detection problem and ensure system reliability:

- RAID technology (Redundant Array of Independent Disks): RAID is a technique for enhancing data dependability that employs data redundancy, tolerates single or multiple disk failures, and recovers corrupted data using data encryption. It is a technology for passive fault tolerance.
- SMART: SMART data describes the disk's attributes. It is defined as the standard interface for disk management and is an active fault-tolerant technology. During disk operation, the technology monitors multiple parameters, including disk seek errors and sector errors. Active fault-tolerance based on SMART employs a threshold method, which is simple and straightforward, but the early warning accuracy is low.
- Machine Learning: AIOps technology is increasingly used in storage management. Machine learning methods are used to predict disk failure based on SMART data from a hard disk and system I/O error information, which can timely report the failure disk and improve storage system reliability.

The prediction of disk failures can significantly improve the availability and reliability of a system (Wang & Zhang, 2020, p. 418). If a disk failure can be predicted in advance, the risky disk can be replaced and the risks of losing data and performance can be reduced (Wang & Zhang, 2020, p. 418). However, the prediction of disk failures has its challenges (Botezatu et al., 2016, p. 40). SMART data are manufacturer specific (Botezatu et al., 2016, p. 40). That is, their encoding and normalization differ between manufacturers (Botezatu et al., 2016, p. 40). This fact makes the use of one predictive model for all different disks impossible (Botezatu et al., 2016, p. 40). Hence, a separate model for each disk manufacturer would need to be trained (Botezatu et al., 2016, p. 40). Furthermore, the implementation of SMART attributes is not standardized (Botezatu et al., 2016, p. 40). As a result, one must identify those that are indicative of failures (Botezatu et al., 2016, p. 40). Finally, the disk data is highly unbalanced (only about 2% of disks are replaced), making fitting high-quality models difficult (Botezatu et al., 2016, p. 40).

Botezatu et al. (2016, pp. 40-48) and Li et al. (2017, pp. 55-64) propose similar approaches for automatic forecasting of disk replacements, using SMART attributes. The goal of their approaches is to not only automate the disk replacement decision, but also allow administrators to replace risky disks proactively in advance (Botezatu et al., 2016, p. 40). Both approaches are based on a combination of decision and regression trees (Botezatu et al., 2016, p. 40; Li et al., 2017, p. 55).

However, both mentioned studies do not completely solve the problem of manufacturer specific SMART data. Mahidisoltani et al. (2017, pp. 393-401), on the other hand, discovered that random forest-based classifiers can accurately predict hard disk failures. According to the authors, these classifiers are simple to train and parameterize, and the training is robust even if the available training data is limited or the data comes from different disk models (Mahidisoltani et al., 2017, p. 401). Wang and Zhang (2020, pp. 419-423) were also able to show that their machine learning approach can accurately predict hard drive failures independent from different manufacturers. Their ensemble learning approach combines the advantages of multiple classification models and is robust and accurate even for unbalanced training sets and performs well on SMART data from different storage systems (Wang & Zhang, 2020, p. 422).

Besides these studies conducted to predict disk failures in storage systems, Xu et al. (2018, p. 481-491) studied the prediction of disk failures in cloud services. They state, that before a complete disk failure appears, the cloud service could already be affected by a disk error (e.g., latency errors, timeout errors, and sector errors) (Xu et al., 2018, p. 481). If these errors are not detected, the severity of the problem that could arise, is often high and could trigger service interruptions (Xu et al., 2018, p. 482). Hence, it is important to predict such disk errors, to take proactive actions before severe system damage occurs (Xu et al., 2018, p. 482). Proactive measures are e.g., error aware VM allocation (allocating VMs to healthier disks) or live VM migration (moving a VM from a faulty disk to a healthy one) (Xu et al., 2018, p. 482).

Xu et al. (2018, pp. 482-491) propose an approach to forecast such disk errors in order to improve the service availability in a cloud service system. Their approach uses SMART and system-level signals and utilizes machine learning to train a prediction model on historical data (Xu et al., 2018, p. 484f.). Their model can rank all disks to the degree of their error-proneness, to be able to allocate a VM to a healthier one (Xu et al., 2018, p. 488). Furthermore, it can identify a set of faulty disks from which the VMs should be live migrated out, subject to cost and capacity constraints (Xu et al., 2018, p. 488).

Using their approach on Microsoft Azure they were able to save around 63.000 minutes of VM downtime per month, by selecting healthier disks for VM allocation and live migration (Xu et al., 2018, p. 489). The 99.999% service availability, which is a common SLA for cloud services, means that only 26 seconds of VM downtime is allowed per month (Xu et al., 2018, p. 489). Hence, their work significantly improved the service availability of Microsoft Azure.

Other failures that can impact the service availability especially on cloud computing platforms are node failures (Li et al., 2020, p. 1). Cloud service systems contain many physical servers

also referred to as “nodes” (Lin et al., 2018, p. 481). The nodes are arranged into racks and a group of these racks form a cluster (Lin et al., 2018, p. 481). A key technology of cloud computing is virtualization, which means that one physical node can host multiple virtual machines (VMs), which offers improved scalability, maintainability, and reliability (Lin et al., 2018, p. 481). To best suit the end user’s needs, the VMs can be backed up, scaled up and down or duplicated (Lin et al., 2018, pp. 481f.). After a VM location request (i.e., request to be hosted on a node) is sent out, the system determines on which node the VM should be hosted (Lin et al., 2018, p. 482). In a node failing situation, all VMs hosted on that node will fail as well (Lin et al., 2018, p. 482). Live migration, the process of moving running VMs between different nodes, without interrupting the client or application connection is an important mechanism for cloud service management (Lin et al., 2018, p. 482). It enables rapid movement of workloads between clusters with minimal impact on the running service (Lin et al., 2018, p. 482).

According to Li et al. (2020, p. 3) node failures are exceedingly rare. They state that less than one node in thousands of nodes fail in a day (Li et al., 2020, p. 3). However, such failures can have a severe impact on the end users when they occur, because they can impact the availability of the hosted service (Li et al., 2020, p. 3). Therefore, it is crucial to predict node failures in advance, to enable DevOps engineers to minimize the impact by performing preventive actions (Li et al., 2020, p. 3). Predicting such failures is challenging, since the monitoring data is of huge size and the failure symptoms are often complex (Li et al., 2020, p. 3). By applications that scan through the monitoring data and automatically generate alerts, early warning signs for node failures can be detected (Li et al., 2020, p. 4). These alerts can be based on thresholds on the collected system behavior data (e.g., CPU is higher than the determined threshold) or error logs (e.g., a certain error type appears too frequently in a specific time range) (Li et al., 2020, p. 4). Based on such early warnings, preventative actions (e.g., virtual machine migration) can be taken in order to minimize the failure impact (Li et al., 2020, p. 4).

However, DevOps engineers face a significant challenge in the effective use of such alert data (Li et al., 2020, p. 4). On the one hand, these alerts contain helpful hints about possible node failures, as they were created by customized rules based on the engineers’ own expertise and years of experience from previous problems (Li et al., 2020, p. 4). These alerts, on the other hand, can be generated not only from nodes that will fail soon, but also from healthy nodes, because some of the issues reported by the alerts can be transient (e.g., network congestions or low memory availability) and recoverable (Li et al., 2020, p. 4). However, the extremely high volume and intensity of these alerts makes it difficult to manually differentiate between transient issues and issues that would result in node failures (Li et al., 2020, p. 4).

To improve service availability, Lin et al. (2018, pp. 481-489) propose an approach to predict the failure probability of a node before the failure occurs. Their approach uses machine learning techniques to learn the characteristics of historical failure data, and then uses the model to predict the probability that a node will fail soon (Lin et al., 2018, p. 481). However, building such an accurate prediction model is challenging, due to the three following main reasons identified by Lin et al. (2018, p. 481).

- Complicated failure causes: A large-scale cloud system is complex and thus node failures can be caused by several different software and hardware issues. For instance, software bugs can occur, the OS can crash, disks can fail, etc. Hence, there is no simple metric or rule that can predict all different node failures in a straightforward manner.
- Complex failure-indicating signals: Identification of a node failure can be done by many different temporal signals produced by the node itself. But also, by spatial properties that are shared by nodes which are dependent on each other. It is therefore needed to analyze temporal signals as well as spatial properties to best capture signals that early indicate a failure.
- Highly imbalanced data: The data of failing nodes is highly imbalanced. The ratio between failing and healthy nodes is often less than 1:1000, meaning that less than 0.1% of nodes contain failures. This imbalance challenges an accurate prediction.

Various nodes could fail at various times. Lin et al. (2018, p. 482) suggest predicting the propensity of a node to fail based on the analysis of previous fault data, prior to the node failing. Predicting node failures can enhance service availability in two ways (Lin et al., 2018, p. 482):

- VM allocation, which is the allocation of a VM to a node. To facilitate better VM allocation, we can always allocate VMs to a healthier node as opposed to a faulty node.
- Live migration, which is the process of transferring a running VM between various nodes without disconnecting the client or application. To enhance the live migration of nodes, we can shift VMs from predicted defective nodes to healthy nodes before actual node failure occurs.

To accomplish this, a prediction model is constructed, which utilizes previous failure data and machine learning techniques (Lin et al., 2018, p. 482). This model can then be used, to forecast the likelihood of a node failing soon. This prediction model must be able to rate all nodes according to their vulnerability to failure, allowing the service systems to assign a virtual machine to the healthiest node available (Lin et al., 2018, p. 482). Furthermore, it should be capable of

identifying a collection of defective nodes from which hosted VMs should be relocated, subject to cost and capacity constraints (Lin et al., 2018, p. 482).

The node failure prediction approach proposed by Lin et al. (2018, pp. 483-489) based on two classification models can successfully predict node failures, and thus allocate and migrate VMs to healthier nodes. As a result, these VMs are less likely to be affected by node failures (Lin et al., 2018. p. 489).

Li et al., (2020, p. 6-20) present a random forest-based model, which allows DevOps engineers to identify the importance of each of the used features in model training. When their solution predicts a node failure, DevOps engineers can take preventative measures (e.g., live migration) to reduce the impact of node failures on the production system. Furthermore, their solution integrates self-healing strategies to automatically mitigate the effects of node failures and thus improve node resiliency (e.g., by automated control of the node repairing process and automated live migration) (Li et al., 2020, p. 22).

As previously stated, there are several other components in a cloud service system that can cause incidents (Chen et al., 2019b, p. 2659). Zhao et al. (2020, pp. 317-325) propose an approach that predicts general incidents in real time using light weight alert data. They state that predicting incidents based on alert data is challenging because of the many attributes included, the noise in the alert data, and the needed interpretability of the prediction model (Zhao et al., 2020, p. 316). Furthermore, to enable engineers to immediately take actions to prevent incidents, time efficiency is crucial (Zhao et al., 2020, p. 323). If the prediction result of an incoming alert cannot be provided timely, the incident may not be prevented (Zhao et al., 2020, p. 323). Zhao et al. (2020, p. 323) were able to successfully apply their approach to two commercial banks, which state that the solution assists them to anticipate incidents in advance and proactive actions can be taken to prevent the service unavailability.

As shown, incident prediction focuses especially on disk and node errors and failures. Both predictions are challenging since the monitoring data of disks and nodes are highly imbalanced and contain a lot of noise. The studies show that accurate incident prediction can have a significant impact on service downtime reduction. Through proactive measures that can be taken (e.g., VM allocation and live migration to healthier nodes) service failures can be prevented.

2.4.2.4 Automated Resource Allocation

Although cloud computing technologies have advanced significantly, there are still unanswered questions around resource allocation (Liu et al., 2021, p. 1). Low resource utilization has been a significant concern for both cloud service providers and cloud customers (Liu et al., 2021, p. 1). Hence, another application for AIOps is automated resource allocation (Chen et al., 2021, p. 1). Resource allocation is a technique that facilitates the allocation of virtualized resources to users (Liu et al., 2021, p. 2). When a user requests a resource or application, a required number of containers is generated and assigned to the user (Liu et al., 2021, p. 2). A crucial aspect to automate this, is the forecasting of time series data (Chen et al., 2021, p. 1). The forecasting aims to anticipate future operational equipment conditions and thus allocate required resources in advance (Chen et al., 2021, p. 1). To truly realize AIOps, the system must be capable of accurately predicting the collected time series data (Chen et al., 2021, p. 1). Only then will the system be able to better understand future operation status and automatically advance the best deployments (Chen et al., 2021, p. 1).

Chen et al. (2021, p. 2) state four challenges in predicting time series data from operations. First, operations data often contain a variety of hidden variables, which cannot be directly observed (Chen et al., 2021, p. 2). Thus, if these hidden features can be extracted and analyzed, it will improve the prediction accuracy (Chen et al., 2021, p. 2). Second, operations data is available in a real-time data stream and its volume increases over time. To be able to realize automatic control, it is vital that real-time predictions can be made (Chen et al., 2021, p. 2). Third, data processing needs to be fast, due to the rapidity and continuity of the data collection. It therefore is also required to be able to quickly deal with failures (Chen et al., 2021, p. 2). Fourth, the state of each device in operations is strongly related to historical data as well as current data (Chen et al., 2021, p. 2). As a result, if only current data is used for statistical analysis, the information carried by historical data is lost (Chen et al., 2021, p. 2).

To overcome these challenges, Chen et al. (2021, pp. 4-10) present a data decomposition algorithm that decomposes the operations data into various components and thus replaces the problematic hidden variables. It then utilizes a neural network to predict each component, and combines the prediction results, to be able to predict the future operating conditions (Chen et al., 2021, p. 10).

2.4.3 Goals of AIOps

As previously shown, AIOps has different application areas that have been studied. The shown applications are associated with specific goals that companies are trying to achieve by implementing an AIOps solution. The upcoming section aims to show those different goals.

2.4.3.1 High Service Intelligence

One goal that companies try to achieve by using AIOps solutions is high service intelligence and thus reducing the MTTD (Mean Time to Detect) and MTTR (Mean Time to Repair) (Dang et al., 2019, p. 4; Shen et al., 2020, p. 276). In a service driven by AIOps, the goal is to be aware of changes in a timely manner, such as quality degradation, cost increase, workload spike, etc. (Dang et al., 2019, p. 4). Detecting data update errors, for instance, which lead to a pattern and identify a bad network connection would take weeks for a human to diagnose (Lithicum, 2020, p. 3). An AIOps solution may forecast future conditions of a service based on its past behaviors, workload patterns, and underlying infrastructure operations (Dang et al., 2019, p. 4). Such self-awareness and predictability cause a service to engage in self-adaptation or self-healing with minimal human interaction (Dang et al., 2019, p. 4). For instance, the AIOps solution can take pre-defined corrective actions automatically, e.g., restarting a server or disconnecting from a bad network device, which relives the On-Call Engineers (OCEs) of dealing with the issue (Lithicum, 2020, p. 9).

Compared to a traditional IT Operations process, where OCEs would first need to check the troubleshooting guide, assuming the incident already occurred previously, to find a solution for the incident, the AIOps solution saves valuable time (Jiang et al., 2020, p. 1411). Identifying and fixing the root cause (e.g., code defects) without AIOps, then testing and re-deploying usually causes much delay before the service can be recovered (Lou et al., 2017, p. 906). Shen et al. (2020, p. 276) state, that with the assistance of AIOps the MTTD can be decreased from ten minutes to one minute, and the MTTR can be reduced from sixty minutes to thirty seconds, compared to traditional IT Operations. This has a positive impact on the service quality, the costs of running and maintaining the service, and the service performance (Shen et al., 2020, p. 276).

2.4.3.2 High External Satisfaction

The second goal an AIOps implementation aims to achieve is high customer satisfaction (external) (Dang et al., 2019, p. 4). It is associated to the goal of high service intelligence, since an intelligent service can analyze consumer usage patterns and take preventative measures to increase customer happiness (Dang et al., 2019, p. 4). In traditional IT Operations, an administrator usually detects an interrupted service, when it already has an impact on the user (Gulenko et

al., 2020, p. 1). Resolving the issue then takes valuable time, and the user could be affected by an interrupted or not correctly working service (Gulenko et al., 2020, p. 1). By integrating monitoring and service data across systems, AIOps aims to provide end-to-end visibility into the customer experience (McKeon-White et al., 2021, p. 2). This provides organizations with the ability to zero in on root causes and detect problems before they disrupt a service (McKeon-White et al., 2021, p. 2). For instance, a service may automatically recommend tuning suggestions to obtain optimal performance (e.g., adjusting configuration, redundancy level, and resource allocations); a service may also recognize that a customer is experiencing a service quality issue and proactively engage with the customer to provide a solution or workaround, as opposed to reactively responding to customer complaints through human support (Dang et al., 2019, p. 4).

2.4.3.3 High Internal Satisfaction and Productivity

Automation enabled by AIOps also assists IT and business teams with their work and thus aims to increase internal satisfaction and productivity (McKeon-White et al., 2021, p. 2). Its functionalities benefit IT operations employees, service desk staff, DevOps teams, site reliability engineers (SREs) and business leaders (Prasad et al., 2022, pp. 13-14). For instance, service desk employees are often waiting for assistance from operations or repeating the same task for an endless stream of users dozens of times per week (McKeon-White et al., 2021, p. 2). AIOps aims to automate these high-volume, low-complexity tasks via capabilities such as advanced machine learning (ML), root-cause detection, and event correlation (McKeon-White et al., 2021, p. 2). These capabilities should allow the operations teams to analyze the huge amount of data, a task that cannot be done manually (Lyu et al., 2022, p. 2; Prasad et al., 2022, p. 14). Thus, they can learn patterns of system behavior and forecast future system behavior to amend service adaption strategies (Dang et al., 2019, p. 4). For the service desk staff, these automations and forecasts could improve the response time and free up time to address more complex issues (McKeon-White et al., 2021, p. 2). For DevOps teams, on the other hand, the goal is to assist them by allowing them to ingest traces, metrics, and log data, which eases the effort to get an overview over platforms and products and their KPIs (Prasad et al., 2022, p. 14). Also, SREs may leverage analytics functions to enhance service resilience through evaluating multiple IT infrastructure scenarios (Prasad et al., 2022, p. 14). In IT operations without AIOps, an SRE would observe the infrastructure, then evaluate and interpret indicators, investigate the system's response, and then decide on a tuning or recovery action (Gulenko et al., 2020, p. 4). Finally, also the business should benefit from insights into qualitative KPIs such as the efficiency and productivity of technology, people, and existing processes (Prasad et al., 2022, p. 14).

2.4.4 Capabilities of AIOps Solutions

Lithicum (2020, p. 8) states that AIOps tools often evolved from standard ops tools. Through managing and monitoring cloud, multi-cloud, legacy, and even IoT and edge-based systems, AIOps tools try to bridge the gap between on-premises legacy system management and cloud service management (Lithicum, 2020, p. 8). After looking at the different use cases and goals of AIOps, the next section aims to evaluate the needed capabilities of AIOps solutions to be able to solve these use cases and to realize the stated goals.

The most notable change provided by AIOps is a shift in focus from reactive engagements and analytics to proactive, predictive, prescriptive, and preventive engagements (Humphrey, 2020, p. 6). According to Lithicum (2020, p. 9), AIOps solutions should have different functionalities, including the predictive spotting of system failures, self-healing of components, connection to remote components, creating customized views that promote productivity and monitoring, managing, as well as repairing standard infrastructure concepts. Shen et al. (2020, p. 277) define these functionalities as the five typical abilities of an AIOps system, named “Perception”, “Detection”, “Location”, “Action”, and “Interaction”. These five abilities are considered in detail in the following.

