Master of Science in Information Systems – Digital Business Systems

Harnessing Artificial Intelligence Capabilities Through Cloud Services – a Case Study of Inhibitors and Success Factors

Student number: 871609

A report submitted in partial fulfillment of the requirement for the degree of Master of Science in Information Systems – Digital Business Systems

Kristiania University College
Prinsens Gate 7-9
0152 Oslo
Norway

Abstract

Industry and research have recognized the need to adopt and utilize artificial intelligence (AI) to automate and streamline business processes to gain competitive edges. However, developing and running AI algorithms requires a complex IT infrastructure, significant computing power, and sufficient IT expertise, making it unattainable for many organizations. Organizations attempting to build AI solutions in-house often opt to establish an AI center of excellence, accumulating huge costs and extremely long time to value. Fortunately, this deterrence is eliminated by the availability of AI delivered through cloud computing services. The cloud deployment models, Infrastructure as a Service, Platform as a Service, and Software as a Service provide various AI services. IaaS delivers virtualized computing resources over the internet and enables the raw computational power and specialized hardware for building and training AI algorithms. PaaS provides development tools and running environments that assist data scientists and developers in implementing code to bring out AI capabilities. Finally, SaaS offers off-the-shelf AI tools and pre-trained models provided to customers on a commercial basis. Due to the lack of customizability and control of pre-built AI solutions, this empirical investigation focuses merely on IaaS and PaaS-related AI services. The rationale is associated with the complexity of developing, managing and maintaining customized cloud infrastructures and AI solutions that meet a business's actual needs.

By applying the Diffusion of Innovation (DOI) theory and the Critical Success Factor (CSF) method, this research explores and identifies the drivers and inhibitors for AI services adoption and critical success factors for harnessing AI capabilities through cloud services. Based on a comprehensive review of the existing literature and a series of nine systematic interviews, this study reveals ten factors that drive- and 17 factors that inhibit the adoption of AI developer tools and infrastructure services. To further aid practitioners and researchers in mitigating the challenges of harnessing AI capabilities, this study identifies four affinity groups of success factors: 1) organizational factors, 2) cloud management factors, 3) technical factors, and 4) the technology commercialization process. Within these categories, nine sub-affinity groups and 20 sets of CSFs are presented.

Keywords: Cloud Computing; Cloud Services; Artificial Intelligence; Technology Adoption; Critical Success Factors

Acknowledgements

This master thesis presents the academic and professional qualifications in the chosen field of study and is a partial fulfillment of the requirement for the degree in Master of Science in Information Systems – Digital Business Systems, conducted at Kristiania University College. My interest in emerging IT trends and technology´s ability to increase efficiency and streamline business processes if adequately implemented, configured, and used have sparked my curiosity about cloud services´ role in artificial intelligence development. This thesis has significantly increased my knowledge and experience in qualitative research. I hope this paper can spark an interest in advancing the research on AI cloud services and aid practitioners and business leaders in strategic decision-making regarding the adoption and use of cloud services for AI development.

I want to thank the informants that have dedicated time for interviews and valuable conversations. The informants´ knowledge and genuine enthusiasm for the topic have been a massive motivation for me during the writing of this thesis. I am grateful for everything I have learned from such experienced and professional workers.

Although the Covid-19 pandemic has largely impacted teaching and learning methods, I am grateful to the professors of Kristiania University College that have handled the situation well and provided students with additional courses and certifications to augment students' knowledge in relevant fields. I would also like to thank my supervisor for guiding me through this dissertation. Going forward, I am excited to bring the knowledge I have gained during this master´s program into my working career.

Total number of words: 23 985

Table of Contents

List of figures

List of tables

1. Introduction

1.1 Problem definition

As we are progressing into the fourth industrial revolution, researchers within the information systems and computing domain frequently argue that this era will be empowered by machine learning (ML), robotics, and artificial intelligence (AI) (Wang and Siau 2019). AI will become as much part of everyday life as the internet or social media did in the past (Haenlein and Kaplan 2019). As a result, AI will fundamentally transform how firms make decisions and interact with external stakeholders such as employees, customers, and business partners. An Accenture report (2019) analyzed 1500 companies that collectively spent $306 billion on AI applications from 2017 to 2019 (p. 6), which signifies that AI is a high-stakes business priority. Additionally, Harvard Business Review estimate that AI will add $13 trillion to the global economy by 2030 (Fountaine, McCarthy and Saleh 2019). Even though vast amounts of money are invested in AI, many companies do not achieve the expected value. A Deloitte survey of 152 AI projects reported that 47% of senior managers found it difficult to integrate AI with existing people and processes (Davenport and Ronanki 2018), and 85% of AI initiatives fail to deliver on their promises (Rayome 2019). Research literature has also observed challenges related to AI, and some of them are upskilling employees and scarcity of AI experts (Chui and Malhotra 2018), coordinating the AI-augmented workforce (Jarrahi 2018), and the complex and demanding process of adopting and integrating AI (Dutta 2018). Organizations attempting to build AI solutions in-house often opt to establish an AI center of excellence, amassing huge costs and an extremely long time to value on their AI projects (Davenport 2018). The majority of enterprises understand the value of AI. However, Forbes (2020) found that only 14% of firms have deployed AI capabilities in production, which indicates that many organizations still fail to adopt AI and harness its full potential.

1.2 Research justification and scope

The big data generated from consumers, information systems, sensors, and the public web has become a critical business resource, popularly deemed the "new oil." According to Gartner Inc. (2020), public cloud solutions are necessary for 90% of all data analysis innovations within 2022. The public cloud enables businesses to share resources and complex system architecture to implement applications and solutions (Sallehudin et al. 2019). Additionally, the abundance of available data and AI enables to turn static data into actionable insights. Cloud services have thus seen significant uptake as organizations seek to benefit from reduced entry barriers and lower operating costs brought by providers' economies of scale. Information systems research has started to devote more attention to AI due to the claimed transformational potential across industries and sectors (Magistretti, Dell'Era and Petruzzelli 2019). However, while cloud computing has been a popular topic in IT and IS research since the term was proposed by Google in late 2006 (Yang and Tate 2012), literature on AI delivered through cloud services is still in its inception due to the novelty of the technology.

As AI advances and businesses seek to acquire its capabilities, cloud vs. on-premises deployment will become a pivotal issue in the minds of business leaders. Cloud vendors provide various assortments of technical infrastructure that make various applications available as a service, commonly termed "Anything-as-a-Service" (XaaS) (Duan et al. 2015). In recent years, several cloud vendors like Amazon, Microsoft, Google, and IBM have started to offer AI services and products. Cloud vendors provide access to infrastructures (IaaS) that enable users on-demand access to specialized hardware and computing power and platforms (PaaS) that assist data scientists in developing, training, deploying, and managing algorithms and machine learning models. AI services also refer to AI software services (SaaS) and third-party offerings (commonly termed AIaaS), which allow companies to buy off-the-shelf AI solutions and pre-trained machine learning models, removing the burden of training, configuring, and deploying AI (Lins et al. 2021). Recent research debates propose that AI software services are a valuable alternative for organizations that face difficulties implementing in-house AI. Pre-built AI software is significantly easier to use compared to cloud services for custom AI development. However, the lack of customizability and control prompts businesses to rather adopt and utilize AI infrastructure and platform services. The infrastructure and platform deployment models enable companies that are unable to embrace AI due to a lack of either appropriate off-the-shelf solutions or complex IT infrastructure and IT outsourcing processes to adopt, deploy and run effective, customized solutions for their unique use case, data, and business needs. Businesses that have successfully adopted and implemented cloud-based AI and ML solutions have reported more streamlined processes, faster time to profit, and improved customer satisfaction (Girvin 2021). What's more, in a Deloitte (2018) study, 49% of the respondents reported that their company used cloud-based AI development services to acquire and develop AI. Extant literature on cloud-based AI services suggests an optimistic view of the ease of use of AI services. A possible reason for this optimistic view may be that prior research primarily covers AI software services and applications relating to the conventional SaaS cloud model (e.g., Zapadka et al. 2020; Cobbe and Singh 2021; Hanzlik et al. 2021). Additionally, the majority of research on AI delivered through cloud computing services takes a technical perspective (see Lopez Garcia et al. 2020 p, 18682 - 18683). Although cloud vendors offer AI developer tools and AI infrastructure services relating to the conventional platform- and infrastructure models, extant research lacks sufficient evidence of adoption and prerequisites for succeeding with such AI services.

1.3 Research questions and objectives

This study aims to fill the research gap by building on the DOI theory to analyze the drivers and inhibitors for adopting AI infrastructure and platform services. By applying the Critical Success Factor (CSF) method, this study also identifies the most critical factors vital for organizations to harness AI capabilities through cloud services successfully. The rationale is associated with the tremendous complexity of developing AI and ML solutions that meet a business's actual needs. Based on the above discussion, this study aims to answer the following research questions:

RQ1: What are the drivers and inhibitors for adopting and succeeding with AI through cloud services?

RQ 2: What are the critical factors for bringing AI ideas to production in a cloud environment?

RQ 3: How can organizations ensure continuous business value capturing with cloud AI projects?

The overarching objective of this study is to understand the challenges and pitfalls of AI services adoption and identify and categorize critical success factors for harnessing AI services capabilities. Based on the research questions of this study, it is likely to identify many potential technological and organizational mechanisms at different levels, interacting in different ways. The sub-objective is to identify the most complete and logical CSFs given the business conditions and use cases. In addition, this study aims to contribute to the existing body of knowledge on technological and organizational development in information systems, data analytics, and cloud computing research.

The remaining of this paper is structured as follows. A systematic literature review is conducted, including a detailed description of the literature collection strategy. Then, the

Diffusion of Innovation theory and the Critical Success Factor method is described. The theoretical framework subsequently shaped the methodology of this study, which is elaborated on next. Findings are presented in addition to a thorough discussion and analysis of the findings. Finally, implications are drawn and concluding remarks sum up the study's key results.

2. Literature Review

This section provides an overview of the literature collection strategy and the key topics and concepts discussed in the collected research articles. First, the literature review briefly covers the background of the topic field, followed by the AI cloud delivery paradigms and some core aspects of adopting and utilizing cloud-based AI services. This literature review also sheds light on the pillars of AI services and describes what research comprehends from a technical and business perspective. Due to the novelty of sophisticated cloud-based AI services, research lacks evidence on critical success factors for its adoption and use. Lack of research is both a blessing and a curse. While inadequate information limits the ability to conclude a given question, this paper takes advantage of the research gap by forming lacking information into questions investigated in the empirical data collection. Therefore, this literature review contributed towards refining the problem definition for the empirical data collection.

2.1 Literature collection strategy

This literature collection strategy aims to synthesize research findings transparently and reproducibly. A systematic literature search is conducted based on the protocols of Webster and Watson (2002). The literature search is conducted in several databases that are deemed relevant (Springer, ACM digital library, and IEEE) with the following search string in the title:

("artificial intelligence" OR "AI" OR "machine learning" OR "ML") AND ("as a service" OR "aaS" OR "cloud" OR "cloud computing")

The search is limited to publication titles because preliminary searches with the above search string resulted in numerous publications referring to AI in abstracts or keywords without explicitly focusing on cloud-based AI services. Selected literature is filtered and selected based on inclusion and exclusion criteria. The credibility of the resources linked through the publishers is assessed on impact score, and only papers from peer-reviewed academic journals published between 2015 and 2022 are considered. The final step of selecting relevant literature is through a forward literature search to identify articles that cite an original article and a backward literature search by reviewing the initial selected article's reference lists to find additional papers not found through keyword search.