Perception. The foundation to enable all AIOps functionalities is sensory data (Casanova et al., 2021, p. 4). Without the ability to access the raw data emerging from all the physical, logical, and conceptual elements of the environment, there is no AIOps functionality (Casanova et al., 2021, p. 4). Without this data foundation, no conclusions can be drawn, no analytics can be utilized for action, and no patterns can be identified for intervention (Casanova et al., 2021, p. 4). The quality and completeness of this data foundation has an impact on all other aspects of AIOps (Casanova et al., 2021, p. 4).

On top of the foundation lies the ingestion capability to filter and preprocess data (Casanova et al., 2021, p. 4). It is necessary to manipulate data from diverse sources with varying levels of granularity (e.g., infrastructure, networks, applications, the cloud, and existing monitoring tools) and in various formats so that it can be processed efficiently and effectively (Casanova et al., 2021, p. 4; Prasad et al., 2022, p. 3). Along with other data transformations, deduplication efforts must support both traditional extract, transform, and load (ETL) processing and stream-based, real-time, or near-real-time processing (Casanova et al., 2021, p. 4).

Detection. The next part is to make sense of the data to be able to detect anomalies (Shen et al., 2020, p. 277; Casanova et al., 2021, p. 4). Here, the fundamental transformation from element-level data to actionable intelligence takes place (Casanova et al., 2021, p. 4). The ability to as-

semble disparate data elements and add context to the resulting story is essential for making the data actionable (Casanova et al., 2021, p. 4). The goal is to reduce the noise generated by IT environments and to present IT professionals with actionable next steps (Casanova et al., 2021, p. 4).

Location and Action. The insights must then be turned into action (Casanova et al., 2021, p. 4). After the data is transformed from its raw state, it must be prepared for use (Casanova et al., 2021, p. 4). To do that, the AIOps platform must be able to leverage machine learning (ML) and data analytics including real time analysis at the point of ingestion (streaming analytics) and historical analysis of stored data (Prasad et al., 2022, p. 3). The AIOps solution automatically correlates and compresses events across domains or sources, reducing the need for manual intervention (Prasad et al., 2022, p. 3). The correlation combines time and topology to group together events that are related (Prasad et al., 2022, p. 3). Furthermore, the AIOps solution analyses event and telemetry data to predict significant incidents or events and to locate their root causes (Shen et al., 2020, p. 277; Prasad et al., 2022, p. 3). Historical data is used as a starting point and the solution then continuously learns and refines important event patterns based on historical data, real-time streaming data, operator input, and reinforcement mechanisms (Prasad et al., 2022, p. 3).

Interaction. A portion of this processed data is then forwarded to the traditional IT action processes via dashboards, alerts, and a variety of other distribution and notification mechanisms (Casanova et al., 2021, p. 4). Additional data can then enable the AIOps solution to provide advice, automate a response, or activate an external automation system, without human intervention (Casanova et al., 2021, p. 4; Prasad et al., 2022, p. 3). Here the interaction between user and system can be simplified by e.g., facial recognition, gesture recognition, and Natural Language Processing (NLP) (Shen et al., 2020, p. 277).

Finally, there are methodologies for ensuring operational dependability (Casanova et al., 2021, p. 4). Organizations' operational environments are based on the processes and procedures that keep them running (Casanova et al., 2021, p. 4). They rely on data and knowledge to deliver the agreed-upon level of service quality (Casanova et al., 2021, p. 4). Utilizing the processed and prepared data improves the human judgment of all practices (Casanova et al., 2021, p. 4). This context, along with insights powered by AI/ML, elevates and modernizes the practice's capabilities (Casanova et al., 2021, p. 4). Through either explicit operator specification or observation, the AIOps solution continuously learns and improves associations between each significant event and the operations response (Prasad et al., 2022, p. 3). Figure 1 shows an AIOps platform enabling continuous insights across IT operations.

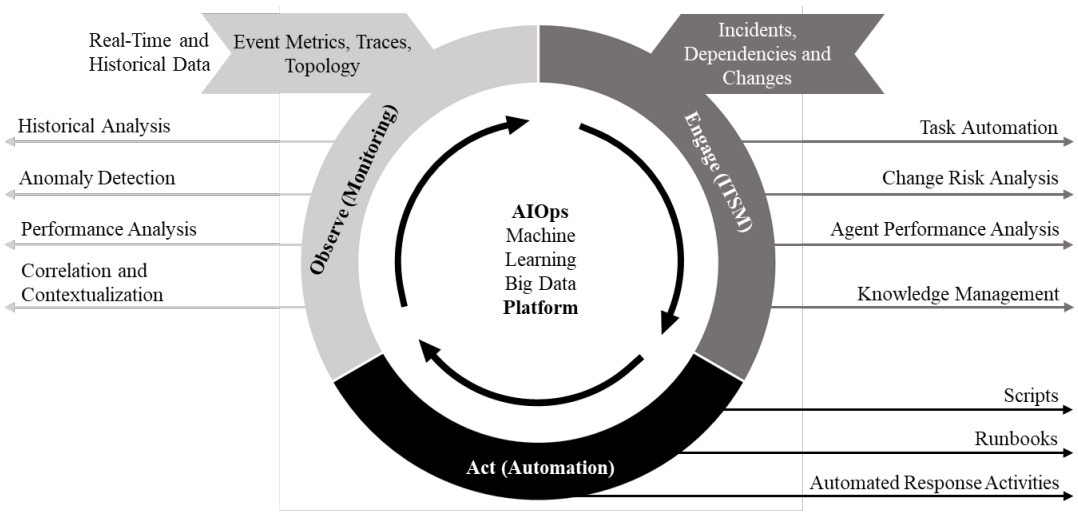

Figure 1: AIOps Platform Capabilities (Prasad et al., 2022, p. 4)

To be able to provide the shown functionalities, an AIOps platform uses the following analytical approaches (Prasad et al., 2022, pp. 8-9):

- *Statistical analysis*. A mix of univariate and multivariate analytic techniques such as correlation, clustering, classification, and extrapolation are used.
- *Automated discovery and prediction of patterns*. Patterns, clusters, or groups of correlations discovered in historical and/or streaming data. These patterns can be utilized to predict anomalies with varying probabilities.
- *Anomaly detection*. Using the patterns discovered by the previous components to determine normal behavior and then to identify univariate and multivariate deviations from that normal behavior. Beyond the simple detection of outliers, they must be correlated with business impact and other concurrent processes, such as change management, to not only generate more alert noise.
- *Probable cause determination***.** Reducing the network of correlations generated by automated pattern discovery and data ingestion to establish causality chains connecting cause and effect.
- *Topological evaluation***.** Providing contextualized analysis based on application, network, infrastructure, or other topologies. Patterns derived from data within a topology will establish relevance and reveal concealed dependencies.
- *Adaptive prescriptive advice***.** Providing potential resolutions to an issue. These recommendations may be derived from a database of past solutions (institutional knowledge) to recurrent issues, or they may be determined through crowdsourcing. Over time, the tool should determine the best solution among multiple possibilities.

Besides these functional capabilities, Li et al. (2020, p. 1) and Prasad et al. (2022, p. 7) emphasize that AIOps solutions must be trustable, interpretable, maintainable, and scalable. Trustable means that an AIOps solution, instead of just employing sophisticated models on raw data, must incorporate years of field-tested, engineer-verified domain expertise into their ML models (Li et al., 2020, p. 2). The interpretability of the used ML models allows its users to reason about model recommendations, to gain management support for implementing such recommendations, and to improve the current status (e.g., through optimizing the used monitoring solutions) (Li et al., 2020, p. 2). Maintainability comes down to the requirement of minimal maintenance and fine-tuning needs, as the users of AIOps tools (e.g., DevOps engineers) are typically not ML experts (Li et al., 2020, p. 2). Scalability means that the AIOps solution must be scalable and efficient, since it must be able to analyze vast amounts of data (Li et al., 2020, p. 2).

2.5 Concluding the Multivocal Literature Review

The conducted multivocal literature review highlights that the increasing complexity of IT operations due to the advent of cloud computing and infrastructure virtualization has necessitated the development of Artificial Intelligence for IT Operations (AIOps). The different systems generate massive amounts of data and alerts, unmanageable without the support of technology. To be able to provide a high Quality of Service (QoS), reduce service downtimes and the associated financial losses, despite the increasing system complexity, companies started using AI to support their IT Operations. AIOps aims to help organizations achieving high service intelligence as well as internal and external satisfaction by reducing MTTD and MTTR and enabling self-adaptation and self-healing.

One AIOps application area is incident management. It is critical to maintain system availability, and research has focused on incident diagnosis, linkage, prioritization, and classification. The presented studies focused on different approaches that aim to automate incident management. As performance issues and system failures can negatively impact user satisfaction and project success, it is essential that such issues are mitigated rapidly. Linking and mapping incidents streamlines the incident management process, reduces time to resolution, and ensures that incidents are treated with the right priority. Also there, several approaches have been researched that improve the accuracy and efficiency of incident triaging and prioritization.

Furthermore, anomaly detection is essential for preventing system failures and improving reliability, with various methods such as statistical, supervised, and multimodal approaches providing accurate results. Moreover, predicting failures is an essential AIOps task. The focus there lies on disk and node failures and various machine learning techniques have been used to improve service availability. Additionally, automated resource allocation is a major concern in

cloud computing, and AIOps aims to predict future operational states of assets by predicting time series data.

The literature shows that there is no future of IT Operations without AIOps. The presented application areas, models, and capabilities have implications on the later presented case study. It focuses on a provider and its implementation partners of AIOps solutions in order to show the real-world adoption of AIOps. The findings will enable a recommended course of action for implementing an AIOps solution for mission-critical applications to address today's IT operations challenges, which is the purpose of this empirical study. Furthermore, it allows to validate the practical applicability of the shown approaches.

3 Theoretical Background

After reviewing the AIOps-related literature, this section provides an overview over the topic of business AI alignment. First a brief introduction into business IT alignment is given, to then show the importance of creating AI plans collaboratively with business strategies to create value from AI. This theory forms the basis of the developed and later presented business AIOps alignment model.

3.1 Business IT Alignment

Reich and Benbasat (1996, p. 82) defined alignment as "the degree to which the information technology mission, objectives, and plans support and are supported by the business mission, objectives, and plans". IT has transformed how businesses operate (Njanka et al., 2020, p. 334). It has an impact on business processes, how companies deliver services, and how they communicate with customers, suppliers, and employees (Njanka et al., 2020, p. 334). The first explanation of the interrelationship between business and IT was presented by Henderson and Venkatraman (1993) with the Strategic Alignment Model (SAM). The SAM has two dimensions: The "strategic fit" and the "functional" integration (Henderson & Venkatraman, 1993, p. 476). Strategic fit refers to the ability of IT to shape and support the business strategy (Henderson & Venkatraman, 1993, p. 476). Functional integration, on the other hand, refers to the link between organizational infrastructure and processes and IT infrastructure and processes (Henderson & Venkatraman, 1993, p. 476). Henderson and Venkatraman (1993, p. 481) state that both dimensions must be considered to achieve alignment of business strategy and IT infrastructure.

De Haes et al. (2020, p. 86) identify three main dimensions of business IT alignment, namely the "strategic dimension", the "operational dimension" and the "individual dimension". Chan and Reich (2007, p. 4) state that strategic alignment refers to the extent to which business strat-

egy and IT strategy complement each other. The operational dimension deals with the implementation of the strategy in day-to-day business (De Haes et al., 2020, p. 87). The individual dimension focuses on the match between the IT infrastructure and the user needs (De Haes et al., 2020, p. 87).

3.2 Business AI Alignment to Create Value

Artificial intelligence (AI) can be seen as a subset of IT (Stecher et al., 2020, p. 2). In contrast to conventional IT, AI emphasizes the capacity of IT systems to learn from experience and act independently (Stecher et al., 2020, p. 2). Adopting AI technology becomes increasingly important for companies as the technology matures (Brynjolfsson et al., 2019, pp. 50-51). AI technology has evolved to the point that it provides real-world commercial benefits and has a large economic impact in several industries (Stecher et al., 2020, p. 2). As Stecher et al. (2020, p. 2) state, AI is more than a minor technological development. A successful AI strategy has the potential to affect all layers of an organization, including its business strategy, organizational structures, processes, workforce, data and information system architecture, and technical infrastructure (Stecher et al., 2020, p. 2). However, there is often a lack of strategic understanding of where AI can be used and for what strategic purpose AI must be aligned (Engel et al., 2022, p. 1). Since AI can be seen as a subpart of IT, the alignment of AI and business strategy is also considered a subpart of business IT alignment. According to Stecher et al. (2020, p. 2), business AI alignment refers to "the extent to which organizations' AI and business units share a common understanding of its strategic goals and contribute towards achieving these goals."

Stecher et al. (2020, p. 13) show that business AI alignment is a crucial prerequisite for successful AI adoption. It ensures that an organization's AI plans reflect business plan objectives and support business strategies (Stecher et al., 2020, p. 13). On the other hand, the business should also have reasonable AI expectations which refer to AI plans, technologies, and techniques (Stecher et al., 2020, p. 13). A lack of strategic alignment can result in the loss of synergy effects between existing and new AI use cases (Engel et al., 2022, p. 1). One challenge is to identify existing or design new AI use cases that meet the organization's business needs and generate the desired business value contribution (Engel et al., 2022, p. 1). A poorly planned application of an AI use case can not only fail to generate value, but also result in the destruction of value (Engel et al., 2022, p. 1).

Creating value with AI, however, has its challenges. Alsheibani et al. (2020) identified six challenges when trying to create value with AI.

1. *AI business case*. There must be a solid business case that has to be aligned with the current business strategies. A solid business case specifies what an AI technology will do and

demonstrates how its algorithms will improve the execution and outcomes of a business process or set of business processes.

2. *Relative benefits of AI.* The relative benefits of AI must be clear. Understanding what AI does and its performance benefits is often not difficult; deciding whether to use AI principles, however, is more complicated. Early observations of AI implementation show that such technologies can deliver no results or uncertain or unpredictable ones, raising new challenges and questions about the long-term effects of AI investments in organizations.
3. *Top management support*. AI technology must be aligned with business and AI owners. Gaining advantages from AI innovation necessitates not just the organization-wide adoption of these innovations, but also the commitment and involvement of senior management.
4. *Effective use of data*. A clear strategy must be developed for obtaining the data required for AI operations. The data must contain both quantity and structure for AI systems to work.
5. *AI talent*. An organization needs AI skills and must therefore invest in the right workforce capable of working with AI solutions.
6. *AI compatibility*. AI adoption needs not only the technical skills to build an AI algorithm, but also domain experts who understand the activities, workflows, and considerations of present business processes and can evaluate if AI programs can improve them.

As shown, aligning business and AI includes several factors and challenges that need to be considered when starting to use AI in business. The challenges identified by Alsheibani et al. (2020) are taken to evaluate them in a real-world context to see if they also apply to the field of AIOps and aligning those initiatives to the business goals.

4 Research Method

The upcoming section describes the chosen research methodology to examine the two research questions. First, the concept of Single-Case Study Research is stated. Second, the selected case is described, and its choice justified. Third, the data collection and data analysis processes are shown.

4.1 Single-Case Study Research

The case study is an approach for conducting research that focuses on gaining knowledge of the dynamics present in single contexts (Eisenhardt, 1989, p. 534). As Yin (2018, p. 33) states, case study research is relevant, if one wants to explain something, e.g., with questions “how” or “why”. Researchers should use case study research if they want to understand a real-world case and assume that such understanding is likely to include important contextual conditions relevant to the case (Yin, 2018, pp. 45-46).

This single-case study aims to examine the implementation of AIOps solutions developed by a leading company in the field of AIOps. To also get non-vendor specific insights, two third-party companies were also chosen, which implement the provider's AIOps solutions. The study will focus on the experiences of experts from the provider and the implementation partners, who have worked on multiple AIOps projects and can provide valuable insights into the benefits, challenges, and best practices for implementing and using AIOps solutions. The findings derived from the expert interviews are then evaluated against the conducted multivocal literature review.

4.2 Case Selection

To assess which companies were eligible for the case study from provider and implementation partner perspective, the following filter criteria were used: (1) The provider had to be named a leader in the Gartner magic quadrant for AIOps solutions. (2) The provider had to be accessible by the researcher, thus allowing access to company-specific information and expert knowledge that would otherwise have been difficult to access. (3) The implementation partners had to have experience with implementing AIOps and related solutions at their clients, thus being able to provide insights into challenges and best practices.

The provider and implementation partner are not mentioned by name in the following to prevent conclusions from being drawn about the companies or persons within them. Thus, supplier and implementation partner are named as such. Table 5 shows an overview of the supplier and implementation partners including type, number of employees, revenue and headquarter country.

ID	Company facts (2022)
Supplier	• IT and Consulting Provider • ~ 280.000 employees • ~ 57 billion $ revenue • USA
Implementation partner A	• Service Provider • ~ 1.000 employees • ~ 150 million € revenue • Germany
Implementation Partner B	• Service Provider • ~ 20 employees • ~ 5 million CHF revenue • Switzerland

Table 4: Overview of Cases and Companies

Out of these companies ten experts have been interviewed who work in the field of AIOps. Details on their positions and experience are stated in Appendix B.

4.3 Data Collection

The data collection is based on qualitative interviews. According to Myers and Newman (2007, p. 2), qualitative interviews are used in all kinds of qualitative research including case studies, action research, grounded theory, and ethnographies. The conducted interviews are based on a semi-structured interview guide following the methodology of Adams (2015, pp. 495-502). The interview guide in a semi-structured interview is incomplete, thus the researcher needs to improvise during the interview (Myers & Newman, 2007, p. 4). Semi-structured interviews are conducted with one respondent at a time, and the questions in the script are followed-up by "why" and "how" questions (Adams, 2015, p. 493). Thus, the interviewees are not limited in their answer possibilities and important insights, which may not be covered by the questions in the guide, will not be neglected.

The interviews were planned to last one hour and were conducted between January and April 2023 following the guidelines of Myers and Newman (2007, pp. 15-17). Their guidelines are based on the general theory of Goffman (1959), where social interactions are seen as a drama with actors that perform on a stage using a script. Myers and Newman (2007, pp. 11-17) adapted this theory and developed a seven-step guideline for conducting qualitative interviews which is described below. The interview can be seen as a stage and the interview guide as a script.

1. *Situating the researcher as actor*. In a first step the researcher and interviewee should situate themselves. Useful questions can be: What is your role? What is your background and your experience? This information allows the readers assessing the validity of the findings. In the conducted interviews every interview started with an introduction of the interviewer and the research topic. The first questions asked always aimed to understand the position and experience of the interviewee and his or her view on the research topic.
2. *Minimize social dissonance*. This is referred to as minimizing anything that could make the interviewee feel uncomfortable. Typically, this entails attempting to manage first impressions, dressing appropriately, using proper language and to act based on the interview situation (e.g., being aware of differences between interviewing a CEO and a shop floor worker). In this study, language and dress code were chosen according to the type of organization in which the interviewees work. Since the interviewees were all at similar levels in the organization, the interviewer did not have to act differently in the various interviews.
3. *Represent various voices*. In qualitative research it is suggested that within the organization a variety of people should be interviewed because not all interviewees are the same. This is also referred to as "triangulation of subjects" (Rubin & Rubin, 2005, p. 67), with the idea to not have one voice that emerges and thus trying to avoid elite bias (Miles & Huberman, 1994). To represent various voices in this research, interviewees from different companies (provider and implementation partner), departments (development, consulting, IT) and positions (solution architects, engineers, consultants, C-level) were chosen. This allowed the researcher to get a holistic view and avoid elite bias. The decision, which personas should be interviewed, was based on a "key informant" methodology, where the interviewees are selected on the assumption that they are knowledgeable about the research topic and willing to discuss it (Kumar et al., 1993, p. 1634). Before conducting the interviews, the personas were contacted to identify if they can be selected as "key informants" in terms of experience, current role in the company, and availability during the research period.
4. *Everyone is an interpreter*. It must be considered that the interviewees are creative interpreters of their field, and the interviewer is as well. Interviews are mostly rare and artificial events for the interviewees, and lead to the creation of one or more texts, with the interview transcript forming the basis. Using the transcripts of the interviews conducted, the researcher was able to analyze the interviews and derive findings, which were then reproduced as continuous text representing the interpretations of the interviewees.
5. *Use mirroring in questions and answers*. The term "mirroring" refers to the practice of using the same words and phrases that respondents use, to formulate subsequent questions or comments. This allows the researcher to focus on the respondent's world and utilize their language, as opposed to imposing the researcher's language. The aim is that the respondents

explain and describe their field in their own words. It is recommended to use open questions and to go from general to more specific questions. In the conducted interviews, if subsequent questions were asked, the interviewer used the same words and phrases as the respondents used. The interview guide was organized in such a way that the topic was first presented in general terms (e.g., the interviewees view on AIOps and the differences between AIOps and traditional IT operations), and then specific subtopics (e.g., benefits, challenges, limitations, and business AI alignment) were discussed in greater depth during the interview.

6. *Flexibility*. Since semi-structured interviews use an incomplete script, they require flexibility, improvisation, and openness. Thus, the interviewer should be prepared to investigate intriguing research avenues and be on the lookout for surprises. In addition, it is necessary to consider the diverse attitudes of respondents and to respond accordingly. Most respondents in the conducted interviews provided detailed answers on most of the asked questions. In interviews, where the answers given were less detailed, more sub-questions had to be asked.
7. *Confidentiality of disclosures*. The records and transcript of the interviews need to be kept confidential and secure. In the conducted interviews, all respondents agreed on recording and transcribing the interviews. Personal and company-related information (e.g., names of individuals and solutions) was anonymized so that no conclusions could be drawn about the organization and individuals.

Myers and Newman (2007, p. 24) state several benefits of using their methodology. One important benefit is that the researcher is sensitized to the complexity of the interview. Furthermore, the difficulties of interviews are explored to reduce potential problems and pitfalls and ensure good performance. In addition, it helps the interviewer to know how to minimize social dissonance and it shows the importance of the interviewer's flexibility and improvisation skills when using incomplete scripts. It also reveals the importance of leading questions and mirroring on the respondents' perception and shows how the respondents' words and phrases can be used more effectively. Ultimately, it greatly improves the potential for greater disclosure, which in turn leads to data being collected in greater quantity and quality (Myers & Newman, 2007, p. 24).

4.4 Qualitative Content Analysis

Once data collection is complete, the next step is to analyze the data to derive findings. Qualitative content analysis is defined by the fact that the entire data is coded, i.e., systematically processed based on a category system (Kuckartz & Rädiker, 2022, p. 71). Categories can be created deductive (mostly independent of the data collected) or inductive (based on the collected data) (Kuckartz & Rädiker, 2022, pp. 71,82). However, fully deductive, or inductive procedures are rarely found in research projects (Kuckartz & Rädiker, 2022, p. 129). In most qualitative research projects, a multi-stage category development and coding technique is utilized (Kuckartz & Rädiker, 2022, p. 129). The initial coding phase is done along major categories with a manageable number of codes (deductive) (Kuckartz & Rädiker, 2022, p. 129). In the subsequent step the categories are refined and differentiated on the data (inductive) (Kuckartz & Rädiker, 2022, p. 129). The coded data will then be analyzed by category and the results will be compiled in preparation for writing the study report (Kuckartz & Rädiker, 2022, p. 129).

4.4.1 Coding the Data

Coding can be used for several types of data, e.g., interview transcripts, journals, field notes, but also for photographs, videos etc. (Saldaña, 2013, p. 3). In this work the coding is done in two cycles, in which the first cycle aims to initially summarize the data, and the second groups those summaries into a smaller number of categories (Saldaña, 2013, p. 58). In the first cycle the collected interview data is coded following a provisional coding approach (Miles et al., 2014, p. 83). Miles et al. (2014, p. 83) note that this coding approach is appropriate for qualitative studies that build on prior research, which is the case in this Master Thesis as it is based on a multivocal literature review. Starting with a list of previously generated codes based on the multivocal literature review, the provisional codes can be revised, modified, deleted, or expanded, to include new codes that appear in the interview data (Miles et al., 2014, p. 83). This first cycle aims to initially summarize segments of the interview data (Miles et al., 2014, p. 90). To further detail and enrich each entry, a sub-coding approach is used (Miles et al., 2014, p. 85). As Miles et al. (2014, p. 85) note, sub-coding can be used for all qualitative studies, especially qualitative content analysis and studies with multiple participants and sites.

The summaries derived from the first cycle are then classified into a smaller number of categories or constructs by a second cycle of pattern coding (Miles et al., 2014, pp. 90-91). These pattern codes organize the information from the first cycle of coding into more understandable and logical units of analysis (Miles et al., 2014, pp. 90-91).

4.4.2 Analyzing the Data

After each transcript has been worked through and coded according to the procedure outlined previously, the data had to be analyzed to provide findings. Due to the length of interview transcripts, it is often difficult to look at multiple variables simultaneously (Miles et al., 2014, p. 106). Therefore, Miles et al. (2014, p. 106) suggest displaying the data using either matrices or networks.

Following the recommendation of Miles et al. (2014) matrices have been used to display the data and ease the analysis process. First, each coded segment was paraphrased to then create a summary of each code and interview. The qualitative analysis software, which was also used to code the data, provides a feature that then visually represents each code and its associated summary. Based on this visual overview, summaries of all the individual interview statements for each code were written, and the findings were derived. The findings are supported by direct quotes from the interviewees and summarized at the end of each subchapter.

To show the reader how the researcher progressed from raw data to terms and themes Gioia et al. (2012, p. 20) suggest visually presenting the structure of the data. Therefore, the data structure is shown in Figure 2.

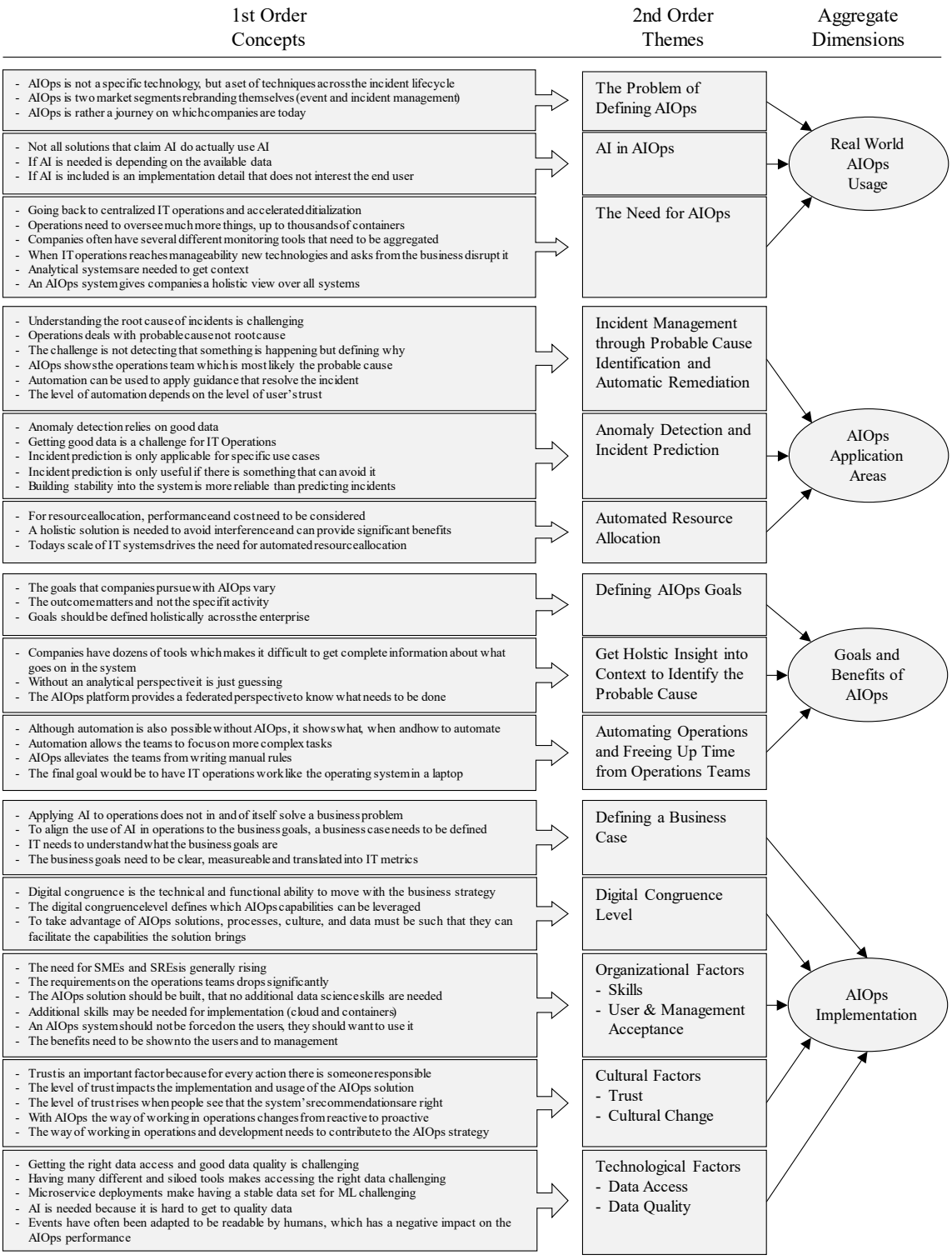

Figure 2: Data Structure (Gioia et al., 2012, p. 21)

The first order concepts shown in Figure 2 reflect statements of the interview partners, which are then summarized in second order themes. Based on this, the four dimensions presented can be aggregated. These are considered in the following chapter.

5 Findings from Case Study Research

The conducted expert interviews have provided valuable insights into the topic AIOps. In this chapter, the findings derived from these interviews are presented. They have allowed to gain a deeper understanding of AIOps from the perspective of experienced professionals in the field. The data gathered from the interviews has been analyzed and organized in a way that provides a clear presentation and interpretation of the key themes and patterns that emerged from the interviews. Through these findings, the aim is to contribute to the existing state-of-the-art and -practice and provide a foundation for the later presented discussion and further research in AIOps.

All the experts interviewed have been working in the field of AIOps and related technologies for several years. The experts come from different areas such as technology, development, solution architecture, system management and consulting. The insights of this differentiated group of experts provide a holistic view of the topic of AIOps, its market adoption, its benefits, challenges and limitations, and its implementation in IT environments.

5.1 Real World AIOps Usage

The following paragraphs show how AIOps is used in real world scenarios. First the problem of defining AIOps is addressed. After that, it is looked at how much AI is in AIOps. Additionally, it is stated where the need for AIOps originates from and what the application areas are. Finally, the objectives and benefits of an AIOps solution are highlighted to then show the current limitations.

5.1.1 The Problem of Defining AIOps

AIOps is seen as the natural evolution of IT operations. IT has always been about automating different things, and AI is cognitive automation, that allows to make decisions based on patterns found in data, which would not have been found manually. The opinions on what AIOps really is do vary. The different analyst's definitions apply to several markets that already existed for a long time (Interview CB, pp. 126-127).

> *"[…] Gartner, as an analyst has this opinion, that AIOps is a platform, which you apply to existing sources of data, right? Now, that's actually aligning much more closely to the previous market, which was event management and incident management. Right now, if you look at what Forrester, a different analyst, is talking about, they talk much more around AIOps is the application of AI and algorithms to existing tools, right? So, they're much more focused on observability vendors and how they're adding AI into what they already do."* (Interview CB, p. 127).

Rather than being a specific technology, AIOps can be seen as previous market segments re-branding themselves (event management, incident management and observability). Those segments are widely adopted by the market already and now include AI capabilities. However, not all solutions that claim to have AIOps capabilities, do really have AI included under the surface (Interview CB, p. 127).

> *"[...] both of those groups refer to themselves as having AIOps capabilities. Now when you peel under the surface, do they have actual AI or machine learning? In some cases yes, in some cases no. [...] if you actually dig into it and say, how much machine learning is there, you'll find very little."* (Interview CB, p. 127).

The interviews showed that AIOps cannot be seen as a specific technology, but rather a journey on which companies are today. AIOps itself as well as most companies are at the beginning of this journey. They are most often using only a sub-part of the capabilities an AIOps platform has and are in a learning phase on how they can leverage the capabilities for their specific use cases (Interview JY, p. 203).

> *"AIOps is very much more of a journey than it is a specific technology [...]. [...] there's a lot of different things that you can realize in AIOps platforms [...]. Very few, if any are all the way here sort of at the proverbial end of it, [...] they're not even at the point to be able to take advantage of anomaly detection in metrics or key performance indicators or they don't have a centralized log management strategy [...]."* (Interview JY, p. 203).

Furthermore, AIOps can be seen as a combination of different techniques across the incident lifecycle. From finding things that can be done automatically, e.g., configuring the system, moving away from manual rules, to finding anomalies in data that would not have been found, and to automatically correlate data. From evaluating if something is wrong, to getting context and understanding where it went wrong and then figuring out what to do to remediate the problem (Interview IS, p. 99).

> *"It's a combination of different techniques mainly applied across the incident life cycle. So, from finding out, if something is wrong to getting context on the data to understanding where it went wrong, like, what is the true root cause to what it actually was."* (Interview IS, p. 99).

As shown, AIOps can be seen as a rebranding of previous areas, which are already widely adopted. AIOps is not one specific technology, but rather a journey and a combination of differ-

ent techniques that are used in IT operations across the incident lifecycle. Most companies today are at the beginning of this AIOps journey, only using specific capabilities of a holistic AIOps solution.

5.1.2 AI in AIOps

Not all solutions that fall within the scope of AIOps include AI functions. Whether AI is included depends on the use case and the amount and quality of the available data. If high quality data is available, more algorithmic capabilities can be used. Where there are gaps in the data, AI is needed to fill those gaps (Interview NB, p. 140).

> *"[…] AI models. They are in general good if you have a lot of data, you have a way of quantifying the attributes in that data and […] you don't really know necessarily what patterns you're looking for. So, AI is very good at that. On the flip side, if you know exactly what you're looking for, and there's a limited set of conditions […], then a more rule-based or policy-based approach might help better […]."* (Interview NB, p. 140).

AI is generally well suited when there is a lot of data, the attributes in the data are quantifiable, but the patterns are unknown. An example is grouping events based on hidden patterns in a large set of events to find instances of related events. However, if it is already clear what patterns one is looking for and the set of conditions is limited, an approach based on statistical rules is more appropriate. For example, this could be a policy-driven grouping of events if a CPU event and a memory event occur within X seconds. Thus, depending on the use case and the available data, either an AI or policy-based approach is better. However, combinations of both approaches are also used. And this is often done in AIOps solutions, e.g., to generate policies with AI (Interview NB, p. 140). The difference between policy-based and AI-based approaches is that in a policy-based approach, a human can still understand what the system is doing. With AI, e.g., Deep Learning, the human no longer understands why the system recommends a certain action (Interview IA, p. 116).

> *"[…] but when it comes to true AI about deep learning and so on, they will not understand why the system is recommending that there will be an incident, why the system is recommending that the change is risky, or why the incidents are related to each other […]."* (Interview IA, p. 116).

Whether AI is included in an AIOps solution is an implementation detail that does not interest the end users. Their focus is on whether the solution solves their IT operations problems and how effectively it does so (Interview CB, p. 127). AIOps is about supporting traditional IT op-

erations to make them have less manual work. This can be done with AI but also with statistical approaches (Interview IS, p. 100).

> *"[…] whether it's technically machine learning it doesn't actually make any difference to the end user, they don't care whether there's actually AI in the box, they care about whether it solves the problem and how effectively it solves the problem."* (Interview CB, p. 127).

Not all solutions that fall into the area of AIOps do use AI to provide their capabilities. If AI is needed depends on the use case and the available data. The better the quality of the available data is, the less AI is needed to reach the desired outcome. However, the end user does not care if there is AI included or not. Their focus is on the results.

5.1.3 The Need for AIOps

Around 2015 it looked like the traditional way of operations, with network operations center, operators sitting there 24/7, having an on-call routing, was gone. Companies started doing DevOps and Site Reliability Engineering (SRE) and thought that they do not need central IT operations anymore. However, over the last years, companies shifting away from central operations realized that DevOps and SRE not solely work. Hence, they went from centralized operations, to decentralized, or to a mixture of both. Also, the increase of new digital services led to an increased need for operations. Thus, a combination of going back to centralized operations and the accelerated digitalization has led to the increased need for AIOps solutions (Interview IS, p. 89).

> "*Over the last couple of years what has happened is, people have realized, like, just that their DevOps and SRE style thing, I think doesn't solely work. It has its merit and its importance, but companies went back from, they were centralized, I mean the existing ones, centralized, decentralized, and now it's a bit of a mixture."* (Interview IS, p. 99).

Also, the scale in IT operations has changed. IT operations now needs to oversee much more things. The same people that looked at a couple of servers in the past, now need to oversee up to thousands of containers. This makes it unmanageable by humans and they need tools to support them (Interview NB, p. 139).

> *"They can't do the work that they did before. The troubleshooting or just keeping eyes on all the system. You know, it simply doesn't scale anymore into modern workloads."* (Interview NB, p. 139).

Medium to large enterprises usually have more than one system of truth (Interview JY, p. 199). They often use different monitoring tools for different components of their IT landscape. Even one application can have several monitoring tools (Interview CB, pp. 132-133). They work with their own cloud providers that have tools, they are using different tools for application monitoring, network performance management, business activity monitoring, infrastructure management plus their own custom solutions. Hence, there are often up to twenty different tools in place that answer the question why the system is exhibiting a certain behavior (Interview JY, p. 199). Getting a combined data model out of all these tools is nearly impossible (Interview CB, pp. 132-133).

> *"[…] there's a credit card company in the US that I do a bunch of work with, right, and just for one of their applications, they use a different monitoring system for their infrastructure. […] So, the underlying machines and virtual machines, to what they use for databases. So, there's different monitoring for the database to the machine that the database is on to the machines that their application servers are on. […]. They use another tool again to monitor the network connections between them. So, they have five different monitoring tools for one application. And that means that getting a combined data model is basically impossible*." (Interview CB, pp. 132-133).

> "*On average, a customer that is in the medium to large enterprises is going to have more than one system of truth. Sometimes they'll have their own cloud providers that have their tools that are getting literally shut down their throats to use for free. Others are using APM tools or network performance management tools or business activity monitoring tools, a security event and infrastructure management tools […], including their own custom and bespoke sources. […] The number often goes into the double digits."* (Interview JY, p. 199).