Figure 1 – Overview of the literature collection and filtering process

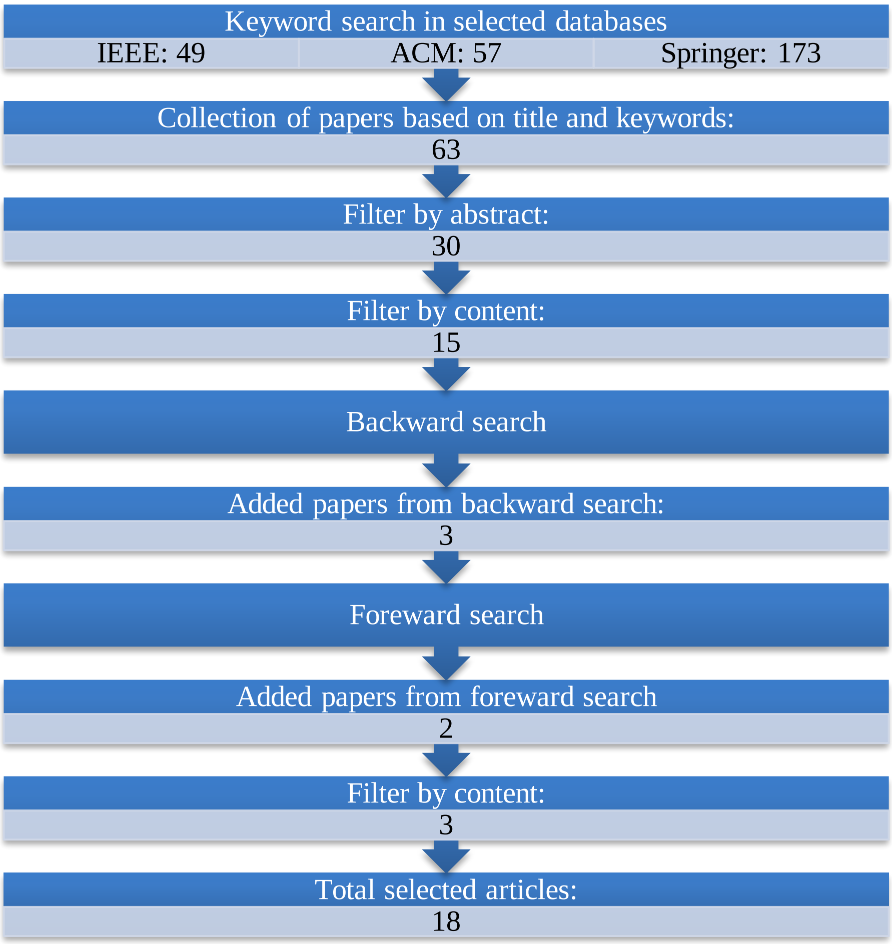


*Figure 2 – Number of published papers by research area**

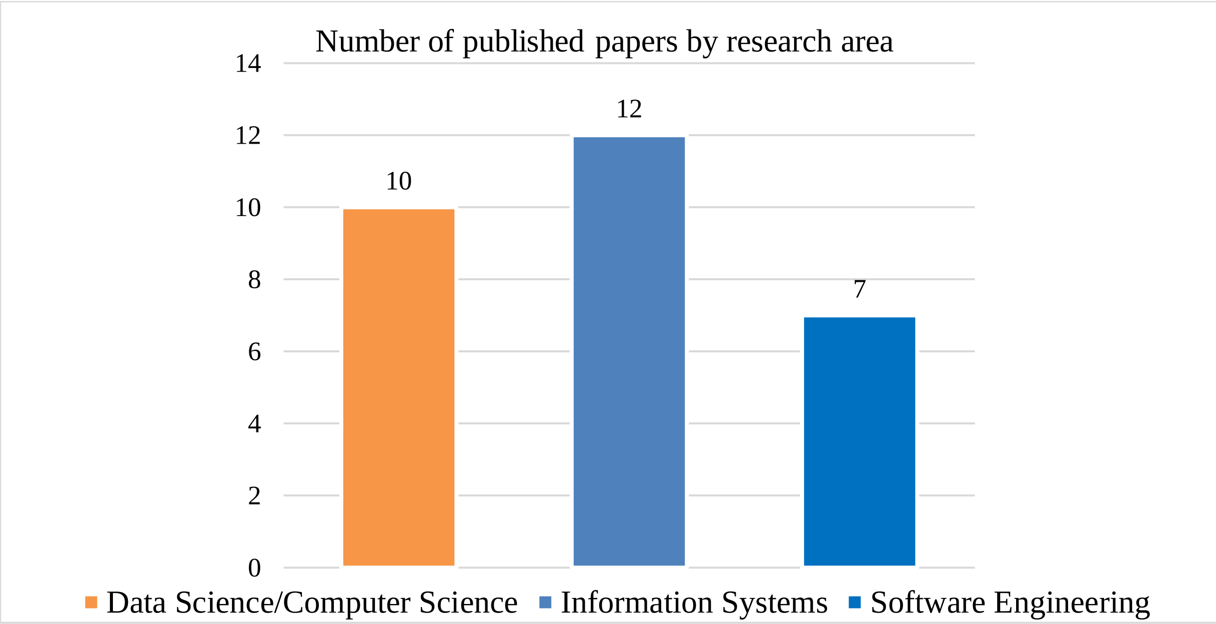
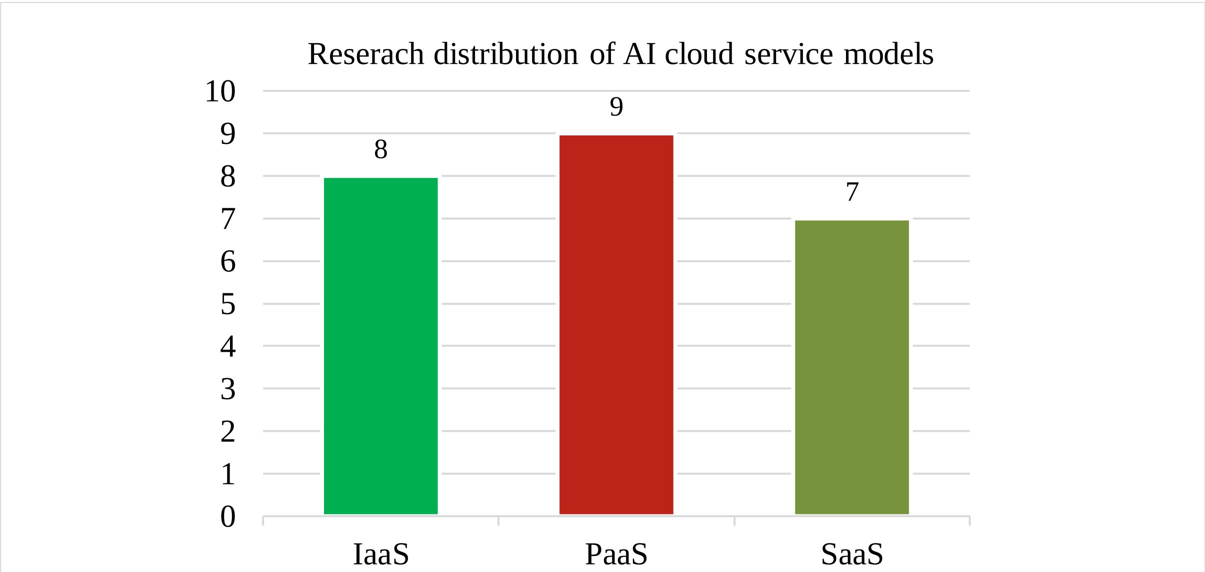


*Figure 3 – Research distribution of AI cloud service models**



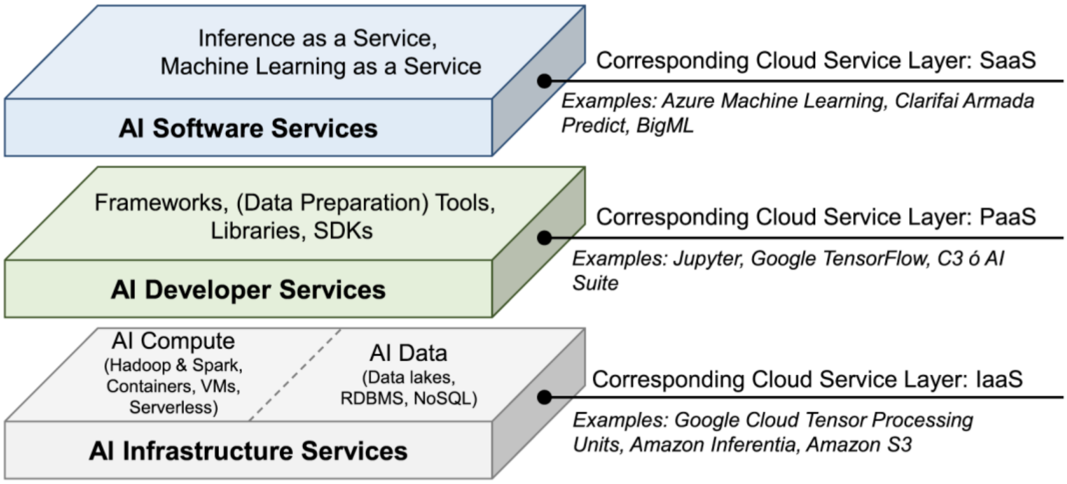
**Several of the reviewed papers comprise more than one research area and more than one cloud delivery model*

2.2 Subsets of artificial intelligence and cloud computing

Artificial intelligence is a broad branch of computer science, typically referred to as an umbrella term for data-related topics and cognitive technologies, including machine learning, deep learning, and artificial neural networks (Jordan and Mitchell 2015). These subsets of AI include additional cognitive computing capabilities, such as computer vision and natural language processing (NLP) (Girasa 2020), to name a few. Machine learning consists of algorithms whose performance improves as they are exposed to more data over time, while deep learning is a multi-layered Neural Network that learns from a vast amount of data. Artificial neural networks consist of algorithms that recognize underlying relationships in a data set through a process that mimics how the human brain operates. AI can thus be defined as "the ability of a machine to perform cognitive functions that we associate with human minds" (Rai et al. 2019, p. iii).

Cloud computing can be defined as an "IT deployment model, based on virtualization, where resources, in terms of infrastructure, applications and data are deployed via the internet as a distributed service by one or several service providers. These services are scalable on-demand and priced on a pay-per-use basis" (Böhm et al. 2011). Cloud computing is designed as a layered architecture. It is often described in terms of the following service models: Infrastructure as a Service (IaaS), which involves «lower-level» computing resources, for example, virtual servers and virtual machines; Platform as a Service (PaaS), which provides various components that support application development and deployment, for example, database systems; and finally, Software as a Service (SaaS), which involves complete, managed applications, such as a pre-built webmail service. However, there are many cloud offerings, and these distinctions may blur. Therefore, the term Anything as a Service (XaaS) (Duan et al. 2015) is increasingly used, reflecting that almost any computing and software infrastructure can be accessible as a service. AI services are subsets of the cloud service models explained above. Figure 4 provides examples of various AI services according to the conventional cloud deployment models.