Also, for incident, problem, and change management there are mature technologies and tools that are acting in the environment, e.g., application resource management, Ansible scripts, and run book automation. Whether or not it is fully automatic or human assisted, these tools perform tasks on the environment. This acting cannot be done in isolation, so these tools stack as blocks on top of each other to act properly and derive appropriate recommendations to build a set of actions to mitigate the situations (Interview JY, p. 199). However, companies today do not have a catalog of automations they can use to remediate. Even if it would be possible to automate the problem detection and the solution would tell the user what to do, a human still needs to act on it (Interview CB, p. 133).

The issue that IT operations has, is that just when they reach manageability of the environment, new technologies come out, new requirements from the business are presented, they need to be faster, and new features should generate more revenue. Just when they have the IT environment in a stable position, new technologies come in that disrupt the stability. And this leads to the fact that managing the environment becomes so complex that it is unmanageable the way it was managed some years ago. Operations wants to be in a state of being dynamically driven from an automation and compute perspective. But every time they get close to that state, new technology comes in (Interview JY, p. 204).

> "[…] *we've been on a, […] almost a rebranding effect in the industry at large between approaches and operations to try to par up manageability with the complexities of the estate and it seems as though we are on this continuing roller coaster ride […], where just when operations has a handle on manageability of the environment, new technologies come out. New asks from the business are presented to the technologists to say, go faster. Here's a new particular option that we can deliver a new feature function to our market to generate additional revenue."* (Interview JY, p. 204).

However, there is a lot of concern when picking additional technologies for IT operations. Operations is the nerve center of the business. Everything else can fail, if operations is up and performing, it is possible to recover but if the systems that are providing insights to operations fail, the business loses the ability to sense. On one side operations knows that they need to do something to be able to manage the environment at large. The goal is to bring manageability back to the complexity of the environment. But at the same time, operations cannot afford a circumstance of being down. Thus, the challenge for providers is to meet the clients where they are in this AIOps journey, so that they can use the new solution seamlessly on top of what they are using, and they do not take a second of downtime. Operations needs to be able to continue using the technology they trust, but in addition they get insights from the new AIOps solution. This is a delicate situation because it is about making a shift in the way their culture, processes, and data occur in the most critical heart of the business (Interview JY, p. 205).

> *"It is a nerve center of the entire business. Everything else can fail, as long as that room is still fine and up and performant, you can recover. If that room, if the systems that are providing insight into that room fail or are taken out of commission, the business loses its ability to sense and get insight into what's going on. […]. You do lose all of your sensory perception simultaneously if that room fails."* (Interview JY, p. 205).

The fact that there are so many different systems used that monitor the whole IT environment, analytical systems are needed to provide insight in context. To get this insight in context, it is substantial to look across as much information as possible, to ingest all perspectives. Without an analytics perspective, it is just guessing what is going on and what incidents could be related. So, to get context, either monitoring or observability or both is needed to get to the insight, why the system is exhibiting the behavior it is exhibiting (Interview JY, p. 200).

> *"I need to look across as much information […] I can possibly do to ingest all of that perspective […]. Without an analytic perspective, I'm grabbing trophies, I'm grabbing post-it notes, I'm grabbing Scrap papers and I'm doing swivel chair operations going from one source of truth to another to try to do that manually, but without that, I can't get that context unless I have either monitoring or observability or both in order to measure the state of what's going on in the insight from the systems exhibiting the behavior, they're exhibiting."* (Interview JY, p. 200).

To also act on the environment is impossible unless it is clear what needs to be done to remediate the issue. The challenge there is, that a company can have dozens of operations tools, and it is difficult to get the answer that might or might not be complete about what is going on in the system. This is where an AIOps platform comes in (Interview JY, p. 200). In giving them holistic information, AIOps makes IT operations more productive, so that they get more time to work on automations for repeating problems (Interview NB, p. 143). Because in the end the reliable way of avoiding problems is not predicting incidents but building resilience into the system. The aim is to have highly resilient systems, which are fault tolerant and scalable to load. That is far more reliable than predicting problems before they happen, because a prediction could always be wrong (Interview CB, p. 135).

5.1.4 AIOps Application Areas

The following paragraphs show the different application areas for AIOps that have been discussed during the interviews.

5.1.4.1 Incident Management

The main effort in IT operations today is focused on problem triaging and determination to identify which component or process needs to be fixed (Interview IA, p. 113). Because you cannot have 100% reliable systems, the main goal is to recover fast if something happens. That is a higher priority than trying to avoid the problems (Interview CB, p. 128). While there is currently a lot of investment in predicting and filtering alerts to reduce noise and improve mean time to resolution (MTTR), the challenge remains in understanding the root cause of incidents

(Interview IA, p. 113). In IT operations today, the failure is already there, and one tries to identify where is the root cause. The actual goal, however, would be to prevent the incident from happening, to achieve failure free operation without interruptions (Interview CK, p. 161).

> *"I think when you look at the lifecycle of an incident, like, detect, isolate, triage, mitigate, restore, learn. I think there's a lot of activity at the front. And I would like to pivot a little bit to focus to […] the back end of the life cycle. Because this is where most of the time is spent, right, and this is where the hard engineering work takes place and this is where, I think an AI system could help."* (Interview IA, p. 124).

There is an opportunity to improve incident management by focusing on the tail end of the incident lifecycle, such as finding similar strategies used to mitigate previous incidents and providing suggestions for how to prevent them from happening again. Learning from incidents can help improve incident response and reduce the need for hard engineering work at the end of the lifecycle. Here an AI system can assist in these efforts (Interview IA, p. 124).

IT operations today is done through events that are submitted by broken components (Interview IA, p. 112). Companies often have a network operations center (NOC) that gets these events as first level and second level. With an event management solution, they quickly see that there is an issue, without having to dig through thousands of events to get to the problem (Interview KS, pp. 171-172). With a directed graph it is then possible to do probable cause and effect analysis (Interview IA, p. 112). Every component sends events and the relationship between them is known, so it is possible to build a remediation plan, where e.g., first the disk space needs to be cleared, then the database needs to be restarted to then restart the application server (Interview CB, p. 131). However, it could be that the system fails again in ten minutes because it is not clear what fills the disk space. Although the service was restored, it could not be identified why the system ran out of disk space (Interview CB, p. 135). Hence, the challenge is not detecting that something is happening but to find out why, to be able to mitigate at speed and restore the service quickly (Interview IA, p. 112). Understanding which of the events caused the underlying problem to then be able to restore the service as quick as possible is a challenge that AIOps could help to solve (Interview CB, p. 135).

> *"[…] the challenge starts right on identifying it. There's a problem in the first place. Then there is, you know, I've got twenty events, which of the systems is actually the closest to the underlying problem, right? And then using the sequence of broken systems to understand how to restore service as quickly as possible. All of those are really problems that AIOps will solve eventually […]."* (Interview CB, p. 135).

The goal of an AIOps solution would be to identify which is the root cause and show the relevant people in IT operations proof that the issue is on their end. Being able to narrow it down and to have proof that the issue resides in a particular part of the system. Here AIOps can help a lot (Interview NB, p. 147). However, AIOps currently deals with probable cause and not with root cause (Interview JY, p. 209).

> *"[…] AIOps deals with probable cause, and root cause analysis is very much in the problem management state. Probable cause is based on all the insight that's been collected. This is the suggested most likely probable cause of the circumstances that are transpiring* […]." (Interview JY, p. 209).

> *"[…] part of our solutions […] is an algorithm to suggest probable cause. We explicitly don't call it root cause. Because it's, that's brave, saying you can detect root cause. I don't think that's sincere. So, we have an algorithm that auto classifies events and says this is like, one of the golden signals, plus information, plus exception."* (Interview IS, p. 108).

Thus, AIOps tells the operations team what the most likely probable cause of the transpiring circumstances is. Although confidence levels are improved with better algorithms, it is still a probable statement and not a "yes this is the root cause" (Interview JY, p. 208). Automatically detecting the root cause is not yet possible. The algorithms auto classify events and state that this is one of the golden signals, plus information, plus exception. Based on the event class and where it resides in the service tree, a score is calculated to identify the probable cause (Interview IS, p. 108). This is one of the big challenges, how far is it possible to go towards root cause. An event management system can only do probable cause if there are events and even then, it does not know why that event was generated. So, there may always be another step towards root cause (Interview CB, p. 132).

> *"Automation typically comes in as well […]. We can let the user now decide that they want to run it or not and then they can either manually step through and do the commands themselves or the system can call out and do it for you, or we can go to fully automated where we say, you know, just run it always, it depends on where the client is on their journey. In essence of how much trust that they want to give to the system."* (Interview NB, p. 141).

Another area of incident management is around automation. When a problem occurs, what is done to automatically take care of the problem, or what can be done to prevent it before it occurs. Holistically, the goal is to find the problem, make sense of it for the user to then provide guidance on how to solve it. And then automation could be used to apply that guidance. However, AIOps solutions are only capable of running an automation if there is one. The system could run it automatically or let the users step through it and do the commands themselves. Hence, an AIOps solution could self-heal problems. If this is implemented and used depends on the level of trust the user has in the system. Furthermore, the automation must be there already. An AIOps tool is not just automatic. For the automation part to work, it must be written in the first place. If the engineers do not get the time to write such automation scripts and the processes are manual, an AIOps tool cannot not do much about it (Interview NB, pp. 147-148).

In summary, incident management is mostly done through events generated by the system. The challenge here is not to detect that an incident is occurring, but to find out why. The goal of an AIOps solution is therefore to determine the root cause of the problem. However, today's solutions deal with the probable cause and not the root cause. Another area that an AIOps solution supports is automatic incident remediation. But these systems are not simply automatic. The automation must first be in place before it can be triggered by the AIOps solution. Furthermore, user's trust in the system's recommendations has to be built.

5.1.4.2 Anomaly Detection and Incident Prediction

When the system accurately tells that there is a problem, the next step is to have the system predictively say that there will be a problem and to provide warnings for upcoming problems (Interview NB, p. 137). The biggest challenge in IT operations is getting good data. And this is also a problem for anomaly detection, which is used to determine if a problem is about to happen. The difficulty is that an anomaly is just atypical behavior. Atypical behavior does not necessarily mean a problem. Such behavior might happen through other things. E.g., the marketing team launched a campaign and now there is more demand on the system than usual. This cannot be detected through seasonality. So, it is difficult to identify if this anomaly is causing an issue or if it is just noise (Interview CB, p. 120).

> *"[…] people in general are using anomalies to try and determine if a problem is about to happen. Now, the problem with that is, so, an anomaly is actually just, should we say, it's atypical behavior, right, now the fact that it's atypical behavior doesn't mean it's necessarily a problem […]. So, how do we know that this anomaly is actually going to lead to a problem versus it being noise itself? And this is some of what we're actually working*

through from a product perspective, is we have to have high confidence that the things that we can detect are actually signals for real problems." (Interview CB, p. 130).

Predicting a problem is only useful if there is something that can be done to avoid it. One area of prediction that is successful, is time series forecasting, but only in specific domains (e.g., disk usage). Many processes do not work well when they run out of disk (e.g., databases or application servers) and clear action that can be taken to remediate and avoid the problem. The volume can be expanded or unused files can be deleted. By applying time series forecasting to disk space, it is possible to predict that the system runs out of disk in the future. But that also does not clearly say that there is going to be a problem, what the problem is and how to resolve it (Interview CB, p. 130). This prediction part is a focus on the provider side, to find ways that the solution can predictively see what problems could arise (Interview NB, p. 148).

"[...] predictions will only work in areas where we can really claim high confidence is going to lead to a problem. And [...] One of the things that we've kind of learned is, if you predict things, which turn out to not occur, then pretty much after that, they ignore the predictions. They very, very quickly lose trust." (Interview CB, p. 135).

"[...] for the clients to be able to trust their predictions, we have to be right and be able to explain to them, why we say this is going to happen." (Interview NB, p. 148).

The focus is on predicting running out of disk and running out of memory because these are problems that break a system, and they are also actionable with zero downtime (Interview CB, p. 130). For the prediction part, the users need to have trust in the system and the providers must make sure that they get highly accurate results and need to be able to explain to the users why something is going to happen (Interview NB, p. 148). However, building stability into the system is far more reliable than predicting problems before they happen, because a prediction can always be wrong. Predictions only work when there is high confidence. If predictions of errors are made that do not occur, the users quickly lose trust in the system and ignore the predictions (Interview CB, p. 130).

5.1.4.3 Automated Resource Allocation

A challenge in IT is to keep the performance of applications healthy. With infinite resources the performance would always be good. But there is performance and there is cost, and both need to be considered. When using different tools for different resource allocation problems, e.g., capacity, storage, compute, or offloading, these solutions can interfere with each other. One says "A", the other "B" and then the user does not know what to do. Thus, a holistic platform that looks at all these problems in the same way is needed (Interview DF, p. 183).

> *"[…] IT can be on premises, cloud, public cloud, and you can have, let's say, cloud native applications, right, with containers and so on that can be both places. And there is a question of how are you keeping your performance of application healthy? […]. And sometimes by healthy, we mean with kind of the least amount of costs that you need in order to keep the performance of the application because, you know, if you had infinite resources, then you could always have a good performance".* (Interview DF, p. 183).

> *"[…] because there are different tools, they thought, okay, sometimes they are in conflict to each other, maybe a tool is telling you to do A, and the other is doing B, and A and B are not the same and you don't know what to do."* (Interview DF, p. 183).

Another factor is scale. In some companies three million decisions must be made every ten minutes for three million objects. The human is not capable to do that. Companies then often use some home written scripts for particular problems, but they do not see the whole picture. The solution here is a fully automated system that is trading of cost and performance and can apply policies. E.g., that a workload must run in a specific region. It also considers sustainability by saving power, e.g., by finding hosts that do not need to run. In addition, the carbon footprint of an application can also be considered (Interview DF, p. 183).

> *"[…] if you step back and say, okay, application has a performance issue, the reason could be two things. One is resourcing […]. The other could be, there is a bug, there is something that was badly written that defect or something that is causing you to always need more and more resources. So, the combination of <an application performance management solution> solving that problem and <an application resource management solution> solving the resourcing problem can be very powerful because it could cover all the application performance issues […]."* (Interview DF, p. 187).

Performance issues are often resourcing problems or application defects that cause the need for more resources. Combining application performance management (APM) and application re-

source management (ARM) can cover the application performance issues. Before these automated resource allocation solutions there was always a human in the loop. This solution tries to completely remove the human from the process, because they can do better jobs in other areas (Interview DF, p. 187).

> *"[…] for example, in your operating system, you don't have a human deciding where to run Excel in which CPU right? It just does it for you right? The same way in a data center should have an automated system to do that for you."* (Interview DF, p. 188).

> *"[…] you think that they're doing a good job […]. But it's such a big problem that no human can really do it. […] There are so many details, for example, […] they had wasted storage that was allocated but nobody was consuming it. You may not notice that right? But the automated system, we will notice it."* (Interview DF, p. 194).

People think that they are doing a fantastic job also without an ARM tool. However, there are things that a human will not recognize, and the system does (e.g., allocated storage that nobody is using and thus is wasted). Furthermore, proactiveness is also implemented in the ARM solution. Even if there is no congestion yet and there is a host that is empty, workload is put there to balance it. That prevents a potential incident on the first host (Interview DF, p. 194).

The need for automated resource allocation became more present due today's scale of IT environments. Companies often use different tools for different problems that can interfere with each other. Hence, using an ARM tool that looks at all problems holistically can provide significant benefits. Combining ARM and APM allows companies to keep the performance of their applications healthy while also considering other factors such as cost, carbon footprint and location of the infrastructure.

5.1.5 Goals and Benefits of AIOps

AIOps can have value for all companies with a set of significant applications that go beyond a simple website. With AIOps they can start looking at what to automate and how to run everything smoothly. So, every company with services that have business criticality can benefit from AIOps (Interview IS, p. 110).

> *"As soon as you have to run services. And as soon as that goes beyond, I don't know a simple website and, and maybe an app […], I would look at how can you automate? How can you run everything smoothly? And once you start doing that, you automatically get into AIOps […]."* (Interview IS, p. 110).

However, the goals that companies pursue with the use of AIOps vary. They are using different capabilities out of the whole AIOps spectrum (Interview CK, p. 155). What matters most is the outcome. Does it affect the user experience? Does it reduce the number of incidents? Hence, companies should chase for the outcome and not for the specific activity. So, it is important to think about it holistically across the enterprise (Interview IA, p. 121).

> *"[…] the outcome should really matter. […]. How many incidents do we have a month? […] And can we reduce the number of incidents? […] Does it affect the user experience? Those are elements to it, right? So, chase for the outcome, not for the activity, and also think about it holistically, right? Don't treat AI just something within ops, right? Maybe developers need to expose the right metric that the system could then leverage […]."* (Interview IA, p. 121).

The focus in operations today is often on freeing up time from operations teams, which are spending their time with managing incident tickets and restoring systems. With an AIOps solution the requirements on the operations team drops significantly. They can now focus on avoiding the problems in the first place because they are freed up to do platform engineering and build resilience into the system (Interview CB, p. 135). It allows companies to shift to an SRE model where the people have more time to fix problems and implement automations to avoid the problems when they occur the next time (Interview NB, p. 143).

> *"[…] but the goal for the organizations […] is really what they're trying to do is free up time from operations teams, who are currently spending all of their time, trying to manage incident tickets and restore systems to get to the point that they can focus on, how do we avoid those problems in the first place?"* (Interview CB, p. 135).

> *"[…] what you would see is there's less repetitive work that the teams would have to do. So, part of what it enables is a bit of a shift to like, an SRE model, […] you could shift people to more of an SRE model where ideally, they would have more time to fix problems and to implement automations to avoid them next time."* (Interview NB, p. 142).

Another challenge for companies that AIOps is trying to solve, is that medium to large enterprises can have dozens of different tools they use in IT operations. Hence, it is difficult to get the answer, which might or might not be complete, about what is going on in the system (Interview JY, p. 200). The output of all these different tools should be aggregated into one solution that provides a holistic view (Interview NB, p. 146). That is where an AIOps platform comes in. To get that holistic insight in context, it is needed to look across as much information as possi-

ble, to ingest all different perspectives. In this case, the AIOps platform can look at the different sources of truth from a federated perspective to provide a single frame of reference to know what to do next or what automation actions to take, and to do the right thing at the right time (JY, p. 200).

> *"The challenge though is if I've got dozens or more of these tools, you've got to know really well your business and really well the particular tool in order to get the answer and that answer might not be even complete about what's going on in the given system, but that's where we come in with regards to our actual product […]. That's looking from a federated perspective across those different sources of truth and get that single frame of reference so that I have confidence to go where to go next or where confidence to know what automation to provide or to act on to be able to engage the right thing at the right time to be able to deal with those situations".* (Interview JY, p. 200).

Through this holistic view, the AIOps platform allows companies to identify the probable cause of the incident. It allows having an actual relationship model of the components, where the events are coming from. The events can then be grouped together. Based on directional analysis it is possible to point out what is the probable cause (Interview CB, p. 129). Over time the people and the AI start working closer together. The people feed the system information and in return get guidance, which allows them to be more productive (Interview NB, p. 143).

> *"[…] if we have an actual relationship model between all of the components that the events are coming from, right? So, a topological graph, a directed graph of the relationships, then we can immediately take those events, we can group them together based on degrees of freedom, and based on directional analysis we can actually point out what is the likely probable cause, versus which is effect […]."* (Interview CB, p. 129).

The next step is around automation, for instance run book automation (RBA), which allows companies to automate operations processes. Instead of manually performing a task, a run book allows the person in the NOC to perform tasks automatically, e.g., automatically restarting a host (Interview KS, p. 173). RBA, however, does not necessarily need an AIOps platform, because it can also be done with other tools (Interview CK, p. 164). But the AIOps platform shows when to automate, what to automate, and how to automate (Interview IA, p. 112). It looks at the whole picture from a federated view to have confidence to know where to go next or what automation to provide, to engage the right thing at the right time (Interview JY, p. 200). Through the automation the people get time for more complicated tasks. However, RBA is not suitable

for every use case. It is important to draw a line for what fully automated systems are used and where a human should take the final decision. (Interview CK, p. 164).

> *"[…] es gibt ja die Möglichkeit manuelle Runbooks oder auch automatisierte Runbooks laufen zu lassen. Das halt auch irgendwann zu einem gewissen Teil auch automatisiert Fehler behoben werden können, dass du jetzt zum Beispiel einen Host, wenn er nicht mehr pingbar ist, erst mal rebootet wird, vielleicht das dann schon mal das Problem behoben wird und die Person im NOC sich da um erstmal gar nicht kümmern muss und nur erstmal beobachtet, ob sich das Problem von selbst behebt."* (Interview KS, p. 173).

> *"[…] everything that alleviates you from writing, boring manual rules is useful. And that is, from my perspective, in the area of anomaly detection in the area of data processing, data enrichment and data correlation […]."* (Interview IS, p. 104).

In general, AIOps solutions are most useful if companies want a cross pillar, cross domain, cross silo overview, which is especially the case in large enterprises. It alleviates companies from writing manual rules in many different areas (Interview IS, p. 104). Thus, it frees up time from the operation's teams, that they can use for writing automations for remediation of recurring issues. The final goal would be to have the IT infrastructure operate like the operating system in a laptop, where it is not needed to think about anything. Companies do not want to spend their money on people to do this (Interview DF, p. 195).

> *"Aber letztendlich ist da schon der erste Schritt raus mit den Externen. Wir brauchen diese Spezialisten nicht mehr, wir haben die selbst im Haus, das ist für mich auch ein erster Schritt […] dann der nächste wird sein intern. […] das Ziel ist es immer, die Kosten zu reduzieren. Jedes Jahr, das war schon immer so und wird auch immer so bleiben."* (Interview JK, p. 231).

In the end, the goal for all companies is to save costs, internally and externally. And that is something that AIOps solutions enable. By automating operations processes, costs can be saved for internal employees and external experts (Interview JK, p. 231). To do that, manual processes must be automated. The final goal would be to have self-healing systems that need no human intervention (Interview PM, p. 212).

5.1.6 Current Limitations and Missing Factors

In IT operations today the failure is often already there, and the teams try to identify where the root cause is. However, the goal of AIOps should be identifying indicators in the data that currently are lost in the noise, which will cause a future outage. Hence, the goal would be to move from reactive to proactive operations. As long as the AIOps solution does not have enough context information it will be difficult to identify outages in advance. It must go further towards observability. Troubleshooting and finding the probable cause works well, but AIOps does not yet identify small warnings that then cause a massive outage in the future, e.g., this event could cause an outage of this network node which leads to that system failure (Interview CK, pp. 160-161).

> *"[…] eigentlich ist die Sache, die man jetzt macht, da ist die Störung ja schon da und man muss das gucken, wo kommt sie her. Aber eigentlich möchte man die Störung ja verhindern. Das ist eigentlich das grosse Ziel. Man will ja störungsfreien Betrieb. Ich will ja nicht, dass es erst mal zum Ausfall kommt."* (Interview CK, pp. 160-161).

In the future, such a solution should be able to predict future events that one can react in time to mitigate the event. To do this, logs and metrics need to be included and the solution must identify the correct root cause every time (Interview CK, p. 169). So, another limited area is the accurate identification of the root cause and automatic remediation of incidents. Here good data is the main thing that is currently missing to go further. The relationships between components need to be understood to direct to the root cause. To improve the ability to do holistic analysis and start building remediation plans a complete data model would be needed. The problem there is that data is often siloed and thus difficult to aggregate (CB, pp. 132-133).

> *"[…] you can only ever detect probable or root cause for things that you've got good data from. And you have to understand the relationship between components in order to basically follow the trail of evidence. And for lots and lots of companies today, their data is siloed. […] we have to get to a complete data model, if we really want to improve our ability to do that analysis and start to build remediation plans."* (Interview CB, pp. 132-133).

Besides the fact that the data within enterprises is siloed, another limiting factor is that companies are closed to their IT and do not want to expose their incidents (Interview IA, p. 119). AIOps models could be much more sophisticated and effective if they could get data from different companies. It could then be possible to have open-source management artifacts, thinking outside of corporate boundaries. Companies should be open to share such data, because then

standards for mitigation actions or standards to respond to incidents could be built (Interview IA, p. 122).

> *"Instead of me as one enterprise have a data point of X, there are thousands of companies that put in information into a model, then the model would be much, much more, sophisticated and effective right but that would require us to think outside of corporate boundaries, right? That we are open to share such data, right?"* (Interview IA, p. 122).

This missing data hinders the further advancing of AIOps. The models cannot learn outside of corporate boundaries. However, applications will always be specific, even in the same sector. But below there is commodity, e.g., Kubernetes, a Postgres database, or Apache for which standards could be built. These things have commonality and there the chance would be to get AIOps models specialized for such commodity solutions (Interview IA, p. 122).

> *"[…] it limits the ability for the model to learn to just my corporate boundaries, right? And my application will always be specific. […] but there is like, commodity on below. Right? So, let's say Kubernetes, or let's say, Postgres database or Apache or MGNX […] They have some commonality, and we could learn from it.*" (Interview IA, p. 122).

AIOps vendors are currently working on better algorithms, better out of the box integrations to seamlessly connect to the data and automatically categorize, cleanse, and de-duplicate it. However, access to real customer data to be able to train and evaluate algorithms and models is challenging. To further advance AIOps capabilities, more real customer data would be needed, because every customer is different and they all have different data and different expectations (Interview IS, p. 107).

> *"I think all vendors are trying to have better out of the box integrations. I think that's the biggest piece to be able to really seamlessly connect to your data and automatically categorize, cleanse, de-duplicate the data. […] some of the challenges that we have is access to customer data. So, we can test algorithms and test models and test ideas."* (Interview IS, p. 107)

Another limiting factor is culture inside and outside of the organization. Today, an external managed service provider (MSP) that takes care of incidents only reacts if an incident is already there. The question that arises with AIOps is, whether a future incident is already an incident or not. If the culture of the company and their MSP is not aligned, the MSP would only intervene

if the incident was there, not if it is predicted to be there in the future. Hence, the contract or the definition of incident must be changed accordingly (Interview PM, p. 214).

> *"The problem is that the MSP, as I said, before, the culture, had an SLA with the customer and a future incident is it an incident? Well, you go to legal, they will tell, you no. It is not an incident because it is not there. You preview that is there, but it's not there. And the rule says, I will intervene if it is an incident not if it's a preview, if it's a prediction so then the MSP, they don't react."* (Interview PM, p. 214).

In conclusion, the current focus in IT operations is on identifying the root cause of an outage. The goal, however, would be to identify indicators in the data that may cause an outage in the future. Accurately predicting future outages is difficult, and AIOps solutions require more context information and observability to do so. To improve AIOps, good data and a complete data model are needed, but data silos and closed IT environments hinder progress. A solution could be to share data across companies and create standards for mitigation actions for commodity solutions (e.g., Apache or Postgres databases). AIOps vendors are currently working on improving algorithms and data integrations, but access to real customer data is challenging, and more data is needed to train and test models. Also, culture can be a limiting factor. It must be ensured that the culture and definition of incidents are aligned between the parties of a service contract.

5.2 AIOps Implementation

The following paragraphs look at AIOps implementation factors. First, the importance of defining a business case is shown. Then, different AIOps triggers and the influence of digital congruence is considered. After that organizational, cultural, and technological factors are presented which impact an AIOps implementation.

5.2.1 Defining a Business Case

The overall objective of having software products is solving a business case. Applying AI to operations, however, does not in and of itself solve a business problem (Interview CB, p. 127). IT measurements are often not tied back to the business needs, so that there is a disconnect between IT goals and business outcomes. Today's value of modern IT operations is mostly on improving IT numbers, instead of serving actual business goals (Interview CB, p. 137).

> *"[...] IT really isn't aligned to the business, right? Because really the IT goals are around availability, throughput, error rates, and that's not actually tied back to the business that it supports in any form other than lots of companies will categorize this as a tier 1 system, that's a tier 2 system. That outlines a set of IT goals as a result. So, there is a*

huge disconnect between IT goals and business outcomes to begin with." (Interview CB, p. 137).

To align usage of AI in operations with the business goals, companies need to define a business case and someone who takes ownership. First, they need to define what they are trying to accomplish. Then, it needs to be identified what is required to get there. And finally, the desired outcome and how it can be measured must be clear (Interview IS, p. 109). To ensure this, it is suggested to define a project owner that looks at it holistically across the enterprise (Interview PM, p. 215). Furthermore, it must be assured that IT understands what the business goals are and that these goals can be translated to a metric that IT can measure (Interview NB, p. 149). Moreover, the IT target (e.g., reducing the number of tickets by factor two) should be linked to the business target (e.g., having X number of goods sold) (Interview PM, p. 218).

"Right down what you're trying to accomplish, write down the required investment and the desired outcome, and then see what you can do to measure that, because if it's not measurable, it's pointless." (Interview IS, p. 109).

"[…] you must have in each of the AI projects an owner, which is above all the parts. If you don't, you are going to lose as my opinion. You must really have synergy between the various stakeholders." (Interview PM, p. 218).

Additionally, the willingness to make the right investment must also be there. If 50% of the applications are defined as mission critical but only 20% are monitored, it is not lining up. It needs to be ensured that the investment and what is monitored aligns with the business goals (Interview NB, p. 149).

"[…] if you say that 50% of your applications are mission critical, but you only want to invest in 20% of that, you know, that's not going to line up." (Interview NB, p. 149).

The specific business case is really depending on the enterprise and the industry it is in. A bank that needs 24/7 service availability should focus more on the incident side. A startup that wants to provide a lot of innovation, however, should focus on the change side, to be able to get faster changes with good quality. Hence, it is needed to find the right use case that best supports the business (Interview IA, p. 125).

"So, I'm a bank and whenever you do a transaction, it will go through and it's always available it's 24 by 7 available right? Then the system should probably focus very much on the incident side, right? If I'm a startup and I want to provide a lot of innovation, right? I want to be, every week there is a new function on my system. Right? Then you probably want to rather look on the change side to say, okay, how can I expedite the changes that they are getting faster and faster and giving me good quality." (Interview IA, p. 125).

In the end, the desired outcome should be an improvement of system reliability and not simply a reduction of the number of incidents. Some incidents are more severe than others. Hence, only looking at the number of incidents does not serve the business goals (Interview IA, p. 125).

"The outcome should not be a reduction of incidents, but rather improvement of the reliability. Right. Um, and then if it's handled incidents that each take a minute, or one incident it takes two hours, doesn't really matter." (Interview IA, p. 125).

Since applying AI to IT operations does not inherently solve a business problem, and there is often a disconnect between IT goals and business outcomes, it is important to first identify a business case before looking at AIOps. The desired outcome must be clear, measurable, and translated into IT metrics. Therefore, identifying a project owner, who has a complete picture is necessary. Furthermore, the planned investment in a solution must also correspond to the expected outcome, which should finally align with a specific business goal.

5.2.2 Triggers for an AIOps Implementation

Having overall a lot of outages is a good first trigger for an AIOps implementation. The more outages, the more actions must follow (Interview JK, p. 225). Also having many people who just do simple manual work is a trigger for an AIOps solution (e.g., manual configuration and manual rules). AIOps is most useful for everything that alleviates companies from writing such manual rules. And this could be for instance around anomaly detection, data processing, data enrichment and correlation. So, AIOps will be looked at by any company with services that have business criticality for the company (Interview IS, p. 104).

"Viele Ausfälle. […] Je mehr Ausfälle, desto mehr muss eine Aktion folgen. Muss irgendwas passieren, damit das reduziert wird." (Interview JK, p. 227).

"I would probably say if you had like a large team that does nothing, but constantly writing manual rules […] that's a good trigger." (Interview IS, p. 104).

Companies should start with AIOps where they have a lot of data volume. For instance, if volume in metric data or events is high, a company could start leveraging AIOps capabilities there. But it really depends on the enterprise, their technology stack, their performance, and their current challenges (Interview IA, p. 119).

> *"So, if you have a lot of volume on metric or on events or on incident records, that's probably a good opportunity to look into. […] Some might start because they have too much noise in the event system. And some struggle because they have so many changes and maybe that changed performance."* (Interview IA, p. 119).

Furthermore, triggers can also be large amounts of events, the possibility to see the relationships in the infrastructure, the possibility to automate with run books, and the possibility to easily group events (Interview KS, pp. 173-174). Besides that, skill can also be a trigger. Operations is less attractive than development, thus less skills are there. If the skills in development and operations are disconnected it can also be a trigger to complement these deficiencies (Interview IA, p. 120). And then there is also the factor of scale. The increasing number of containers makes today's IT environments unmanageable by humans, so they need tools to support them (Interview NB, p. 144).

> *"I think skill is probably also a key element that you want to look into. […] if there's a big disconnect between the skill that you have an in the Dev organization, and the skill that you have in the Ops organization, then this might compliment a little bit the skill deficiencies."* (Interview IA, p. 120).

> *"In practice what you're going to see though is the number of containers and things to manage is exploding at such a pace that, you know, in essence, if you want to have a life as an IT person you need to embrace this kind of help in essence."* (Interview NB, p. 144).

In the end, the drivers are the same as for event management, monitoring, and observability solutions; improve uptime, reduce ticket numbers, and reduce the cost of operations. The solutions are now called AIOps, but they are still event and incident management (Interview CB, p. 136).

> *"[…] the business drivers are still back to the same, you know, we need to improve our uptime, we need to reduce the number of tickets that we've got, we need to reduce the cost of our operations teams."* (Interview CB, p. 136).

It can be said that the triggers for implementing AIOps vary among companies. Generally, having a lot of manual tasks and writing rules is a good trigger for AIOps. Companies should start implementing AIOps where they have high data volume and focus on their current challenges. Other triggers include large numbers of events, relationship visualization, automation with run books, grouping events, and skill deficiencies. Finally, the drivers for AIOps are the same as for event management, monitoring and observability, which aim to improve uptime, reduce ticket numbers, and decrease operational costs.

5.2.3 Digital Congruence Level

Digital congruence is the organization's technical and functional ability to move with the business strategy. A specific tool is often not the answer for being able to define the appropriate strategy and what is needed to be able to execute from an AIOps perspective. It is a combination of data, processes, culture, and tools that must be reassembled and reassessed on a regular basis to make sure a company can be successful. Based on the digital congruence level of the company it is possible to define of which AIOps capabilities it can take advantage of. A company, for instance, can currently take advantage of event analytics, threshold management, and topology management. But now they must work on data, processes, culture to get to the next level (Interview JY, p. 203).

> *"[...] based on quantitative assessment, you are at a digital congruence level to be able to take advantage of event analytics, application assurance, threshold management, topology management capabilities, now you've got, you collectively have work to do across data, process, tools, culture, to be able to take advantage of the next waves and then continually assess that to figure out what are the right things to do because there's a lot of different things that you can realize in AIOps platforms. Doing them smartly is the key to be the most successful."* (Interview JY, p. 203).

Looking at runbook automation as an example. To take advantage of the next best recommended action and perform a ticket to run book automation process, it is needed to tie runbooks to incidents, take that information and immediately act on it. To do that and find the next best recommended actions, the solution scans the solved tickets using NLP and machine learning to find something that worked in the past for a similar problem. This includes a lot of complexity to integrate accordingly to e.g., ServiceNow. However, if the process and data in operations is not appropriate, e.g., the people in operations just write "fixed the issue" in the remediation field, then there is no algorithm that gets any meaningful derivative out of that type of statement. To take advantage of AIOps solutions, it is needed that the processes, culture, and data are such that they can facilitate the capabilities the solution brings. If the culture, processes, and

data are not assessed in advance the solution will recommend “fix the issue”, because that was the data fed to it (Interview JY, p. 202).

> *“It’s extremely powerful. Except if your process and your data in operations are such that your team types in “solved the issue” in the remediation field and close the ticket […]. There isn’t an algorithm on the planet that’s going to get any meaningful derivative from that type of statement that has been put in, because of the process and the culture that had been established.”* (Interview JY, p. 202).

So, to figure out what are the right things to do from an AIOps perspective, the company’s digital congruence should be continually assessed. A lot can be realized in an AIOps platform, but doing it smartly is the key to success. Data, processes, culture, and tools must be in an appropriate state to get the most value from an AIOps platform.

5.2.4 Organizational Factors

The following paragraphs show which organizational factors should be considered when implementing an AIOps solution.

5.2.4.1 Skills

With AIOps and automation in general, companies want to reduce the number of low skilled people. Here, a big part can be automated with AIOps (Interview IS, p. 103). Hence, the requirements on the ops team drops significantly. Those teams are then freed up to do platform engineering and resilience, which is a different set of skills. If a company wants to get to highly available platforms, then the skills need to change (Interview CB, p. 137). On the other hand, the need for SREs or SMEs is generally rising and AIOps is not going to reduce that because the people that build running services are still needed (Interview IS, p. 103). These skills in the operations team remain mostly the same because the companies still need the people that know their systems (Interview CK, p. 156). But level one support, for instance, may not be needed anymore because the system does a lot of their work automatically (Interview IA, p. 118).

> *“What they want to reduce is basically the low skilled people that sit there and just follow list-based instructions or documentation. […]. I think that is a big area where you can potentially automate. I think in general the need for, SMEs, so subject matter experts in their respective domain is in general rising, and I don’t think AIOps is going to reduce the need for that.”* (Interview IS, p. 103).

An AIOps platform should be built in a way that the client does not need any additional data science related skills (Interview CK, p. 156). The solution must abstract a lot of the complexity and aim at IT people and not at data scientists (Interview NB, p. 144). Furthermore, the system should already be tailored to the specific domain. The vendors know what data is coming and how to optimize the models to generate the needed output. However, if a company wants to build their own solution, then they would need such skills (Interview IS, p. 103).

> *"One of our goals as a vendor, and I think the other vendors stood the same way, we claim you don't need a data scientists and machine learning engineer, because we have tailored the model to your specific domain. We know what data is coming. We know how to optimize that model automatically and generate the output that you need."* (Interview IS, p. 103)

Data science people are scarce, and the companies need to do more with less. That is also why companies need AIOps tools that correlate the thousands of events automatically, so that the operations team only needs to look at the probable cause of incidents (Interview CK, p. 156). Furthermore, companies do not want data scientists to monitor their data center, because they cost much more (Interview NB, p. 144). When it comes to metric data prediction and trend analysis, this is straightforward, and no data scientist is needed. However, for more complex cases, where there are e.g., relationship models, chats, trouble ticketing system, JIRA entries etc., then somebody is needed that takes care of it and decides what should be put into the AI system and how good the performance is. In general, the more different capabilities are included the more need for data scientists is there (Interview IA, p. 119). Having people with data science skills supports ensuring that the right KPIs are injected into the solution. Out of thousand KPIs, ten could be the most relevant. If this knowledge is missing, all KPIs must be provided to the AIOps solution. This knowledge, however, does not have to be company internal. It is often a combination of provider, partner, and customer skills (Interview PM, p. 218).