Figure 4 – AI services stack according to the conventional cloud service models (adopted from Lins et al. 2021)



1. AI infrastructure services can include AI compute resources such as Hadoop and Spark, containers, virtual machines, and serverless computing. IaaS also encompass memory, disk, network, operating systems, and specialized hardware such as central processing units (CPUs), graphic processing units (GPUs), and tensor processing units (TPUs) (Lins et al. 2021).
2. AI developer services can include frameworks, data preparation tools, libraries, and software development kits (SDKs). PaaS also encompass data science services such as Jupyter notebooks and data catalog services, technical environments, tools, and resources that enable data scientists to develop, operate, deploy, and maintain AI solutions (Elshawi et al. 2018).
3. AI software services relate to pre-built software and inference (e.g., prediction) services for generic use cases. SaaS encompass pre-built ML models, such as chatbots using natural language processing, provided to users on a commercial basis. Clients of pre-build AI solutions can essentially "plug" the AI software into their applications (Cobbe and Singh 2021).

There are several possible AI service arrangements, and some services may represent a hybrid of the above. This paper focuses specifically on (1) and (2) and defines AI services as cloud-based systems providing on-demand services to organizations and individuals to design, train, deploy, manage, and maintain AI solutions (Lins et al. 2021).

2.3 AI cloud delivery paradigms

Cloud providers are gearing up to offer a comprehensive stack to deliver AI as a service for business, which aims to democratize the development of AI solutions. For example, major cloud providers such as Google, Microsoft, IBM, and Amazon typically offer machine learning, deep learning, big data analytics, and inference-as-a-service to foster AI diffusion and application (Lins et al. 2021). AI and ML researchers and practitioners turn to the cloud mainly because AI services inherit the strengths and typical cloud characteristics, such as broad network, on-demand self-service, resource pooling, rapid elasticity and scalability, and measured service (Mell and Grance 2011).

2.3.1 AI infrastructure services

AI infrastructure services deliver virtualized computing resources over the internet and enable raw computational power and specialized hardware for building and training AI algorithms (Lins et al. 2021). IaaS is the supporting infrastructure for customers to undertake machine learning with, for example, ML frameworks and GPUs (Javadi et al. 2021). With IaaS, users can use remote physical or virtual machines and quickly scale up the compute resources based on user requirements instead of establishing their own data center (Wang et al. 2018). AI services users typically have a wide choice of provisioning computing resources, such as physical servers, virtual machines, containers, or AI-specialized hardware such as central processing units (CPUs), graphic processing units (GPUs), or tensor processing units (TPUs) for computations (Pandl et al. 2021). Due to both efficient hardware utilization and economics of scale, cloud service providers have the resources to deploy expensive, specialized hardware. Besides, they have the expertise required to develop and maintain efficient AI infrastructures and handle performance peaks dynamically (Romero et al. 2019). For example, applying complex deep learning and neural networks requires complementing CPUs with GPUs for faster calculations (Lins et al. 2021). AI infrastructure services typically provide access to relational or NoSQL databases and the capability to upload and integrate external data lakes as input to train AI models (Pandl et al. 2021). IaaS enables to automatically adapt and optimize the underlying hardware concerning the unique demands of an algorithm (Elshawi et al. 2018). With data residing in the cloud, exploiting the compute infrastructure to train machine learning models enables developers with the required expertise to individually create, configure, customize, optimize, and control their AI models to their needs fully (Lins et al. 2021). IaaS enables complete control over an AI solution that can run independently on a cloud platform. According to Pandl et al. (2021) and Lins et al. (2021), this also prevents vendor lock-in.

When using AI, organizations' hardware requirements typically change frequently and quickly. Fortunately, the scalability of the cloud, combined with the number of available hardware resources, results in a large amount of processing power provisioned by the cloud and enables AI services to respond to extensive requests with scalable and responsive utilization of CPUs and GPUs (Bao et al. 2018). For example, the training of ML models can require powerful GPU resources for a certain period of time (e.g., weeks), while the hardware requirements for the inference of ML models are typically much less. By using infrastructure AI services, organizations can share hardware resources building on a multi-tenant architecture, thus, utilizing the hardware resources more efficiently (Shaukat et al. 2018). The downside of AI infrastructure services is a complicated setup, the responsibility for maintenance, security, access controls, and other details that turn the focus away from solving business challenges using AI (Wang et al. 2018). Figure 5 pictures the shared operation and maintenance responsibilities between a customer (company) and cloud vendor in the IaaS model.

Figure 5 – Shared responsibilities in the respective cloud service models

On-premises	**IaaS** **Infrastructure a Service**	**PaaS** **Platform as a Service**	**SaaS** **Software as a Service**
Applications	Applications	Applications	Applications
Data	Data	Data	Data
Runtime	Runtime	Runtime	Runtime
Middleware	Middleware	Middleware	Middleware
Operating system	Operating system	Operating system	Operating system
Virtualization	Virtualization	Virtualization	Virtualization
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

Vendor-managed

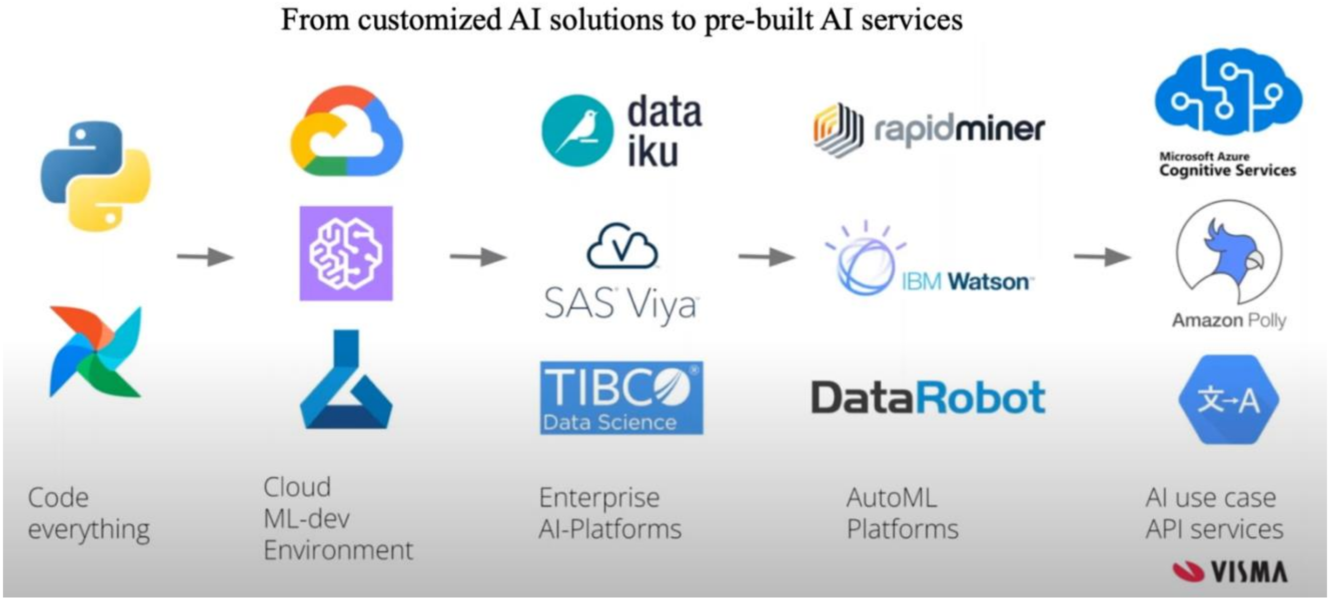
Client-managed

Duong and Sang (2018) also imply that setting up and maintaining a distributed computing cluster is time-consuming and resource-intensive and thus reduces the time for core business activities. Users must also spend a significant amount of time selecting an appropriate configuration for training. If not appropriately configured, training ML models on the cloud could incur high costs and time (Duong and Sang 2018). Due to various cloud resource types, their performance levels, and the pricing offered by public cloud providers, there are many potential cluster setups. Accordingly, efficiently choosing a suitable cluster configuration remains a considerable challenge for businesses without the required expertise. Besides, setting up an AI infrastructure and integrating it into the existing information systems can be challenging, especially for less experienced workers (Pandl et al. 2021). According to Lopez Garcia et al. (2020), the adoption of cloud computing infrastructures is still limited. Besides, the IaaS offers are considered too complex and heterogeneous to be efficiently exploited by end-users (Caballer et al. 2018). A step ahead from low-level IaaS offerings is needed to exploit the potential of cloud computing. It is required to provide users with high-level tools, like a Platform as a Service stack, which can implement the flexibility to deliver tailored solutions to satisfy the business needs while also providing transparent exploitation of the infrastructure resources (Salomoni et al. 2018).

2.3.2 AI platform services

Machine learning "as-a-Service" can be considered one of the most demanded services when large-scale computing infrastructures are adopted (Lopez Garcia et al. 2020). AI platform services abstract some of the complexity associated with AI infrastructure services. Abstraction enables users to achieve shorter time-to-market of their AI applications because they do not have to start from scratch and spend a lot of time planning, developing, and setting up the required hardware or developer tools (Javadi et al. 2020). Several commercial platforms exist, such as Google Cloud AI and Google Hub AI; Azure Machine Learning Service and Studio; Model Asset eXchange from IBM; and AWS Marketplace – Machine Learning (Lopez Garcia et al. 2020). These platforms provide different levels of functionality and various AI services. Most of the commercial platforms include the ability for users to access resources for developing, training, testing, and deploying ML models. Figure 6 (adapted from Visma webinar 2021) shows the abstraction levels, from fully customizable platform environments to pre-build AI software services.

Figure 6 – AI platform abstraction levels (adapted from Visma 2021)

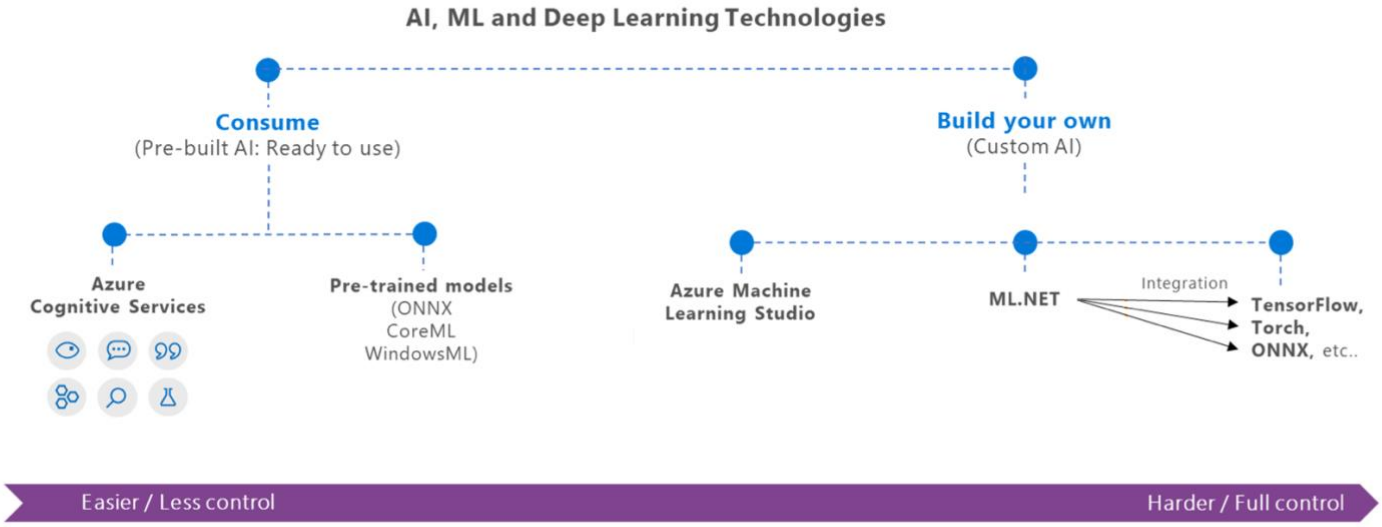


Development tools and running environments that assist AI developers in implementing code to bring out AI capabilities, such as Cloud ML-dev Environments, enable businesses to develop fully customized solutions. The development toolkits and running environments can be adjusted automatically based on business demand (Wang et al. 2018). AI platform services also provide various components that support application development and deployment, such as database systems (Lins et al. 2021). Clients of AI development tools and running environments can focus on creating and training deep learning and machine learning models adapted to their data and use cases. For instance, cloud-based distributed deep learning platforms and advances in hardware enables deep learning to scale and can enable developers and customers to share deep learning infrastructure. Distributed cloud-based software platforms can handle the scheduling, orchestration, elasticity, and resilience of deep learning jobs. Such cloud platforms also aim to reduce the barrier to entry by enabling developers to focus on training neural networks and choosing hyper-parameters rather than focusing on installation, configuration, and fault tolerance (Boag et al. 2018). As the AI solutions are trained from the very beginning, users have complete control over the algorithms and the data. Control and customizability over AI model configuration enable knowledgeable users to build higher quality models because feature, model, and parameter selection can significantly impact the performance of a machine learning task (Yao et al. 2017). In addition, the platforms can provide tools for insight into how an ML model and the data affect the decisions and the results the model provide (Wang et al. 2018). As such, it is easier to determine whether the ML model is fair and accurate (e.g., Thiebes et al. 2021). However, while developer services enable customizability and system control, they also require higher domain knowledge and might not be suitable for all users (Elshawi et al. 2018). For instance, successfully optimizing feature, model, and parameter selection requires overcoming significant complexity that is difficult without in-depth knowledge and experience (Yao et al. 2017). Pandl et al. (2021) found that AI developer services imply a challenging deployment process and thus require cloud-based AI services specialists.

Some AI platform services abstract additional layers of complexity. For instance, AutoML Platforms and open-source AI frameworks that comprise various tools for exploiting algorithms (e.g., TensorFlow) reduce users' efforts in designing, developing, training, and deploying data analytics models. These services reduce the need to learn complex algorithms and technologies and are often available via API access and typically priced based on the number of API calls (Elshawi et al. 2018). AutoML Platforms thus remove the need for installation, maintenance, and related management problems. AutoML Platforms is, therefore, a middle ground between flexibility and complexity. However, while AutoML Platforms require less knowledgeable and experienced users, AutoML faces data and model application problems. Businesses may have customized requirements for the machine learning process, for example, for which only a specific type of data processing tool can be used. Such requirements cannot be met in a black box AutoML solution (Karmaker et al. 2021).
An example of an industry-leading AI service provider is Microsoft Azure. Microsoft Azure provides a cloud-based environment for either constructing customized data analytics workflows or acquiring pre-built, ready-to-use AI applications. The spectrum of Microsoft's AI, ML, and Deep Learning technology offerings is visualized in Figure 7. The figure depicts the trade-off between easy to use and lack of control of pre-built AI applications versus more complex, custom AI development tools which enable more user control. A comprehensive set of tools for data scientists is available in Azure Machine Learning Service. The Azure AI platform also provides data scientists with a Web-based machine learning IDE for creating and automating machine learning workflows (Elshawi et al. 2018). The custom AI services enable IT professionals to build and deploy AI solutions and create machine learning models with tools like Jupyter Notebooks, Visual Studio Code, and open-source frameworks like TensorFlow and PyTorch for domain-specific use cases and business problems. The finished products of custom AI development are typically published as web services or applications

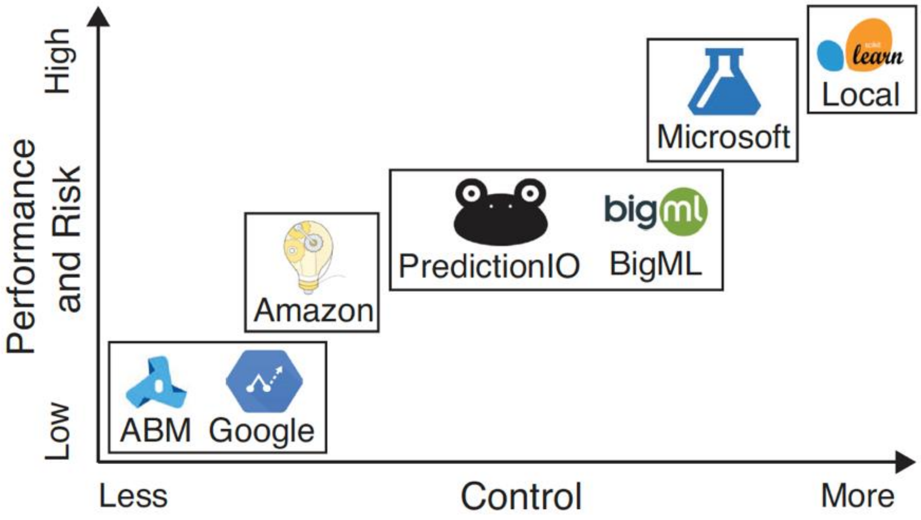
for end-users. End-users can then send inputs and receive back the results of an ML process, such as predictions and classifications (Javadi et al. 2020).

Figure 7 – Microsoft's AI platform: ease of use vs. control of ML models (adopted from Hyon and Apolinario 2019)



Since custom AI software is developed for a single business, it needs to satisfy the business' specifications and expectations (Ransbotham et al. 2017). Furthermore, Yao et al. (2017) imply that more user control over the platform and solutions suggests more performance gains but also higher underperforming risk or poorly performing models due to bad configuration settings (figure 8). Besides, when users are required to upload the training data onto the provider's platform, it risks feeding low-quality input data into an AI model's training process, which will largely influence how the AI-based systems will behave (Thiebes et al. 2021). Training an ML model for an application, decisions regarding hyper-parameters related to the model structure, optimization algorithm, and data pre-processing operations are vital for optimizing the performance of the AI algorithm. All hyperparameter tuning algorithms work via empirical trials (Wang et al. 2018). Hyper-parameters are critical for the convergence of the training algorithm and the final model performance. However, Wang et al. (2018) imply that manual hyperparameter tuning requires rich experience and is tedious.

Figure 8 – Overview of control vs. performance/risk trade-offs in MLaaS platforms (adopted from Yao et al. 2017)



As MLaaS allows users to upload their data, run various data analytics or model building processes, and deploy the trained models to their own services, researchers have found interest in the potential vulnerabilities of such machine learning services (Truex et al 2019). The adoption of AI cloud services raises serious privacy concerns in applications where the risk of data leakage is not acceptable (Rouhani et al. 2018). A barrier to adopting AI services thus concerns data and model privacy. On the one hand, AI models are usually trained by allocating significant computational resources to process massive amounts of training data. Hence, the trained models are considered companies' intellectual property, which requires confidentiality to preserve the competitive advantage. On the other hand, clients do not desire to send their private data in plain text to cloud servers due to the risk of information and training data leakage (Rouhani et al. 2018). Similarly, MLaaS systems that allow users to train models on potentially sensitive data and charge others for access on a pay-per-query basis motives Tramér et al. (2016) to study model extraction attacks due to the tension between model confidentiality and public access. Further, Hanzlik tries to mitigate these challenges by proposing MLCapsule, a guarded offline deployment of MLaaS, which executes the machine learning model locally on the user's client so the data never leaves the client.

2.3.3 AI Software Services

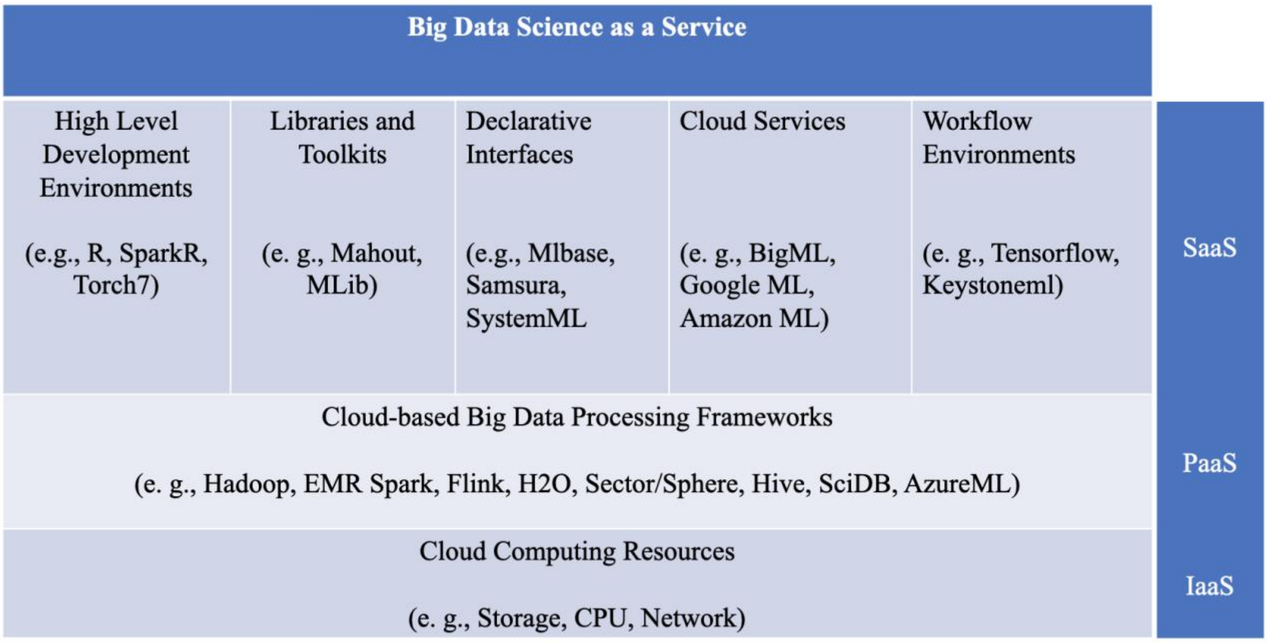
The cloud computing model Software as a Service is a cloud delivery paradigm in which the software is hosted off-premises, typically on a third-party vendor's server, and delivered to the user via the web (Godse and Mulik 2009). AI delivered through SaaS refers to off-the-shelf AI tools and pre-trained models provided to customers commercially (Javadi et al. 2020). Lins et al. (2021) relate AI models already trained by the cloud provider as inference as a service (e.g., prediction APIs). AI software services includes access to a range of pre-built models and related services, such as SaaS data analysis frameworks. SaaS data analysis frameworks are tailored for implementing data analysis and machine learning applications as compositions of high-level services to reduce the programming burden and complexity (Elshawi et al. 2018). AI software services also shift the complexity from the user to the provider by supporting users through automatic hyper-parameter tuning and automating the selection of an optimal classifier (Lins et al. 2021). As pictured in figure 6, AI software services are described as AI use cases and API services. Access to AI services and models is offered via the provider's APIs, which facilitate the interaction between systems (Zapadka et al. 2020), send input data to the applied models, and return the results (Javadi et al. 2020). However, service providers' offerings tend to be fairly generic, and examples include object detection, text to voice synthesis, facial recognition, and text translation (Javadi et al. 2020). The users know only about the input and output formats of the respective API, but the model and the dataset on which the model is trained remain private to the provider (Truex et al., 2019).