> *"When it comes to, I have metric data and I do some prediction and trend analysis. […] I don't need a data scientist for this. Right? But when it comes to, okay, I have a CMDB, and I have additional relationship models, I have chats, and I have my trouble ticketing system, I have my JIRA entries. You probably need to have somebody who's taking care of it."* (Interview IA, p. 119).

> *"A data scientist must be part of the team. Otherwise, […] you might inject the KPIs, which are not correct. I give you an example […] in the case of […] the analysis we did with the Linux OS, we identified more than one thousand KPIs. If you have a data expert,*

he will tell you actually you need ten to make a good […] understanding of the system." (Interview PM, p. 218).

When working with an AIOps solution, the people automatically get new skills. They will learn what data works well for the AI models. Bigger companies also already have some contingent of data scientists, which could be used to build company specific models if needed or to provide custom ways of integration (Interview NB, p. 144). But even if no new skills are needed based on the use case and the solution, the users must adapt to the new processes. Before they just had their screen, looked at events that came in, and opened a trouble ticket. Now they are faced with event groups and must learn what to do with them (Interview CK, p. 156).

"I think they will naturally pick up some data science skills. But I'm not going to call it traditional data science. […] over time they will kind of learn how, what kind of data works well for AI models […]." (Interview NB, p. 144).

Depending on the kind of AIOps solution, additional skills can also be needed for the implementation. To implement the considered AIOps solution on-premises, Kubernetes skills are needed to understand how a container platform works. Such skills need to be built or insourced if they are not already there. If the company decided to move to the cloud, then cloud skills are needed. But this is really depending on the company. Some use their own hardware on which the container platform then will be installed. Some clients implement it themselves; some need an implementation partner that does it. Based on this decision, different skills are needed (Interview CK, p. 157).

"Wo er zusätzliche Skills braucht, wenn er sie sich nicht einkaufen möchte, ist halt durch den Plattformwechsel entstanden. […] dieser Wechsel natürlich in die Cloud hinein, der sorgt natürlich auch dafür, dass der Kunde sich auch in der Cloud auskennen muss, ja." (Interview CK, p. 157).

With AIOps, enterprises can automate parts of their operations, reducing the demands on the operations team. This allows the team to focus on platform development and resiliency, which require other skills. However, AIOps does not eliminate the need for SMEs or SREs. While no additional operational skills are required to deploy an AIOps solution, certain skills such as understanding how container platforms work or cloud skills may be required depending on the type of AIOps solution used. Furthermore, using an AIOps solution should not require additional data science skills. The bought solution should be already tailored to the specific domain, and

it should aim at IT people and not data scientists. However, for more complex use cases people with data science skills are needed that ensure that the right KPIs are used by the system.

5.2.4.2 User and Management Acceptance

Another factor to consider is the user and management acceptance. Pushback from the operations team is seen as a challenge for AIOps. The teams are often not open to change because they think that with their reactive approach, they already have enough work to do (Interview JK, p. 224). To strengthen the acceptance, it is needed to show the users the new functionalities and how they can benefit from it (Interview KS, p. 176). Especially level one operators will fear that they could lose their job. Hence, it is necessary to show them that with AIOps their work gets more interesting and productive (Interview PM, p. 213).

> *" […] die Teams, mit denen wir jetzt mit den <Hersteller> Produkten arbeiten, die sind da sehr vorsichtig. Weil die mit ihrem reaktiven Modus schon genug zu tun haben und sagen, reicht mir, und das andere wollen sie eigentlich gar nicht so richtig kennenlernen […]."* (Interview JK, p. 224).

It is important that such systems are not forced on the users. They should naturally want to use them. If they do not want to work together with the AI, they feed the system bad data then they get even worse suggestions. To make the users want to use the system they should perceive it as help. E.g., when they see that a problem that usually took four hours to be resolved, and now it can be resolved in one hour because they get the needed information from the AI system, they perceive the value (NB, Pos. 80-84).

> *"So, if you can start showing them, you know, hey, it took you four hours to solve this problem and but look the AI can give you all of these extra snippets of information you could have solved it in an hour that, you know, that helps them. Right? It makes them look better in essence."* (Interview NB, pp. 143-144).

The benefits derived from the AIOps solution need to be shown to the users, but also to management (Interview KS, p. 179). These benefits need to be shown live and in practice, and not solely through presentations and theory (Interview JK, p. 225). However, lack of management support is usually not an issue if the administrators are able to show the benefits (Interview KS, p. 179). There is often pressure on operations teams (Interview IS, p. 225). In the end, management is responsible that the systems are available, and they are aware of that. So, management buy in is often not an issue (Interview JK, p. 225).

"Die Theorie, wie gesagt, die Personen, die das dort machen […] sind nicht interessiert an Präsentationen, Folien von Powerpoint, die wollen das sehen. Deswegen ist eigentlich das einzige, die einzige Chance wie man sie überzeugen kann, ist wirklich zu zeigen." (Interview JK, p. 225).

"Lack of management support not so much. These days, there is so much pressure on operations teams and AIOps is everywhere, and they're being questioned, like, oh, when do you implement an AIOps solution? So, I haven't seen an instance of that at all in the last, like, two to three years […]." (Interview IS, p. 108).

Hence, the benefits that can be derived from an AIOps implementation need to be clearly shown to user and management to gain their acceptance and ensure that the solution then is also being used by the individuals in operations.

5.2.5 Cultural Factors

Besides organizational factors, different experts also stated cultural factors that have to be considered. The following paragraphs show the findings on cultural factors.

5.2.5.1 Trust in the AIOps system

The first cultural factor is the user's trust in the AIOps solution. Trust in the system is a crucial factor because every action in operation is tracked in terms of who gave the authority to make a particular change and that also extends to the AIOps system. In the end a person is responsible for the automations running in the back. No one wants to be responsible for a false negative or a false positive when there are incomplete or imprecise perspectives that are happening. So, it is a risk for the person to solely rely on automation because if it fails, they will be held accountable and could lose their job (Interview JY, p. 207).

"Every action that occurs in operations is tracked in terms of who gave the authority to make that particular change and that extends to the automation systems. Someone is behind as a person's name is behind the set of automation that are running in the environment, and nobody wants to have the fault be put on them by a false positive, a false negative, a false narrative overall when there's incomplete or in precise perspectives that are happening." (Interview JY, p. 207).

The level of trust that people have in the system also impacts the way the AIOps platform is implemented and used. Here, the level of trust is very individual. Some users want to do all commands by themselves, whereas others just let the system manage it when something goes

wrong (Interview NB, p. 141). However, most often the insights coming from the analytics perspective are managed manually and not fully automated. The operations wants explain ability and to be able to validate the system's decision before it is taken. That is why it is challenging to get to a fully automated state (Interview JY, p. 207).

> *"So, some people are, you know, they don't trust anything. In essence, there has to be somebody who goes typed in every command, you know, to the other end of the spectrum where you have people that, you know, when something goes wrong, just run an automation and do it. It, it very much depends on your clients."* (Interview NB, p. 141).

> *[…] what ends up happening all too often […] is, yes, I see the insight that's coming from the algorithms and analytics and perspective, but I'm not going to let automation just do things without the explain ability, the manual exception handling that ends up going on in this space to be able to validate and verify that the insight that was found by the analytic engines is confirmed by, by either myself or other members of the team that might have subject matter expertise."* (Interview JY, p. 207).

During the journey of AIOps people then start to increasingly trust the solution. When they can run the automation manually in the beginning to see that it works, the automation can then be run semi-automatic and fully automatic. But it is important to first build trust in the system (Interview KS, p. 179). Once a human validates the recommendations they can be run independently (Interview JY, p. 208).

Trust is also especially important in the prediction part. There the providers must make sure that they get high accuracy results and need to be able to explain to the users why something is going to happen. If it is not possible to explain the system's recommendations, they will not trust them. Hence, they will not touch their production system unless they believe that the tools' suggestion is the right decision (Interview NB, p. 148). Furthermore, if predictions of errors are made that do not occur, the users quickly lose their trust in the system and ignore the predictions (Interview CB, p. 135).

> *"[…] if you can't tell the client why, then they're gonna be: Okay so, you know, you're telling me my storage is going to fail. I'm going to have to spend real money and time to go replace the storage, when right now there's nothing wrong with it. So, being able to explain to them, why that's going to happen, it's going to be key."* (Interview NB, p. 148).

As soon as AI techniques are used, e.g., to analyze the data, to drop irrelevant data or to correlate data into event groups, people are no longer able to comprehend what the system does. Thus, to gain the user's trust the used algorithms need to be transparent and show how they came to a decision (Interview IA, p. 116; Interview IS, p. 109). This explain ability is key for AIOps because lack of explain ability can be seen as a barrier. If the end users in operations do not trust the system, they will not use it. Especially in cases when the system recommends doing something, or running an automation, users trust the system becomes essential (Interview NB, p. 148).

5.2.5.2 Cultural Change

When a company starts using AIOps, their operations team need to change the way of working. Changing people is a bigger challenge than the technology (Interview CK, p. 166). In today's IT operations culture, first an incident occurs and then the people act on it. With an AI system that states there will be an incident in three days, the problem should be avoided before it occurs. The people, however, are not trained to think like that. Hence, a cultural shift is needed to start thinking differently and to take advantage of the findings the AI system exposes (Interview IA, p. 118).

> *"[…] there is a lot of cultural change required to start thinking differently to respond to all the capabilities that an AI system will expose. But unless you […] change the attitude, people might not take advantage of it. […] we had a pilot with science logic, which is like an AI monitoring system, and it had a lot of predictive alerts. Right? But people did not respond in time because they said, well, it's not burning yet, right? It's not an incident yet. So, I don't need to take an action."* (Interview IA, p. 118).

Another factor to consider is that today people often do not do proper analysis or proper investigation steps. The way of working needs to change that it contributes to an AIOps strategy, meaning that detailed information about the problem resolution needs to be put into the incident tickets. Otherwise, the solution cannot leverage that information to provide, e.g., a next best recommended action. The problem there is, that the teams often do not have enough time to do such proper analysis (Interview IA, p. 123).

> *"If I'm so busy that whenever I solve one incident, I need to move on to the next incident, I will probably not have enough time to really write down a proper analysis or proper investigation steps. […]. So, I need to change the way of working in a way that it contributes to an AIOps strategy."* (Interview IA, p. 123).

Also, in development there are changes needed. They need to build the code that it exposes the right information, log information, exposing the dependencies, on which other microservices the application depends. Applications need to be built to manage. The developers should not only build executables but also management artifacts about dependencies, change, metrics, logs, and that is also a change of way of working (Interview IA, p. 123).

> *"[…] also in development, right, that I expose data, that I have good log information, but also, I expose programmatically my dependencies, right? […]. So, it should be, ideally programmatically, that you say, okay, this application depends on those other micro-services."* (Interview IA, p. 123).

To summarize, when a company adopts AIOps, their operations team needs to change their way of working to take advantage of the capabilities the AI system exposes. Proper analysis or investigation of incidents is often not done today. To contribute to an AIOps strategy detailed information about the problem resolution must be written into the incident tickets. Then the AIOps system can leverage this information to provide next best recommended actions. Changes are also needed in development culture to build the code that exposes the right information.

5.2.6 Technological Factors

The third area of factors are technological factors. The challenges there relate primarily to data access and data quality. An AIOps solution needs good data to learn from, the more the better (Interview CK, p. 152). The most difficult part is getting to the right data, ensuring it contains the needed information and making sure it is of high quality. How the data can be accessed should be considered upfront. If it is not clear where the data comes from and whether it is in the correct format, it is very time-consuming to clarify this. Although AIOps solutions have out of the box connectors, data access is still challenging. It can be hard, for instance, to get access to the needed monitoring tool because it is run by a different team and they do not want anybody interacting with it and are worried that it might impact performance (Interview IS, pp. 104-105).

> *"[…] if you don't know where you want to pull the data from, and if it's in a format that you can use, it takes a while. And that's what we see with our customers. I mean, even if we have out of the box connectors, which we don't always do, sometimes, it's hard to get access to that monitoring tool. Because it's run by a different team. And they don't want to let you lose because they're worried you might impact performance."* (Interview IS, p. 105).

To get to the needed data, the AIOps solution must often be integrated to up to thirty different systems. Each of these systems has to be treated as a single case. Many system providers do not want that others interfere with their solution, because it could cause an incident on their end. This adds to the integration challenge because it is not possible to look inside such solutions to consider all available KPIs (Interview PM, p. 215).

> *"[…] the caveat is that they say we will never provide you any information about what happens in SAP in our cloud, and […] we will not disclose you any information. You can't put agents on our servers. Why? Because they want to keep full control. If you accept these rules, […] AI approach collapses. Because I mean it's a black box. I mean, it's like if you were running on a car, which is a managed by, you don't know who, because it's a black box and you just trust it."* (Interview PM, p. 215).

The fact that companies often have many different siloed monitoring tools adds to the data access challenge. Getting the data out of these siloed tools to get a holistic view is difficult (Interview CB, p. 136). Furthermore, it is also determined that the data needed is not available, the system is not integrated, or the system does not expose enough data or not at the right speed (Interview IA, pp. 122-123).

> *"One of the underlying challenges that customers have […]. It's, they've got five tools today to monitor one application."* (Interview CB, p. 136)

> *"[…] that the data is not available or not accessible, that systems might not be integrated yet. That you are not enabled to get the right data like the metric, right? […] I get my resource data every five minutes: is that sufficient for an AI system to then look into it? Or should I have it every two seconds, right?"* (Interview IA, pp. 122-123).

Most ML models require significant data sets to work. Having big enough and stable data sets is a challenge especially in microservice deployments. It is difficult to get a stable data set because changes are made so often. By the time the model is built it could already have been invalidated by the changes to the system (Interview CB, p. 129).

> *"Now, one of […] the big challenges for that is most machine learning requires significant data sets to work off […]. If I've got a deployment that is scaling in and scaling out over time, and that changes are being made to the system, because I have an agile development team that is releasing changes on a daily or weekly basis. It becomes very hard*

for me to get a stable dataset from which to learn. And by the time I built the model, it may well have been invalidated by changes to the system." (Interview CB, p. 129).

There is a balance and a trade off by what needs to be done through structured understanding of the data and what can be learnt through interference. And this brings challenges because some data sources send the events in batches (e.g., every five minutes) and not as they occur. Hence, all these events seem to be correlated and related, even though they could come from unrelated systems. So, relying on ML is limited. That is why a topological graph is needed to be able to identify which of the events is likely to be the probable cause and which is effect. Problems could be solved better with complete data and deterministic relationships, than with relying on AI to infer relationships in the data. However, AI is needed because for operations it is a challenge to get good data (Interview CB, pp. 129-130).

"[…] so, the biggest problem for operations is getting good data in the first place […]. […] we can solve the problems in general better if we have complete data and things like deterministic relationships, than we can, if we are relying on AI to infer relationships between data." (Interview CB, p. 130).

Another issue with the data is that in the last fifteen years, events have been modified to be readable by humans, the people in the network operations center (NOC). This is contra productive for a machine because information that could not be read by humans was neglected. The system, however, now could make sense out of it. A machine could read the event traps and system logs without filtering. These changes made negatively impact the AIOps usage. So, for companies that are new in this space of using monitoring tools it works well, but for companies with adapted systems the AIOps solutions can get problems because the events do not contain the exact information needed by the system (Interview CK, p. 152).

"Wobei ich habe jetzt auch immer das Problem, wir haben, […] ungefähr mal mindestens 15 Jahre lang oder so, Event Anpassungen gemacht für Menschen. […] im Grunde haben wir als Eventmanager oder Event Management Consultants immer die Aufgabe, so ein Event auch lesbar für den Menschen zu machen, den NOC-Mitarbeiter, ja. […] das ist für eine Maschine hier kontraproduktiv, denn auf diesem Weg wird meistens Maschinen Information, wo wir sagen das interessiert ja keinen, die wird dann weggelassen […]." (Interview CK, p. 152).

In summary, the primary challenges regarding technology are accessing the right data and ensuring its quality. Despite often having pre-built connectors in AIOps solutions, accessing data

is still challenging, especially when it is siloed across different monitoring tools. Additionally, getting sufficient and stable data sets is difficult, especially in microservice deployments where changes occur frequently. Another challenge is that events have been modified to be readable by humans, which negatively impacts AIOps usage. Hence, AI is necessary for operations as it is difficult to obtain good data.

6 Business AIOps Alignment Model

As shown in Chapter 3, aligning business goals with AI is a prerequisite to successfully adopt AI, as it ensures that the AI strategy reflects the business objectives and supports the business strategy (Stecher et al., 2020, p. 13). This is also true for AIOps. As shown AIOps in and of itself does not solve a specific business problem. Hence, a successful AIOps implementation needs a structured approach of aligning the business goals with AIOps capabilities. Companies must ensure that they are aware of the relevant factors that influence a successful implementation. To support companies with adopting AIOps the in Figure 3 shown business AIOps alignment model is presented. The model builds on the presented theoretical foundation of business IT and AI alignment. Based on this, the findings from the case study were incorporated into the model to ensure its practical applicability. The model aims to overcome the six challenges of aligning business and AI shown by Alsheibani et al. (2020): AI business case, relative benefits, top management support, effective use of data, AI talent, and AI compatibility. In addition, the model should be applicable regardless of the sector the company operates in.

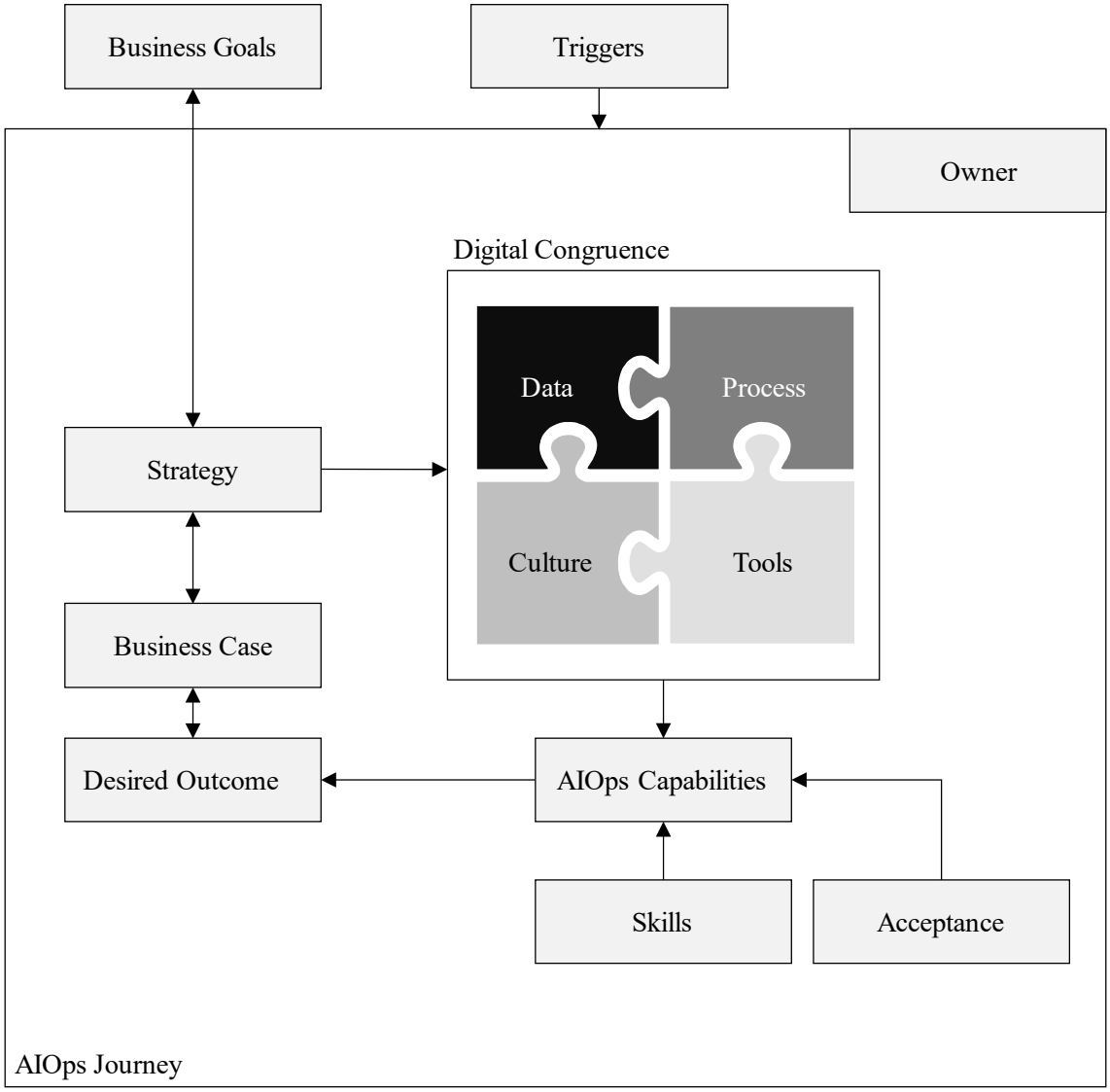

Figure 3: Business AIOps Alignment Model

Triggers. An AIOps project should be triggered by specific problems the company has. These triggers are company specific. In general, having many outages, a lot of manual tasks and data to process are good triggers for AIOps.