Purchasing off-the-shelf AI applications can be quick and convenient, but on the other hand, generic ML models may not suit all use cases and businesses. AI software services are deemed insufficient for specific use cases and large firms operating in a competitive environment. Many extant AI models appear as black boxes and are not easily understandable. Opaque data processing is a challenge associated with of AIaaS. Opaque data processing makes it challenging to understand and audit the results or predictions (Truex et al. 2019). With AI services based on pre-trained models, users do not know which data instances, which model architecture, and which model hyperparameters were used for training that AI model. When a company has no choice but to build from scratch, it is typically due to cases when the data, or the models, are extremely sensitive or proprietary or when commercial tools are not available (Korolov 2019). Every AI problem and use case is completely unique, which requires minute customization. For most companies, this rules out most off-the-shelf commercialized or use case-based solutions that lack the specificity required to achieve significant business impact in the long term.

2.4 Data management and regulations

Data and analytics are becoming the new norm for vertical knowledge. Cloud services provide a simple and easy way to share data, create shared workspaces, and bring cross-functional teams together (Stanciu et al. 2021). Additionally, big data analytics (BDA) capabilities are an organizational requirement for developing sophisticated AI systems (Elshawi et al. 2018). Data management and processing systems such as the NoSQL databases have in recent years increased in prevalence due to its efficient mechanisms and techniques to store and access complex data used to train and create sophisticated AI solutions. The increasing data analysis requirements of several application domains have created a need for designing and building data science tools that can analyze massive amounts of data to elicit valuable information. The three main cloud service models support the analysis and learning from data and are depicted in figure 9.

Figure 9 – Big Data Science as a Service stack (adapted from Elshawi et al. 2018)



Another aspect of data management of AI services is the direct link to cloud storage (Kaplunovich and Yesha 2017). In cases where organizations already use the cloud or are cloud-native, AI services can directly access the data on the cloud, which decreases the effort required to use AI services. However, the provision of necessary computing power and storage capacity and orchestration of various infrastructure components must be sufficiently managed since large-scale data is commonly distributed (Lopez Garcia et al. 2020). Most users have domain knowledge in specific fields, lacking technological or infrastructure knowledge. Therefore, support by the infrastructure layer must break down the complexity of the task and allow scientists to focus on their respective activities, such as modeling the problem and evaluating and interpreting the results of the algorithms. Besides, when expert knowledge or domain knowledge is required to create high-quality training data, it can become costly and time-consuming (ibid). However, companies with a significant scale can use their data to build capable solutions in strategic areas and often have the finances to develop and equip employees with the required skillset (Mikalef et al. 2018).

Literature on AI services frequently highlights the legal, regulatory, and policy challenges AI cloud services present. For instance, some relate to the difficulties of assigning legal responsibilities and liabilities in these complex processing environments. Others relate to the general implications of these powerful technologies working to determine and drive customer applications' functionality (Cobbe and Singh 2021). Pandl et al. (2021) pose that AI services may, in some cases, not be compliant with corporate governance guidelines or regulations. Rouhani et al. (2018) assume that some organizations are concerned about whether a cloud provider has implemented adequate data governance and protection mechanisms to ensure that collected data and AI-generated data about individuals are not used to impede their privacy. Cloud service providers must implement adequate data governance and protection mechanisms for collecting data and AI-generated data about individuals to adhere to privacy policies to gain customers' trust (Thiebes et al. 2021). Trustworthy Artificial Intelligence (TAI) is increasing in relevance in electronic markets and its research community (Thiebes et al. 2021). Research on achieving TAI does not only cover AI-related research domains, such as ethical computing or human-computer interaction, but also information systems, marketing, and management have focused on achieving trust in AI technology adoption. TAI is based on the idea that trust builds the foundation of societies, economies, and sustainable development. Further, organizations will only be able to realize the full potential of AI if trust is established in its development, deployment, and use (Thiebes et al. 2021).

2.5 Cost-efficiency

Cloud-based AI services are assumed to play a growing role in society's technological infrastructure and enable, facilitate, and underpin functionality in many applications. Cloud services and the complementary AI capabilities accelerate time-to-value and time-to-insight in a changing business environment (Stanciu et al. 2021). Besides, a primary generator of cost-saving innovation comes from technologies encapsulating cross-stack software capabilities, faster hardware, and improved energy consumption levels. Stanciu et al. (2019) suggest that transforming the business model into an asset-light operating model contributes to a consistent cash flow while locking capital investments and capital expenditures, which is critical for cost savings. Managed services, such as IaaS, PaaS, and SaaS, are the first shift from a capital expenditure model (CapEx) to an operating expenditure (OpEx) model. Additionally, the innovation factors of cloud services facilitate cost savings due to the elimination of overprovisioning capacity, lowering the power consumption, and increasing utilization with pay-as-you-go models. On the contrary, cloud-based AI services introduce challenges related to difficulty estimating total cost and the ROI of AI projects (Pandl et al. 2021). Oppose to prior research that emphasizes budget flexibility, cost-effectiveness, and cloud providers' efforts to ensure transparent service offerings, the authors highlight that adopting AI services can be challenging, especially for estimating project overhead costs before embarking on a project. While cloud systems are connected to the internet, making AI services systems highly compatible with various information systems, legacy systems may not support the connection to or integration of cloud technologies (Pandl et al. 2021). As a result, legacy systems can potentially increase maintenance costs, lack compliance with governmental regulations, and create data silos that prevent integration between systems (Serrano et al. 2014). The lack of interoperability between legacy systems and cloud technologies, and the subsequent creation of data silos can potentially result in higher costs due to redundant IT and application infrastructure.

AI services may not work well with systems that need to infer decisions in real-time due to potentially high latency (Romero et al. 2020). AI cloud services rely on statistical computations, meaning that deeper exploration yields more accurate results but requires more processing time to perform (Halpern et al. 2019). As a result, cloud providers and developers are forced to make a trade-off between the service's result accuracy and responsiveness. On the one hand, increasing the service response time can provide better results, which is critical in some domains (e.g., healthcare and finance). However, there are responsiveness-critical application domains, such as e-commerce, where slow user experiences can lead to poor customer engagement, subsequently decreasing commercial value.

2.6 Automation

AI services enable users to optimize ML models automatically by selecting the most suitable hardware architectures and handling hardware and software failures in an automated manner. Lins et al. (2021) suggest that automation in AI software, developer, and infrastructure services benefits users by enabling the deployment of AI technologies faster and with higher technical robustness while having little prior knowledge about AI. The authors also state that automation removes the need to invest in AI engineers. Besides, the combination of cloud computing and AI leads to unique advantages over in-house AI solutions. Cloud-based AI services abstract a large amount of complexity compared with in-house AI. Cloud computing accelerates the development of AI compared to development on a traditional technology stack because tools, resources, and components are readily available. On the cloud, resources are interconnected by nature, which removes the need for external libraries and frameworks. Automation thud abstracts the complexity of AI and relates to both the hardware and the software layer. Most people look for automation benefits when deciding on cloud-based AI services (Pandl et al. 2021). On the hardware layer, AI services offerings typically rely on cloud infrastructures where the cloud provider automatically manages hardware resources, for example, maintenance or scaling (Romero et al. 2020). On the software layer, users only need to call APIs to access the service and perform tasks, for instance, along the ML pipeline. Some AI services are capable of automatically pre-processing data, selecting appropriate ML model architectures for a given data set, or tuning hyperparameters for an ML model automatically (Yao et al. 2017). Thus, organizations need less AI expertise, which can be rare and expensive to hire (Truex et al. 2019).

Based on automation, Pandl et al. (2021) found that the on-demand availability of computing resources is the most frequently mentioned driving factor for AI services adoption. A dominant advantage of AI services is scalability. Cloud providers can elastically provision and release hardware resources available to the platform and thus scale horizontally per the user-defined configurations and requirements (Elshawi et al. 2018). Cloud-based AI services are perceived as more resilient than in-house AI applications due to automated handling of infrastructure and software stack failures and sub-services on which the AI product depends (Bhattacharjee et al. 2017). Preventing such failures and achieving effective recovery is crucial for AI-based systems. For instance, training a deep neural network with a large dataset may take days, and losing the progress due to failure could be critical and expensive.

2.7 Summary

Cloud computing equips AI with tremendous power and is an essential catalyst for developing innovative and intelligent applications. The time, therefore, seems ripe for companies to capitalize on AI. However, enterprises cannot realize the benefits of AI services with just a push of a button on a cloud platform. It thus remains to identify the barriers and drivers for adopting AI services and the critical factors enabling businesses to succeed with cloud-based AI services. Although several of the reviewed papers are based on technical issues, they also provide insights into the challenges and benefits of the various cloud deployment models that are valuable in an adoption and use perspective. Table 1 depicts the identified research concepts and refined problem definitions further investigated through the empirical data collection. The refined problem definitions aim to provide answers to the proposed research questions.

Table 1 – Literature concepts and refined problem definitions

Literature concepts	References	Refined problem definitions
The complexity of developing customized AI solutions through infrastructure and platform services	Elshawi et al. (2018), Yao et al. (2017); Mikalef et al. (2018); Pandl et al. (2021); Boag et al. (2018); Wang et al. (2018); Duong and Sang (2018)	What professional resources are required to set up, manage, and maintain a cloud environment and develop customized AI solutions?
Performance, accuracy, and control of AI applications and ML models for specific use cases	Javadi et al. (2020); Yao et al. (2017); Halpern et al. (2019); Wang et al. (2018)	How can businesses ensure stability and robustness of customized AI solutions in a cloud environment?
Data management and data privacy protection	Rouhani et al. (2018); Thiebes et al. (2021); Cobbe and Singh (2021); Pandl et al. (2021); Lopez Garcia et al. (2020)	What are the main concerns of data management in a cloud environment?
Automation and scalability benefits of the cloud platform	Stanciu et al. (2021); Lins et al. (2021); Elshawi et al. (2018); Pandl et al. (2021)	How can businesses automate development of AI solutions on the cloud and scale the solutions after they are deployed into production?

3. Theoretical framework

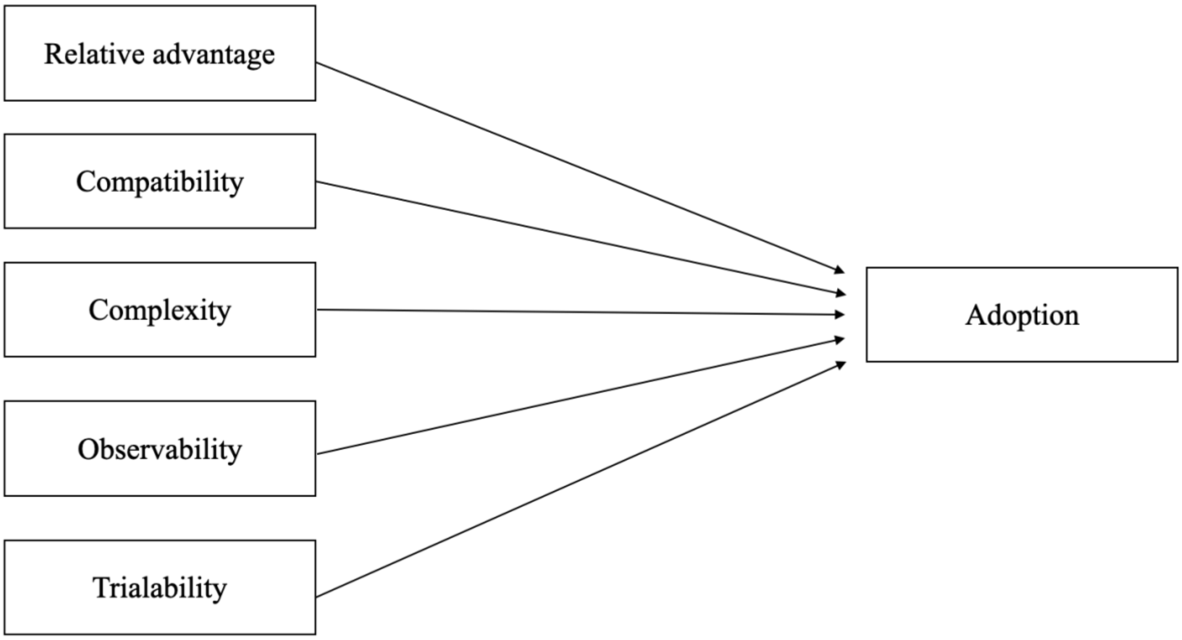
A theoretical framework and method explicitly structure the thinking about the research topic and the process undertaken. The theoretical background in this study is based on the Diffusion of Innovations theory (DOI) proposed by Rogers (1995) and the Critical Success Factor (CSF) method by Rockart (1979). Building on the DOI theory enables a deeper understanding of the inhibitors and drivers influencing an organization's adoption process and thus its ability to harness AI capabilities. Rather than focusing merely on adoption of the technology, this study also apply the Critical Success Factors method which guides and forms a basis for the choice of research methods. Articulating this study's theoretical background and method also forces the researcher to address why and how questions and helps generalize the various aspects of the studied phenomenon.

3.1 The Diffusion of Innovation theory

The DOI theory has been widely used in information systems research and related research domains. The theory allows potential adopters to evaluate an innovation, the perceived efficiencies gained by the innovation relative to current tools or procedures, its compatibility with the pre-existing system, its complexity, such as difficulty to learn and associated pitfalls, its trialability or testability, its potential for reinvention, and its observed effects. Innovation is an idea, object, or practice perceived as new by the members of a social system (e. g., an organization). The social system consists of individuals or organizations that share a common culture and are potential adopters of the innovation. Rogers defines diffusion as "the process in which an innovation is communicated through certain channels over time among the members of a social system" (p. 5). DOI theory proposes five major characteristics of an innovation that influence organizations' adoption intentions: (1) relative advantage, (2) compatibility, (3) complexity, (4) observability, and (5) trialability (Rogers 1995). Relative advantage is the degree to which an innovation is perceived as being better than the idea it supersedes. Rogers' theory suggests that innovations with a clear, unambiguous advantage over the previous approach will be more easily adopted and implemented. Compatibility is the degree to which an innovation fits with the existing values, past experiences, and needs of potential adopters. Extant research suggests that the more compatible the innovation is, the greater the likelihood of adoption (Greenhalgh et al. 2004). Complexity is the degree to which an innovation is perceived as challenging to understand and use. Observability is the degree to which the results of an innovation are visible to the adopters. If there are observable positive outcomes from the implementation of the innovation, then the innovation is more adaptable. Finally, trialability is the degree to which an innovation may be experimented with on a limited basis. Because innovations require investing time, energy, and resources, innovations that can be tested before being fully implemented are more readily adopted. If an organization has systems in place and appropriate skills to monitor and assess the impact of innovation (both anticipated and unanticipated), the innovation is more likely to be assimilated and sustained (Greenhalgh et al. 2004).

The DOI theory proposes an applicable framework for addressing and clustering findings into the above categories to understand the drivers and inhibitors for adopting AI cloud services.

Figure 10 – Diffusion of Innovation Theory (adapted from Rogers 1995)



3.2 The Critical Success Factor method

The Critical Success Factor (CSF) method has been adopted in numerous Information Systems studies since it was proposed in a Harvard Business Review report by John Rockart in 1979. An extensive amount of research in the IS domain aims to understand the key factors that enable organizational- and, more specifically, IS success (e. g., Delone and McLean 1992). Bullen and Rockart (1981) suggest that CSFs are "the limited number of areas in which satisfactory results will ensure successful competitive performance for the individual, department or organization." A widely accepted definition of CSFs is "those few things that must go well to ensure success" (Boynton and Zmud 1984, p. 17). The concept of identifying and applying CSFs to business problems is not a revolutionary field of work. It dates back to the original concept of “success factors” in management literature by D. Ronald Daniel in the

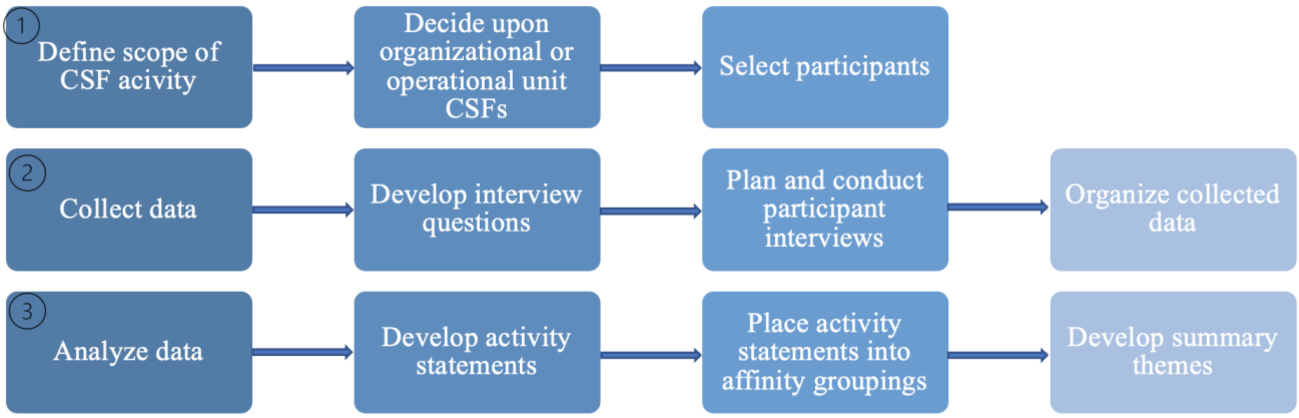
1960s. Most of the work in success factors performed by Rockart and Daniel was focused on refining the information needs of executives. However, as a natural outgrowth of this work, Rockart (1979) suggested the usefulness of the method as a component of strategic planning for information systems or technology (Rockart 1981). The CSF method has found its way into many formalized information or business systems and technology planning methodologies and the CSF concepts and approach are still to this day applicable to many of the challenges presented in the IT fields.

Some researchers have argued that CSFs are not free from bias and may not depict the actual situation (Boynton and Zmud 1984). The CSF method is not a proven mathematical model that guarantees precision in computation. Additionally, the method's validity has been questioned due to the risk of researcher and respondent bias. However, the CSF method facilitates a structured, top-down analysis and planning process. The CSF method enables the researcher to focus on identifying strategic planning and managerial control information required by top management to orient the organization on its current and future activities. A skilled analyst must direct the CSF method. However, the same can be stated about all nonautomated information systems methodologies (Boynton and Zmud 1984). The CSF method shaped the data collection and analysis methods, which are further elaborated in the methodology sections.

3.2.1 The Critical Success Factor method activities

This study performs three basic activities for deriving CSFs, adapted from Caralli (2004). The process of creating activity statements, affinity groups and supporting themes are elaborated in the methodology section.

Figure 11 – Overview of the CSF method activities (adapted from Caralli 2004)



3.2.2 Defining success criteria

One of the ways in which IT departments have addressed challenges is by involving the organization at large in their strategic planning process. To determine priorities, these efforts also include a direction-setting activity such as developing CSFs (Caralli 2004). CSFs define key areas of performance that are essential for the organization to accomplish its mission. Managers implicitly know and consider these key areas when setting goals and directing operational activities and tasks crucial to achieving goals. From a Project Management perspective, CSFs have been described as "characteristics, conditions, or variables that can significantly impact the project's success when properly sustained, maintained or managed" (Milosevic and Patanakul 2005, p. 183). When success criteria are formally defined and then measured, IT project outcomes are improved, and project resources are better utilized (Thomas and Fernández 2008). For instance, Wateridge (1998) suggests that projects must "achieve their main objectives, be commercially profitable for the contractor, and achieve its business purpose in three ways (strategically, tactically, and operationally)" to be considered successful. Success is a critical concept when trying to foretell the future of a project (Christenson and Walker 2004). However, success and failure are difficult to define and measure since they mean different things to different people. Thomas and Fernández (2008) found that it is the act of defining and measuring IT project success that may contribute to success itself. Based on their analysis, three effective practices were identified: (1) an agreed definition of success, (2) consistent measurement, and (3) the use of results. Hence, by knowing what to look for, tracking progress, and being willing to alter the path, chances of finding success are better.

4. Methodology

This section presents the research design, research strategy, the data collection processes and methods, and the qualitative method for analyzing the collected data. This study aims to inductively build theory by answering research questions rather than deductively proposing and testing specified hypotheses. The research design is aimed to achieve a suitable combination of theory and practice for the research results to be utilized by practitioners. Accordingly, the qualitative interpretive research paradigm is adopted in this study.

4.1 Research design

The research design is a plan to answer the research questions, while the research method is a strategy used to implement that plan. A clear research design enables to relate empirical data to the initial questions under investigation and connects it to the study's conclusions, which minimizes the risk of collecting and analyzing irrelevant data that does not satisfy the research questions (Yin 2009). This thesis is carried out through a combination of a literature review and empirical research comprised of an in-depth case study of a large enterprise operating in the transportation and logistics industry. First, a review of extant literature was carried out using a systematic methodology adopted from Webster and Watson (2002). The literature review synthesizes existing knowledge and highlights relevant issues related to this research topic and the current research gaps that guide the subsequent empirical research.

4.2 Research strategy

In addition to the literature review, an initial informal conversation with a representative from the case organization contributed to forming the research questions. The decision to select a single case study as a research strategy is based on the problem definition and the available literature. Darke et al. (1998) state that case study research is the most widely used qualitative research method in IS research. Myers (2015) found that qualitative research is regarded by some of the best scholars in the IS field as being of high quality. According to Thomas (2011), case studies can include analysis of persons, events, decisions, periods, projects, institutions, or any other systems studied holistically through one or more research methods. The unit of analysis in this case study is completed, and ongoing successful AI projects are carried out through the Microsoft Azure cloud service, Azure Machine Learning, a Cloud ML-dev Environment. This study applies an exploratory case study approach. According to Oates (2006), an explorative case study is typically used where there is inadequate literature about a topic.

As previously discussed, there are insufficient empirical studies investigating AI services adoption drivers and inhibitors, and CSFs for harnessing AI capabilities through cloud services. Case studies are appropriate when the researcher has less prior knowledge of constructs and variables (Cavaye 1996). Single in-depth case studies are a useful way of representing unique cases when exploring a new phenomenon and when there is a lack of theory (Eisenhardt 1989). Additionally, Darke et al. (1998) suggest that case study research is suited to understand the interactions between IT-related innovations and corporate circumstances. While single case studies' generalizability is limited (Yin 1994), they can provide an important insight into the direction of any future research.