Business Case. After the triggers have been identified, the next step is to define a clear business case. The business case should be linked to the company's strategy which is derived from the company's holistic business goals. The metrics that IT operations cares about (e.g., MTTD or MTTR) must be tied back to the desired business outcomes, so that IT operations does not simply improve IT numbers but serves the actual business goals instead. The business case must have a desired outcome that can be measured.

Owner. At the same time, an owner must be defined, who looks at the AIOps project holistically across the enterprise. The owner should have a clear understanding of the business goals, the strategy, the business case as well as the desired outcome and how it can be reached (e.g., which investments are needed and how a successful project is defined). Furthermore, the owner must ensure that the desired outcome serves the business goals and is not simply a metric that IT wants to reach.

Digital Congruence. To understand which AIOps capabilities can currently be leveraged, the organization's digital congruence must be continuously assessed. This is an ongoing task of reassembling and evaluating data, processes, culture, and tools to identify opportunities for improvement and maximize the benefits of AIOps. If digital congruency is not at the required level, it is pointless to invest in additional capabilities. If the underlying foundation is missing, AIOps will not make any difference to the problems. Based on the digital congruence level, the AIOps capabilities whose use currently makes the most sense, can be identified. These identified capabilities should then serve the desired outcome.

Data. One part of digital congruence is data. Data access and data quality must be ensured to be able to provide the right data to the AIOps solution. Without the right data an AI system cannot give relevant and accurate recommendations. The final goal would be to integrate all systems to the AIOps solution. The more comprehensive the picture that the AIOps solution has of the entire IT infrastructure, the greater the benefits that the solution will bring.

Processes. Another factor are processes to ensure that high data quality is established. Looking at incident tickets for instance, the people in operations must be sensitized that they put a detailed incident resolution description into the ticket, so that the AIOps solution can make use of this information.

Culture. With AIOps, the culture of IT operations must change from being reactive to becoming proactive. Thus, it is important that this cultural change is encouraged. This requires a change in the way the people in operations think. Incidents should be managed proactively before they trigger an outage. This needs to be ingrained in people's minds, and the benefits of it need to be clearly demonstrated. Furthermore, collaboration between the different operating teams should also be actively encouraged, as they can benefit from each other.

Tools. Most companies are already using several different tools for their daily operations or to monitor their IT landscape. These tools impact the digital congruence level of a company and thus also the AIOps capabilities that can be leveraged. The AIOps solution must integrate with

the tools that are already in place. Thus, the decision for the right AIOps solution with the needed functionalities must be made.

AIOps Capabilities. Based on the digital congruence level it can be decided which AIOps capabilities should be used. The use of automation capabilities, for example, only makes sense if the data, processes, culture, and tools also enable this functionality. Otherwise, it is pointless to invest time and money in it. For this reason, it is recommended to start with e.g., observe functionalities and then, if possible, introduce new functionalities step by step. In this way, the greatest benefits can be derived from AIOps. In addition, it must always be ensured that the newly introduced capabilities lead to the desired outcome.

Skills. Skills also affects the AIOps journey. It must be ensured that the required skills are present within the organization and that they are sufficient to reach the desired outcome. Based on the chosen AIOps solution and use case the needed skills can be different. The more AIOps capabilities are used, the greater the need for data-specific knowledge to ensure that the solution is using the right data. If these skills are not available internally, they need to be insourced from the AIOps provider or an implementation partner.

Acceptance. The last relevant point is management and user acceptance. To ensure their acceptance, the benefits, functionalities, and value of the AIOps solution must be demonstrated. Although management support is most often not an issue, it is still required to highlight the desired business outcome, how it can be reached and what investments are needed. Also, the users must be shown how they can benefit from the support of AIOps in their daily work. This is relevant to avoid the user's fear of losing their jobs as this fear can hinder the correct usage and thus the benefits of AIOps.

By following this model and taking the stated factors into account, companies can effectively align their business goals with AIOps. It helps them overcome the associated challenges and reap the benefits of AIOps technologies.

7 Discussion of the Findings

The following sections aim to interpret and analyze the results of the research conducted. It examines how the results of the study are consistent with and contribute to the previous literature on AIOps. In addition, this chapter attempts to identify any new insights and perspectives gained from the research findings to answer the following two research questions:

How can companies benefit from using AIOps on their mission critical applications?
How can AIOps be implemented in established IT departments?

Overall, this chapter provides a critical analysis of the results obtained through the research process and offers an opportunity to assess the significance and relevance of the study's findings in the broader context of AIOps research and enterprise adoption.

7.1 Theoretical Implications

IT operations is evolving. The industry must constantly adapt and find new ways to optimize and manage the growing IT landscape. Especially the rise of cloud computing and infrastructure virtualization led to increased complexity in IT operations (McCreadie et al., 2022, p. 136). Thus, it got increasingly difficult to manually manage the IT landscape. AIOps aims to solve these challenges (Gulenko et al., 2020, p. 2). Observability as well as event and incident management have been around for many years and companies are now incorporating AI capabilities to it. Here, it is important to mention that AIOps is not a specific technology, but rather a set of different techniques applied across the incident lifecycle. Although Gartner's definition of AIOps implies that AI is a crucial part of AIOps solutions, the discussions with experts showed that AI is just an implementation detail. Not all solutions that claim AIOps capabilities use AI to get to their results. However, finally it should not bother the end user if AI is included in a solution or not. Much more important is the outcome and the benefits derived. The case study showed that AIOps is applicable to the five levels of automation defined by Ganek and Corbi (2003, p. 9). Although, the technology would already be capable of reaching level 4 "adaptive" and level 5 "autonomic", the companies mentioned are currently on level 2 "managed", starting to investigate level 3 "predictive".

The multivocal literature revealed different application areas of AIOps. Using AI and machine learning to simplify and automate incident management was researched by different authors. Lou et al.'s (2014) approach highlights an effective method for evaluating correlations between time series and event sequences. Arya et al.'s (2021) work combines time series and event sequences for root cause analysis. Both approaches, however, are limited to specific use cases and

are not generalizable. The experts confirmed the importance of capabilities to analyze and use event data as well as time series data but also emphasized that AIOps currently deals with probable cause and not root cause.

The interviews showed that the current focus in IT operations and AIOps is on incident triaging, diagnosing, and determining which components or processes need to be fixed. The goal is to recover quickly if something happens in the system. The recent studies conducted in this regard confirm this focus of the industry (e.g., Shi et al., 2021; Chen et al., 2019a; Chen et al., 2020a; Chen et al., 2020b). However, the practical applicability of Chen et al.'s (2020a) approach to automatically prioritize incidents based on incident ticket information is questionable. The expert interviews revealed that incident tickets often not contain the needed information. Mainly because the operations teams do not have enough time to write detailed descriptions in the tickets. The same applies to Chen et al.'s (2020b) approach for linking incidents together. For these approaches to be applicable in practice the processes and culture in IT operations need to be changed first, so that the operations teams are drilled to ensure that they provide detailed incident information in the tickets.

Another area mentioned by theory and experts is anomaly detection and incident prediction (e.g., Xu et al., 2018; Wu et al., 2021). The importance of including contextual information to detect anomalies and predict future incidents shown by Farshchi et al. (2018) is shared by the interviewed experts. The researchers got to high accuracy results in their studies. However, how well the approaches can be adopted in practice has to be further investigated. As the experts stated, predictions of incidents are only useful if there are clear actions in place that can be taken to remediate or avoid the problem. The studies of Li et al. (2018) and Lin et al. (2020) show, predicted node failures have clear actions that can be taken (e.g., VM allocation and live migration to healthier nodes). Hence, node failure predictions are applicable and useful also in practice. The experts also emphasized that predictions only work in areas where there is high confidence that it will lead to a problem. Hence, the used models need to provide high accuracy results and must be explainable to gain the user's trust. This importance of explain ability of the results derived with such models is also emphasized in the literature (Li et al., 2020; Prasad et al., 2022). Although the presented approaches for node failure prediction proposed by Li et al. (2018) and Lin et al. (2020) provide high accuracy results, the applicability of their approaches in real world scenarios must be investigated further.

A current concern for both providers and customers is resource allocation. The literature shows that AIOps can be applied to automate resource allocation, mostly in cloud environments (Chen et al., 2021). One interview partner also stated this focus of IT operations. Solutions that can

automatically allocate resources are necessary to keep up with the scale of modern IT environments. The amount of decision that must made to use the resource as effective as possible cannot be made by humans. The human resources are better used for tasks, such as building resilience into the system. Combining application performance management and application resource management significantly benefits IT operations by covering most application performance issues.

Looking at the goals of AIOps mentioned in the literature, the focus lies achieving high service intelligence, reducing MTTD and MTTR, enabling self-adaptation or self-healing with minimal human intervention (Dang et al., 2019, p. 4; Shen et al., 2020, p. 276). External and internal satisfaction as well as productivity can be increased by automating high-volume and low-complexity tasks (McKeon-White et al., 2021, p. 2; Prasad et al., 2022, pp. 13-14). The experts share these findings. Especially as freeing up time from operations teams allows companies to focus on platform engineering and building resiliency into the system, enabling a shift towards an SRE model. It was stated that building resiliency in the system is much more effective than predicting future incidents. However, how successful the implementation of an AIOps solution is and if the goals are reached depends on the quality of the data and the used algorithms as well as the organizations' ability to adapt and embrace the technology.

Considering the capabilities of AIOps solutions mentioned in the literature, it is shown that the most notable change is a shift from reactive to proactive operations (Humphrey, 2020, p. 6). This is agreed on by the interviewed experts. AIOps aims to prevent incidents before they occur. However, looking at the functionalities stated by Lithicum (2020, p. 9), the case study research showed that especially predictive spotting of systems failures and self-healing of components are challenging. Although such functionalities are provided by AIOps solutions, companies do often not yet leverage them. To be able to accurately predict future incidents an AIOps solution needs context information and a complete data model. Both are difficult to achieve due to data siloes and closed IT environments. Although providers are working on improving algorithms and integration capabilities, the challenge to access real customer data hinders AIOps progress. Sharing data across companies and creating standards for mitigation actions for commodity solutions could be a possibility to address these challenges.

While the literature on AIOps focuses mostly on different techniques and approaches for the different application areas, there was no research found looking at the actual enterprise adoption of AIOps. Although the literature covers the challenges of IT operations, the application areas, goals, and capabilities of AIOps as well as the need for it, no studies were found that consider the actual implementation of AIOps in a real-world scenario. The conducted case study showed

that implementing an AIOps solution in a company comes with challenges that should not be neglected. These challenges are addressed in the next chapter.

7.2 Practical Implications

The scale in IT operations has drastically changed over the last years. Through increased digitalization and going back to a more centralized operation model, people in operations now need to oversee much more things than in the past. Hence, to be able to manage the complex IT environment, humans need tools that support them. The following paragraphs show the practical implication of the conducted case study.

7.2.1 Benefits From Using AIOps on Mission Critical Applications

As shown, AIOps has the potential to revolutionize how companies manage their IT infrastructure. The benefits of AIOps are numerous and can extend to any organization with mission-critical services. By providing a holistic view of IT operations and automating processes, AIOps can free up time for operations teams to focus on platform development and resilience. As having 100% reliable systems is impossible, recovering fast if something happens is the main goal. Hence, identifying the root cause of the incident is the top priority. Through having a holistic overview of the IT environment, AIOps can assist the operations teams in identifying the root cause by narrowing down the probable cause. Holistically, the goal is to find the problem, make it understandable to the person on the operations team, and then provide guidance on how to solve it or solve it automatically. Fully automated operations, however, is still futuristic, as automation must be written by someone in the first place.

Other notable benefits are improvements in incident prediction and anomaly detection, allowing organizations to anticipate and prevent problems before they occur. However, building resilience into the system is better than predicting future incidents, as predictions could always be wrong. Moreover, predictions of incidents are only useful if something can be done to avoid them, ideally without downtime. So, incident prediction is only applicable in specific use cases (e.g., running out of disk or memory).

Another area is automated resource allocation, which can significantly improve IT operations by optimizing resource utilization and detecting potential performance problems early. Manual and distributed resource allocation is not feasible anymore due to the number of decisions that must be made to perfectly allocate resources and the holistic picture needed. A holistic ARM solution allows companies to automatically allocate resources considering cost and performance constraints. Thus, saving computing resources, power, and money.

The fact that companies often have many different operations tools in place, makes it difficult to get a holistic picture about what is going on in their IT landscape. Here, AIOps can provide this holistic insight in context, providing a federated perspective to make decisions on what to do next and doing the correct thing at the right time. Ultimately the different capabilities of AIOps allows companies to save costs as they do need to hire less people in operations and can prevent incidents and performance issues before they impact the production systems and end users.

7.2.2 Implementation of AIOps

Rather than a specific technology companies should see AIOps as a journey. On this journey, AIOps capabilities will allow them to automate the incident lifecycle. However, the journey to get to an automated operations state is long and challenging. The conducted case study showed that most companies using the solutions of the chosen provider are currently at the beginning of this journey, leveraging only some capabilities a holistic AIOps platform would bring. To make use of the full AIOps potential, leverage the benefits and reach the goals associated with it, companies need to be aware of the challenges of implementing it and how to overcome them.

First, AIOps needs data to derive findings out of it. This data needs to be accessible and of high quality. As most companies have several systems and the data is often siloed, accessing the needed data is challenging. Thus, companies need to break down data siloes and establish processes that ensure data quality. The easier the data access and the better the quality, the more successful the AIOps implementation will be.

Second, companies need to ensure that the users trust the AIOps system and its recommendations. On one hand, providers must ensure that the algorithms used provide accurate results and are explainable to the users. False positive predictions of incidents, e.g., will result in not using the solution or not considering its recommendations. Algorithms need to reveal why certain predictions are made, to allow the users to understand the decisions. On the other hand, companies need to make sure that such a solution is not forced on the users. They should naturally want to use it. This can be reached through showing them the benefits with practical showcases rather than with slides and presentations.

Third, companies need to be aware of the cultural change that comes with AIOps. Adopting AIOps requires a shift in IT operations culture, moving from a reactive approach to a proactive one. This can be challenging for the people working in operations as they generally worked in this reactive mode for many years. Leveraging AIOps also needs changes in the process of documenting and managing incidents. To allow the AIOps solution to derive findings out of previous incident tickets to then recommend a certain action when an incident occurs, the resolution

steps must be documented in detail. This is often not done today, because the operators have not enough time to do so. In addition, it makes sense to encourage collaboration between development and operations teams, as this joint work benefits AIOps.

Fourth, skills and adaptability need to be considered. Although using an AIOps solution does not necessarily require new skills, the users still need to adapt to new processes and tools. Depending on the complexity of the use case, it can be beneficial to involve individuals with data science expertise in the project. These skills can come from internal or external resources. As people with data science skills are scarce, it can be easier to count on solution or service provider resources. However, building skills internally is suggested to avoid high costs. Having skilled people in this area will make it easier to ingest the right data into the solution, which saves time and resources and delivers more accurate results. In terms of implementation, some AIOps solutions may also require specialized knowledge of container platforms or clouds. These can also be insourced if required.

7.3 Limitations and Future Research

This Master Thesis provided valuable insight into the theory and practical application of AIOps in IT operations and its opportunities and challenges for enterprise adoption. However, any study has its limitations. The following section aims to critically reflect on the limitations of the study conducted and presents opportunities for further research. The presented limitations should be considered when interpreting the findings.

The findings of this Master Thesis are based on a single case study looking at one specific AIOps provider and its implementation partners. To confirm the findings, it is suggested to conduct further case studies with other AIOps providers and compare the results to validate them.

As shown, many companies are still at the beginning of the AIOps journey, only leveraging parts of the capabilities of AIOps solutions. Hence, it would be interesting to conduct the same study in future research to see how the AIOps journey evolved at the clients of the provider and implementation partners. As mentioned by different experts, AIOps will become a commodity in the next ten years. This statement could be confirmed or refuted by future case studies with companies from different sectors which are using AIOps.

To validate the stated benefits, a quantitative study investigating the long-term effects on productivity, cost savings, and overall system resilience should be conducted. It is important to note here that the level of AIOps adoption in the different industries must first be more advanced to illustrate relevant results. Looking at companies in different sectors and quantifying

the benefits achieved would allow to provide a recommended course of action for companies considering an AIOps implementation.

As AIOps is evolving and new capabilities will be available in the future, further research should explore the emerging capabilities of AIOps solutions including their benefits and limitations in real-world scenarios. Future advancements in AI capabilities will provide new capabilities, e.g., dynamically creating actions with foundations models before incidents occur. If and how such functionalities will be adopted by the market will be interesting to see.

Another relevant area requiring further research is the impact of AIOps on organizational, cultural and technological factors in different companies. Research focusing on resistance to change in the introduction of AIOps, for example, could help organizations overcome the challenges involved.

Different interview partners also mentioned Chat Ops and risk analysis as other use cases for AIOps. Since these use cases were not related to the two research questions, they were not explored further in this work. The interviews showed that with Chat Ops the way operations teams collaborate could change. The transcripts of these chats could then be mined to find solutions for future incidents. Here, it would be interesting to elaborate how and where Chat Ops and AI for risk analysis processes could be applied and how operations could benefit from it.

Finally, the presented business AIOps alignment model is based on the current findings in the literature and the opinions of the interviewed experts. Therefore, it is currently a theoretical model, whose application in practice has not yet been evaluated. To validate the practical applicability of the model it would have to be used as a reference for a real AIOps implementation project.

8 Conclusion

This Master Thesis evaluated how companies can benefit from using AIOps on their mission critical applications and how AIOps can be implemented in IT departments. The conducted multivocal literature review showed that many studies have been conducted focusing on different application areas of AIOps. The real-world adoption, however, has not yet been researched. Hence, this Master Thesis aimed on closing this research gap by conducting a case study focusing on an AIOps provider and its implementation partners. The case study showed that AIOps enterprise adoption is still at the beginning. Many companies have started using different AIOps capabilities to optimize their IT operations.