4.3 Data collection methods

A case study typically combines data collection techniques such as interviews, questionnaires, observation, and documents (Oates 2006). This study conducts semi-structured interviews to allow interviewees to «speak their minds» to discover important elements. Semi-structured interviews are also appropriate because a basic structure is necessary to gather further information based on knowledge from prior research. Interviewing builds a holistic view of informants' knowledge and enables interviewees to express their thoughts on a particular topic (Alshenqeeti 2014). Alshenqeeti (2014) further suggests that interview events are directly "observable." Talking to people is an effective method for attaining and exploring constructs. Moreover, interviews are interactive and can express complete, clear answers and probe into emerging topics. As Oates (2006) implies, interviews are flexible, and the researcher can adjust a line of inquiry as the interview progresses. The interview procedure seeks to help interviewees think about and identify AI services adoption drivers and inhibitors, as well as CSFs for harnessing its capabilities. Following Bullen and Rockart's (1981) suggestions, an interviewer that aims to identify and elicit CSFs should accomplish the following objectives:

1. To understand the interviewee's organization, the interviewee's mission and role (the "world view") within the context of their organization as the interviewee perceives them.
2. To understand the goals and objectives of the interviewee.
3. To elicit CSFs and measures from the interviewee.
4. To assist the manager in better comprehending her own information needs.

The results of CSF studies typically include a list of four to eight CSFs (Rockart 1979). However, former studies have reported substantially more CSFs (e. g., Somers and Nelson 2001). Additionally, previous research has presented CSFs in various formats to facilitate the conceptualization of their findings. For instance, Zahedi (1987) proposed a CSF hierarchy, Somers and Nelson (2001) attempt to prioritize CSFs, while others have provided categorization schemes (e. g., Alazmi and Zairi 2003). To facilitate conceptualization of the findings, the findings are depicted in a conceptual model (figure 15), demonstrating interactions and dependencies of the identified affinity groups.

This study also incorporates method triangulation to allow the findings from one data generation method to be supported or questioned by comparing them with the data from another method. Therefore, data are also generated from document analysis of researcher-generated documents. Researcher-generated documents are put together solely for the purposes of the study. Oates (2006) implies that a researcher can design a document and then ask someone else to complete it. The CSF method initially focuses on a participant's attention on a core set of essential issues and then proceeds to refine the issues to allow an evolving "design" to be continuously examined for validity and completeness (Boynton and Zmud 1984). Accordingly, based on Oates's (2006) and Boynton and Zmud's (1984) suggestions, I compose a list of the initial identified CSF and present it to the professionals from the case company, which help modify and validate the list. In particular, CSFs were identified in a three-step process.

1. The first phase identified CSFs by interviewing four managers involved in the AI projects, three key personnel working hands-on with the technology, and two external IT consultants experienced with AI and cloud computing.
2. The second phase concerned analyzing the data and extracting and refining the individual CSF from the respondents into a collective set of operational level CSFs.
3. In the third phase, the list of CSFs was modified by merging, dropping, adding, and renaming some of the previously identified factors in collaboration with the professionals from the case company.

For a Critical Success Factor to be deemed sufficient and relevant for it to be accepted, at least four informants need to suggest or indicate it. If fewer informants elaborate on a particular phenomenon, supporting literature is required to ensure the validity of the findings. Following the "triangulation of subjects" strategy (Rubin and Rubin 2011), the interviews involved employees and managers with diverse roles and functions and from different departments within the organization. The interviews adhere to the predefined themes according to the interview guide. However, some of the themes are adjusted according to the informants' position and role. Appendix C includes all predefined themes and questions asked during the interviews.

Table 2 – Distribution of interviews

Informant #	Department	Role	Number of interviews	Interview length	Type of industry and size of the organization
i01	BI and Data Warehouse	Program leader of BI and Advanced Analytics	1	41 minutes	- State-owned - Large Enterprise - Transportation and logistics
i02	Data Science	Department leader	1	46 minutes	
i03	IT	Leader of Cloud Platforms	1	32 minutes	
i04	Data Science	Data Scientist	1	30 minutes	
i05	Data Science	Data Scientist	1	42 minutes	
i06	BI and Data Warehouse	Scrum Master and Agile Project Manager	1	37 minutes	
i07	BI and Data Warehouse	Data Analyst	1	67 minutes	
i08	N/A	Individual consultant – Data and AI Associate Manager	1	54 minutes	- Private Sector - Multinational corporation - IT services and consulting
i09	N/A	Individual consultant – AI Product Manager	1	58 minutes	

4.3.1 Sampling procedures

The type of CSFs being developed strongly influences decisions about whom to include in the CSF activity. The scope of the CSF activities in this study are limited to the operational level and some of the organizational areas that are important to the operational unit´s success. For operational unit CSFs, the broad perspective of senior managers in the operational unit is necessary. In Rockhart´s work, he suggests the concept of individual manager CSFs which refers to the CSFs that are important to each manager in the organization. However, depending on a manager´s position in the organization, their CSFs are likely to be the same as the CSFs at the operational unit level over which the manager has responsibility. In lower levels of the organization, line managers, for example, may have their own CSFs, but it is likely that they also reflect most of the CSFs for the operational level (Caralli 2004).

The sampling technique in this qualitative evaluation is purposeful sampling and snowball sampling (Parker er al. 2020). First, four managers from four different units of the organization were asked to be interviewed because they play a particularly important role in the organization's cloud AI projects. Purposeful sampling is a non-probability sampling technique. The researcher deliberately hand-picks the sample and chooses instances likely to produce valuable data to meet the research purpose (Oates 2006). In addition, to recruit more people to participate in interviews, I chose to apply the snowball sampling technique. This technique involved asking the managers for suggestions about relevant workers to recruit for interviews. As a result, three additional informants were interviewed. After seven interviews with the relevant informants from the case company, I received great insights from different perspectives and viewpoints. However, I also reached a state of data saturation. Data saturation refers to the point in the research process when no new information is discovered in data analysis. This redundancy signals that it may not be necessary to collect additional data. However, to minimize bias and get a more comprehensive overview of the topic, I interviewed two experienced professionals within AI management. Including individual consultants outside the target case enables gathering different opinions and perspectives on key issues related to cloud-based AI services.

4.4 Data analysis

The data analysis method describes how I worked from the raw data to the conclusions. Most qualitative data analysis involves text analysis and abstraction of verbal, visual, or aural themes and patterns that help answer the research questions (Oates 2006). It is the primary type of data or evidence generated by case studies. The challenge with qualitative analysis is that it has fewer procedures and is more dependent on the researcher's skills to see patterns and themes within the data (Oates 2006). I try to mitigate this challenge by conducting data coding and analysis techniques to produce contextually grounded findings.

A thematic data coding analysis method are performed to reveal adoption drivers and inhibitors. Thematic data coding and analysis is a method of "identifying, analyzing, and reporting patterns (themes) within data" (Braun and Clarke 2006). It is a frequently used method of data analysis in qualitative data analysis because of the wide variety of research questions and topics that can be addressed. Thematic analysis of transcribed interviews can explore the context at a level of depth that quantitative analysis lacks while also allowing flexibility and interpretation when analyzing the data. The adoption drivers and inhibitors are revealed by coding the interview data, including highlighting and labelling words and phrases that indicates various themes and concepts. Then, an iterative comparison of the codes and the themes aims to ensure validity of the findings.

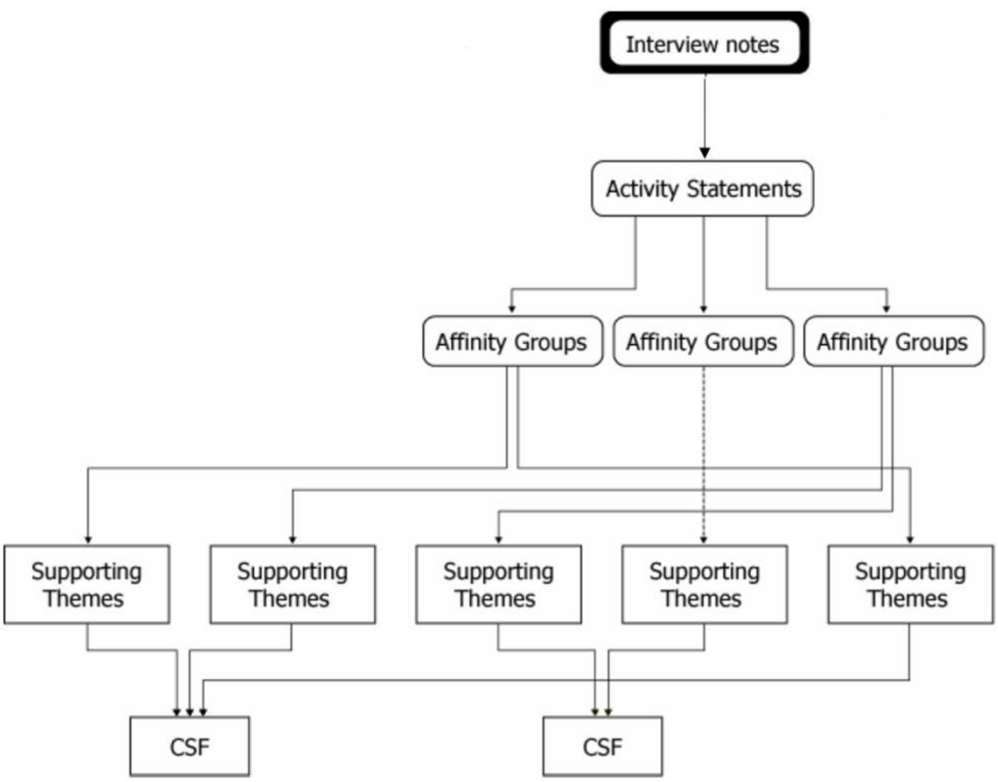
Interviews provide the raw data for deriving CSFs. This information is formed into statements that represent and reflects the activities that managers and key personnel perform or should perform, focusing on, or monitoring. The creation of activity statements is a technique for transforming raw data into manageable, consistent entities that subsequently can be analyzed to derive CSFs (Caralli 2004).

The affinity grouping of activity statements enables to summarize the core thoughts and concepts from managers and key personnel regarding those activities that require significant attention. Affinity grouping enables to categorize data that share common characteristics, traits, or qualities so that a common description of the data can be developed. Caralli (2004) suggests a process for performing affinity grouping of activity statements. This process is depicted in Appendix E.

A final step required before the development of CSFs is to develop supporting themes. Supporting themes represent a group of activity statements and are used as the foundation for creating CSFs. The objective of the supporting themes activity is to identify the underlying concepts or intentions that represent the activity statements in a particular grouping which will eventually represent a CSF. The activity statements and supporting themes that emerged from the raw data are depicted in appendix D.

When supporting themes are developed for each group of activity statements, an additional affinity grouping exercise using the supporting themes is performed. The process aims to bring together similar supporting themes into groups that will result in CSFs. This process resulted in sub-affinity groups.

Figure 12 – Illustration of Affinity Grouping of Supporting Themes (adapted from Caralli 2004)



The analysis method enables uncovering variables and interactions that are not initially anticipated and thus aims to reduce bias. The potential for researcher bias can be reduced by actively involving research participants in checking and confirming the results (Birt et al. 2016). Hence, the analysis was discussed with relevant interviewees before publication to increase the credibility of the analysis through respondent validation.