Rather than being a specific technology, AIOps should be seen as a journey. Successful and holistic AIOps adoption takes time and cannot be done in one single project. To successfully establish AIOps, companies need to ensure that they align the desired outcomes with their business goals. Their digital congruence level needs to be continuously assessed to define which AIOps capabilities can be leveraged. The presented business AIOps alignment model should assist companies with their AIOps adoption.

As AIOps and the associated technologies are evolving it will be interesting to see what the future holds for IT operations. Will AIOps solve all challenges IT operations has today? Should AI take mission critical and potentially dangerous decisions independently? Who is legally accountable for decisions the AI made? Will humans still be needed in the process, or will AI replace the operations teams completely? These are all questions that will be answered eventually.

Providers, implementation partners and clients should work closely together to overcome today's challenges and drive further advancing of AIOps.

References

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of Practical Program Evaluation* (pp. 492–505). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119171386.ch19

Aggarwal, P., Nagar, S., Gupta, A., Shwartz, L., Mohapatra, P., Wang, Q., Paradkar, A., & Mandal, A. (2021, September 5–10). Causal Modeling based Fault Localization in Cloud Systems using Golden Signals. In *2021 IEEE 14th International Conference on Cloud Computing (CLOUD)* (pp. 124–135). IEEE. https://doi.org/10.1109/CLOUD53861.2021.00026

Alsheibani, S. A., Cheung, Y., Messom, C., & Alhosni, M. (2020). Winning AI Strategy: Six-Steps to Create Value from Artificial Intelligence. *AMCIS 2020 Proceedings*, 1–10.

Alt, R., Auth, G., & Kögler, C. (2021). DevOps for Continuous Innovation. In R. Alt, G. Auth, & C. Kögler (Eds.), *SpringerBriefs in Information Systems. Continuous Innovation with DevOps* (pp. 17–36). Springer International Publishing. https://doi.org/10.1007/978-3-030-72705-5_3#DOI

Arya, V., Shanmugam, K., Aggarwal, P., Wang, Q., Mohapatra, P., & Nagar, S. (01022021). Evaluation of Causal Inference Techniques for AIOps. In J. Haritsa, S. Roy, M. Gupta, S. Mehrotra, B. V. Srinivasan, & Y. Simmhan (Eds.), *8th ACM IKDD CODS and 26th COMAD* (pp. 188–192). ACM. https://doi.org/10.1145/3430984.3431027

Bansal, C., Renganathan, S., Asudani, A., Midy, O., & Janakiraman, M. (06272020). DeCaf: Diagnosing and Triaging Performance Issues in Large-Scale Cloud Services. In G. Rothermel & D.-H. Bae (Eds.), *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Practice* (pp. 201–210). ACM. https://doi.org/10.1145/3377813.3381353

Bogatinovski, J., & Nedelkoski, S. (2021). *Multi-Source Anomaly Detection in Distributed IT Systems.* https://doi.org/10.48550/arxiv.2101.04977

Botezatu, M. M., Giurgiu, I., Bogojeska, J., & Wiesmann, D. (08132016). Predicting Disk Replacement towards Reliable Data Centers. In B. Krishnapuram, M. Shah, A. Smola, C. Aggarwal, D. Shen, & R. Rastogi (Eds.), *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 39–48). ACM. https://doi.org/10.1145/2939672.2939699

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brynjolfsson, E., Rock, D., & Syverson, C. (2018). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics. In *The Economics of Artificial Intelligence: An Agenda* (pp. 23–57). National Bureau of Economic Research, Inc. https://EconPapers.repec.org/RePEc:nbr:nberch:14007

Casanova, C., O'Donnel, G., Lynch, A., & Lynch, D. (05 / 2021). *The Forrester AIOps Reference Architecture: The Need for Clarity.* Forrester.

Chan, Y. E., & Reich, B. H. (2007). IT Alignment: What Have We Learned? *Journal of Information Technology*, *22*(4), 297–315. https://doi.org/10.1057/palgrave.jit.2000109

Chen, J., He, X., Lin, Q., Zhang, H., Hao, D., Gao, F., Xu, Z., Dang, Y., & Zhang, D. (2019, November 11–15). Continuous Incident Triage for Large-Scale Online Service Systems. In *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)* (pp. 364–375). IEEE. https://doi.org/10.1109/ASE.2019.00042

Chen, J., Zhang, S., He, X., Lin, Q., Zhang, H., Hao, D., Kang, Y., Gao, F., Xu, Z., Dang, Y., & Zhang, D. (2020). How incidental are the incidents? Characterizing and Prioritizing Incidents for Large-Scale Online Service Systems. In J. Grundy, C. Le Goues, & D. Lo (Eds.), *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering* (pp. 373–384). ACM. https://doi.org/10.1145/3324884.3416624

Chen, L., Wang, W., Yang, Y., & Xu, Y. (2021). A novel robust prediction algorithm based on REMD-MWNN for AIOps. *Knowledge-Based Systems*, *228*, 107038. https://doi.org/10.1016/j.knosys.2021.107038

Chen, Y., Xu, Y., Li, H., Kang, Y., Yang, X., Lin, Q., Zhang, H., Gao, F., Xu, Z., Dang, Y., Zhang, D., & Dong, H. (2019). Outage Prediction and Diagnosis for Cloud Service Systems. In L. Liu & R. White (Eds.), *The World Wide Web Conference on - WWW '19* (pp. 2659–2665). ACM Press. https://doi.org/10.1145/3308558.3313501

Chen, Y., Yang, X., Dong, H., He, X., Zhang, H., Lin, Q., Chen, J., Zhao, P., Kang, Y., Gao, F., Xu, Z., & Zhang, D. (2020). Identifying linked incidents in large-scale online service systems. In P. Devanbu, M. Cohen, & T. Zimmermann (Eds.), *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 304–314). ACM. https://doi.org/10.1145/3368089.3409768

Cruzes, D. S., & Dybå, T. (2010). Synthesizing evidence in software engineering research. In G. Succi, M. Morisio, & N. Nagappan (Eds.), *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM '10* (p. 1). ACM Press. https://doi.org/10.1145/1852786.1852788

Dang, Y., Lin, Q., & Huang, P. (2019, May 25–31). AIOps: Real-World Challenges and Research Innovations. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion)* (pp. 4–5). IEEE. https://doi.org/10.1109/ICSE-Companion.2019.00023

De Haes, S., van Grembergen, W., Joshi, A. & Huygh, T. (2020). Enterprise Governance of IT, Alignment, and Value. In S. de Haes, W. van Grembergen, A. Joshi, & T. Huygh (eds.), *Management for Professionals. Enterprise Governance of Information Technology,* 1–13. Springer International Publishing. https://doi.org/10.1007/978-3-030-25918-1_10

Dick, S. (2019). Issue 1. *Harvard Data Science Review.* Advance online publication. https://doi.org/10.1162/99608f92.92fe150c

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532. https://doi.org/10.2307/258557

El-Sayed, N., Zhu, H., & Schroeder, B. (2017, June 5–8). Learning from Failure Across Multiple Clusters: A Trace-Driven Approach to Understanding, Predicting, and Mitigating Job Terminations. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)* (pp. 1333–1344). IEEE. https://doi.org/10.1109/ICDCS.2017.317

Engel, C., Schulze Buschhoff, J., & Ebel, P. (2022). Structuring the Quest for Strategic Alignment of Artificial Intelligence (AI): A Taxonomy of the Organizational Business Value of AI Use Cases. In T. Bui (Ed.), *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 55th Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.723

Farshchi, M., Schneider, J.-G., Weber, I., & Grundy, J. (2018). Metric selection and anomaly detection for cloud operations using log and metric correlation analysis. *Journal of Systems and Software*, *137*, 531–549. https://doi.org/10.1016/j.jss.2017.03.012

Floridi, L. (2014). *The 4th revolution: How the infosphere is reshaping human reality*. Oxford Univ. Press.

Ganek, A. G., & Corbi, T. A. (2003). The dawning of the autonomic computing era. *IBM Systems Journal*, *42/1*, 5–18.

Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, *106*, 101–121. https://doi.org/10.1016/j.infsof.2018.09.006

Garousi, V., Felderer, M., and Mäntylä, M. V. (2017). Guidelines for Including the Grey Literature and Conducting Multivocal Literature Reviews in Software Engineering, 1–33.

Ghandour, A. (2021). Opportunities and Challenges of Artificial Intelligence in Banking: Systematic Literature Review. *TEM Journal*, *10/4*, 1581–1587. https://doi.org/10.18421/TEM104-12

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Grundy, J., Le Goues, C., & Lo, D. (Eds.) (2020). *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering*. ACM.

Gillis, A. S. (2018, September 27). *mission-critical application*. IT Operations. https://www.techtarget.com/searchitoperations/definition/mission-critical-computing

Gulenko, A., Acker, A., Kao, O., & Liu, F. (2020, August 3–6). AI-Governance and Levels of Automation for AIOps-supported System Administration. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCCN49398.2020.9209606

He, S., Lin, Q., Lou, J.-G., Zhang, H., Lyu, M. R., & Zhang, D. (2018). Identifying impactful service system problems via log analysis. In G. T. Leavens, A. Garcia, & C. S. Păsăreanu (Eds.), *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 60–70). ACM. https://doi.org/10.1145/3236024.3236083

Henderson, J. C. & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM Systems Journal*, *32*(1), 4–16. https://doi.org/10.1147/sj.382.0472

Humphrey, G. (2020). *The Rise of AIOps: Seven Emerging Managed Service and XaaS Delivery Capabilities You Must Invest in Right Now.* tsia.

Illsley, R., & Grossner, C. (2021). *Data Center Automation Strategies and Leadership.* Omdia.

Jehangiri, A. I., Yahyapour, R., Wieder, P., Yaqub, E., & Lu, K. (2014, June 27 – July 2). Diagnosing Cloud Performance Anomalies Using Large Time Series Dataset Analysis. In *2014 IEEE 7th International Conference on Cloud Computing* (pp. 930–933). IEEE. https://doi.org/10.1109/CLOUD.2014.129

Jiang, J., Lu, W., Chen, J., Lin, Q., Zhao, P., Kang, Y., Zhang, H., Xiong, Y., Gao, F., Xu, Z., Dang, Y., & Zhang, D. (11082020). How to mitigate the incident? an effective troubleshooting guide recommendation technique for online service systems. In P. Devanbu, M. Cohen, & T. Zimmermann (Eds.), *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 1410–1420). ACM. https://doi.org/10.1145/3368089.3417054

Kumar, N., Stern, L. W., & Anderson, J. C. (1993). Conducting interorganizational research using key informants. *Academy of Management Journal*, *36*(6), 1633. https://www.proquest.com/scholarly-journals/conducting-interorganizational-research-using-key/docview/199778042/se-2

Kuckartz, U., & Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden* (5. Auflage). *Grundlagentexte Methoden*. Beltz Juventa. http://www.content-select.com/index.php?id=bib_view&ean=9783779955337

Levin, A., Garion, S., Kolodner, E. K., Lorenz, D. H., Barabash, K., Kugler, M., & McShane, N. (2019, July 8–13). AIOps for a Cloud Object Storage Service. In *2019 IEEE International Congress on Big Data (BigDataCongress)* (pp. 165–169). IEEE. https://doi.org/10.1109/BigDataCongress.2019.00036

Li, J., Ji, X., Jia, Y., Zhu, B., Wang, G., Li, Z., & Liu, X. (2014, June 23–26). Hard Drive Failure Prediction Using Classification and Regression Trees. In *2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks* (pp. 383–394). IEEE. https://doi.org/10.1109/DSN.2014.44

Li, J., Stones, R. J., Wang, G., Liu, X., Li, Z., & Xu, M. (2017). Hard drive failure prediction using Decision Trees. *Reliability Engineering & System Safety*, *164*, 55–65. https://doi.org/10.1016/j.ress.2017.03.004

Li, Y., Jiang, Z. M., Li, H., Hassan, A. E., He, C., Huang, R., Zeng, Z., Wang, M., & Chen, P. (2020). Predicting Node Failures in an Ultra-Large-Scale Cloud Computing Platform. *ACM Transactions on Software Engineering and Methodology*, *29*(2), 1–24. https://doi.org/10.1145/3385187

Lim, M.-H., Lou, J.-G., Zhang, H., Fu, Q., Teoh, A. B. J., Lin, Q., Ding, R., & Zhang, D. (2014, December 14–17). Identifying Recurrent and Unknown Performance Issues. In *2014 IEEE International Conference on Data Mining* (pp. 320–329). IEEE. https://doi.org/10.1109/ICDM.2014.96

Lin, Q., Hsieh, K., Dang, Y., Zhang, H., Sui, K., Xu, Y., Lou, J.-G., Li, C., Wu, Y., Yao, R., Chintalapati, M., & Zhang, D. (2018). Predicting Node failure in cloud service systems. In G. T. Leavens, A. Garcia, & C. S. Păsăreanu (Eds.), *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 480–490). ACM. https://doi.org/10.1145/3236024.3236060

Lithicum, D. S. (2020). *Key Criteria for AIOps.* Gigaom.

Liu, R., Yang, P., Lv, H., & Li, W. (2021). Multi-objective Multi-factorial Evolutionary Algorithm for Container Placement. *IEEE Transactions on Cloud Computing*, 1. https://doi.org/10.1109/TCC.2021.3137400

Lou, J.-G., Lin, Q., Ding, R., Fu, Q., Zhang, D., & Xie, T. (2017). Experience report on applying software analytics in incident management of online service. *Automated Software Engineering*, *24*(4), 905–941. https://doi.org/10.1007/s10515-017-0218-1

Luftman, J., & Brier, T. (1999). Achieving and Sustaining Business-IT Alignment. *California Management Review*, *42*(1), 109–122. https://doi.org/10.2307/41166021

Luo, C., Lou, J.-G., Lin, Q., Fu, Q., Ding, R., Zhang, D., & Wang, Z. (2014). Correlating events with time series for incident diagnosis. In S. Macskassy, C. Perlich, J. Leskovec, W. Wang, & R. Ghani (Eds.), *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1583–1592). ACM. https://doi.org/10.1145/2623330.2623374

Lyu, Y., Rajbahadur, G. K., Lin, D., Chen, B., & Jiang, Z. M. (2022). Towards a Consistent Interpretation of AIOps Models. *ACM Transactions on Software Engineering and Methodology*, *31*(1), 1–38. https://doi.org/10.1145/3488269

Mahidisoltani, F., Stefanovici, I., & Schroeder, B. (2017). Improving Storage System Reliability with Proactive Error Prediction. In *2017 USENIX Annual Technical Conference (USENIX ATC '17),* Santa Clara, CA, USA.

McCreadie, R., Soldatos, J., Fuerst, J., Argerich, M. F., Kousiouris, G., Totow, J.-D., Nieto, A. C., Navidad, B. Q., Kyriazis, D., Macdonald, C., & Ounis, I. (2022). Leveraging Data-Driven Infrastructure Management to Facilitate AIOps for Big Data Applications and Operations. In E. Curry, S. Auer, A. J. Berre, A. Metzger, M. S. Perez, & S. Zillner (Eds.), *Technologies and Applications for Big Data Value* (pp. 135–158). Springer International Publishing. https://doi.org/10.1007/978-3-030-78307-5_7

McKeon-White, B., Nelson, L., & Brown, T. (09 / 2021). *Justify Your AIOps Investment Using Forrester's Total Economics Impact™ (TEI) Methodology.* Forrester.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (Edition 3). SAGE.

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.001

Nedelkoski, S., Cardoso, J., & Kao, O. (2019, July 8–13). Anomaly Detection from System Tracing Data Using Multimodal Deep Learning. In *2019 IEEE 12th International Conference on Cloud Computing (CLOUD)* (pp. 179–186). IEEE. https://doi.org/10.1109/CLOUD.2019.00038

Njanka, S. Q., Sandula, G., & Colomo-Palacios, R. (2021). IT-Business Alignment: A Systematic Literature Review. *Procedia Computer Science*, *181*, 333–340. https://doi.org/10.1016/j.procs.2021.01.154

Prasad, P., Byrne, P., & Siegfried, G. (05 / 2022). *Market Guide for AIOps Platforms* (G00750431). Gartner, Inc.

Ragin, C. C. (2014). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=1698820

Reich, B. H., & Benbasat, I. (1996). Measuring the Linkage between Business and Information Technology Objectives. *MIS Quarterly*, *20*(1), 55. https://doi.org/10.2307/249542

Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data* (2. ed.). Sage.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2. ed.). SAGE Publ.

Shen, S., Zhang, J., Huang, D., & Xiao, J. (2020, August 25–27). Evolving from Traditional Systems to AIOps: Design, Implementation and Measurements. In *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 276–280). IEEE. https://doi.org/10.1109/AEECA49918.2020.9213650

Shi, C., Wu, Z., Lv, X., & Ji, Y. (2021). DGTL-Net: A Deep Generative Transfer Learning Network for Fault Diagnostics on New Hard Disks. *Expert Systems with Applications*, *169*, 1–11. https://doi.org/10.1016/j.eswa.2020.114379

Stecher, P., Pohl, M., & Turowski, K. (2020). Enterprise architecture's effects on organizations' ability to adopt artificial intelligence - A Resource-based perspective. *ECIS 2020 Research Papers*, 1–16. https://aisel.aisnet.org/ecis2020_rp/173

Vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & and Cleven, A. (2009). RECONSTRUCTING THE GIANT: ON THE IMPORTANCE OF RIGOUR IN DOCUMENTING THE LITERATURE SEARCH PROCESS. *ECIS 2009 Proceedings*, *161*. https://aisel.aisnet.org/ecis2009/161

Wang, H., & Zhang, H. (2020, January 6–8). AIOPS Prediction for Hard Drive Failures Based on Stacking Ensemble Model. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 417–423). IEEE. https://doi.org/10.1109/CCWC47524.2020.9031232

Wang, J., Jing, Y., Qi, Q., Feng, T., & Liao, J. (2019). ALSR: An adaptive label screening and relearning approach for interval-oriented anomaly detection. *Expert Systems with Applications*, *136*, 94–104. https://doi.org/10.1016/j.eswa.2019.06.028

Wu, Z., Xu, H., Pang, G., Yu, F., Wang, Y., Jian, S., & Wang, Y. (2021). *DRAM Failure Prediction in AIOps: Empirical Evaluation, Challenges and Opportunities.* https://doi.org/10.48550/arxiv.2104.15052

Xu, Y., Sui, K., Yao, R., Zhang, H., Lin, Q., Dang, Y., Li, P., Jiang, K., Zhang, W., Lou, J.-G., Chintalapati, M., & Zhang, D. (2018). Improving Service Availability of Cloud Systems by Predicting Disk Error, 481–493.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition). SAGE.

Zhang, D., Han, S., Dang, Y., Lou, J.-G., Zhang, H., & Xie, T. (2013). Software Analytics in Practice. *IEEE Software*, *30*(5), 30–37. https://doi.org/10.1109/MS.2013.94

Zhao, N., Chen, J., Wang, Z., Peng, X., Wang, G., Wu, Y., Zhou, F., Feng, Z., Nie, X., Zhang, W., Sui, K., & Pei, D. (11082020). Real-time incident prediction for online service systems. In P. Devanbu, M. Cohen, & T. Zimmermann (Eds.), *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 315–326). ACM. https://doi.org/10.1145/3368089.3409672