Figure 13 – Summary of research design

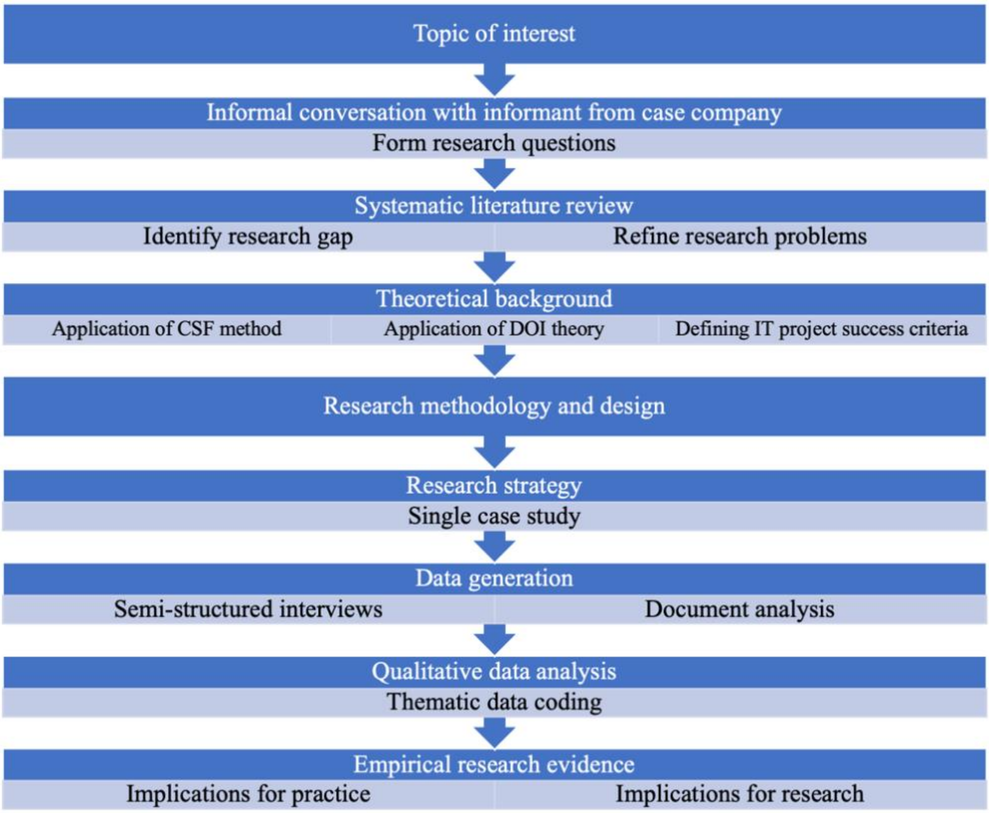
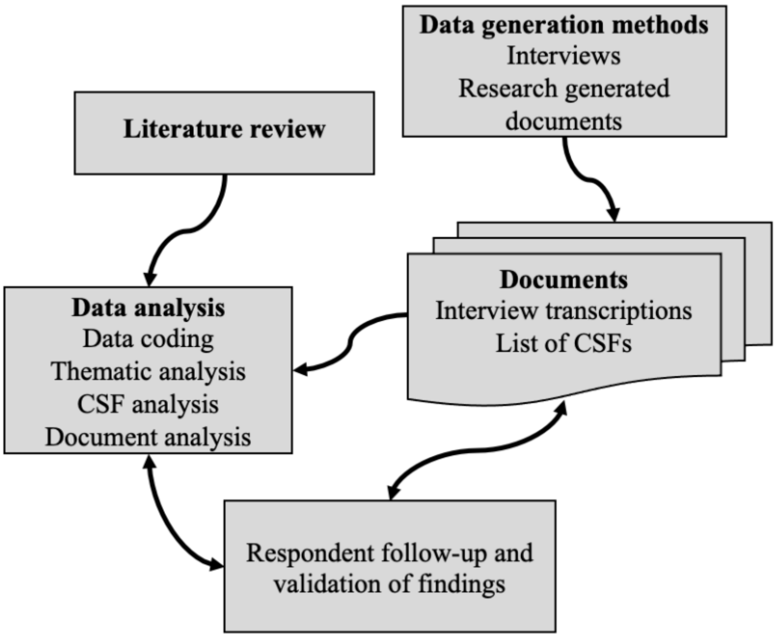


Figure 14 – Data collection and analysis overview



4.5 Target case and background

The Nordic logistics market is becoming increasingly digitalized. Behind the shipping of a single parcel lies a complex value chain and enormous amounts of data. Competition in the logistics industry has intensified, forcing actors to dive into their data to gain valuable insights and enable data-driven decisions. The McKinsey Global Institute has estimated that the transportation and logistics industries have a value potential of $810 billion worldwide. The target case of this study is one of Norway's largest logistics enterprises. The organization can point to tremendous digital adaptability throughout its history. The organization has gathered forces within Business Intelligence (BI) and Data Warehouse in a central department. Business Intelligence is a global collective term that deals with collecting, structuring, and disseminating business information in the form of analyzes and reports. In logistics, this is often related to data within the value chain, such as finances, sales, resource utilization, and production of packages, letters, and goods. In recent years, the company started to combine BI with advanced analytics. While BI is about creating business insight using live and historical data, advanced analytics is related to modern methods such as algorithms, machine learning, and artificial intelligence. Their objective is to create more predictive analyses and form a picture of the future – from describing what has happened to predict what will happen. One of the company's main AI projects is a solution that estimates the delivery time of a parcel. Based on historical data and machine learning, the AI solution predicts how many packages will arrive at their terminals in the coming months and enables online shoppers to know when their packages will arrive at their local pick-up point. The estimates also help to staff more precisely and maintain delivery quality in a cost-effective way.

Prior to the AI initiative, the company's data scientists were not equipped with the suitable tools and infrastructure for machine learning. A former attempt to develop machine learning models on-premises was unsuccessful as they could not access or use the existing infrastructure. Building an internal infrastructure was very demanding, primarily due to the organization's size. External suppliers operated and managed the IT infrastructure as in most large organizations. Outsourcing the IT infrastructure implies that people need to order the required hardware and computing resources from an outsourced supplier. The service thus depends on the suppliers' regimens and competence profiles. In addition, data scientists lacked the competence to use suppliers. The project thus became too challenging to realize. As a result, the company turned to the cloud. They aimed to be an early adopter of AI and cloud technology, at least in their market space. Therefore, together with Microsoft, the initiative began with a hackathon in 2018, aiming to explore the available opportunities. A team of data scientists started to explore and identify what infrastructure they needed to take advantage of the cloud and connect to the data warehouse to access data. Fast forward, the adoption project was successful, and they started production of AI in the cloud as early as 2019. Azure enables the company to design, train, deploy, manage, and maintain customized AI solutions without suppliers. In addition to easy, on-demand access to the suitable hardware to run machine learning models, proper tools, libraries, and frameworks are also highly accessible, with modern programming languages well-known to the data scientists. Today, two departments and four interdisciplinary teams are involved in the AI projects, including professional resources such as data engineers, data analysts, IT architects, and business leaders. The AI solutions have had a remarkable impact on the organization's innovative solutions, which have increased their digital maturity, customer satisfaction, and competitive advantage.

5. Findings and discussion

The following chapters present and discuss the findings, aim to answer the research questions, and the refined problem definitions based on the literature review. The findings and discussion sections are merged to discuss and analyze each set of the findings. The findings are also compared and discussed against the literature review and additional research papers.

5.1 Drivers and inhibitors of AI services adoption

Based on the DOI theoretical framework, findings from the literature review, and empirical data collection, this study has identified 17 inhibitors and ten drivers of AI services adoption.

Table 3 – Drivers and inhibitors of AI services adoption

DOI characteristics	Inhibitors (-) and drivers (+)	Source
Relative advantage	Customizability (+)	Lit, Int
	Control over data and AI solutions (+)	Lit, Int
	On-demand availability of computing resources and access to specialized hardware (+)	Lit, Int
	Direct link to cloud storage (+)	Lit
	Elastic provision of hardware resources and horizontal scalability (+)	Lit, Int
	Project and technology flexibility (+)	Lit, Int
	Cost-effective utilization of specialized hardware resources (+)	Lit, Int
	Data and privacy concerns (-)	Lit
Observability	Lack of measurable results (-)	Int

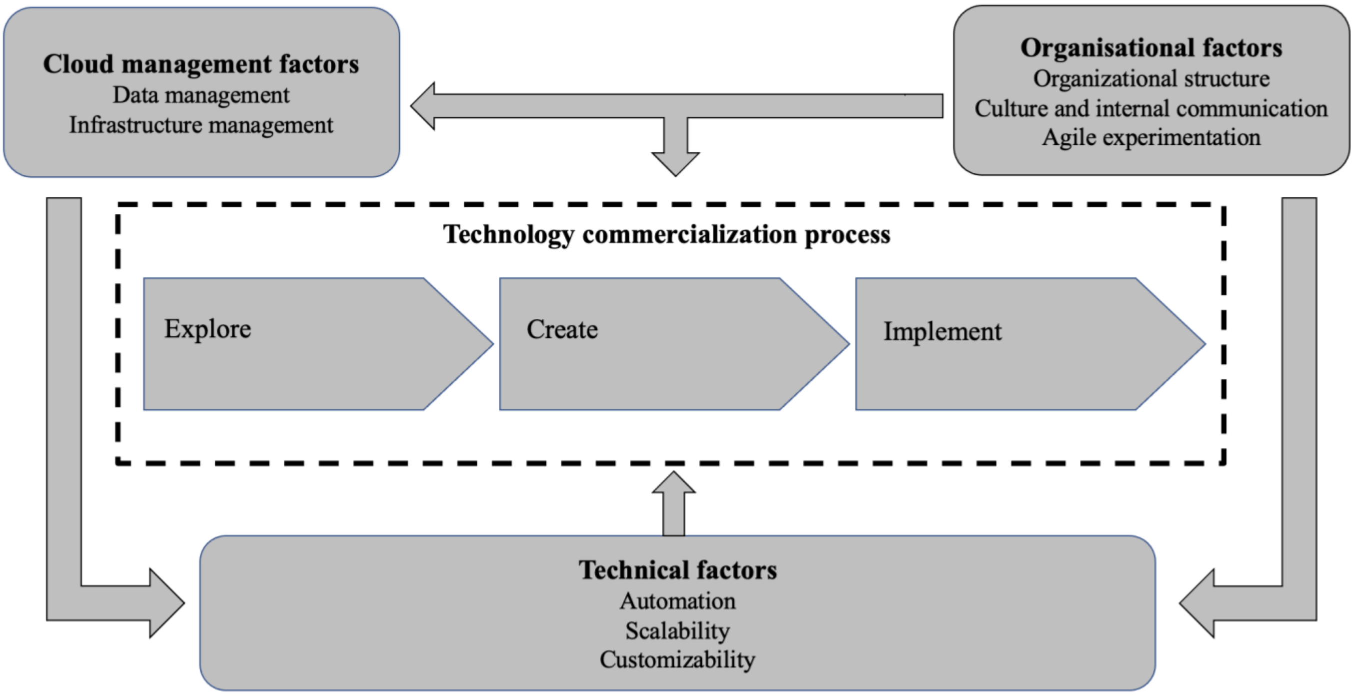
	Difficulty in estimating the total cost of a project (-)	Lit, Int
	Difficulty estimating ROI (-)	Lit, Int
	Lack of proof of economic and commercial benefits (-)	Int
	Lag between the release of new, innovative AI services offerings and the skills and awareness of users (-)	Int
	Fast time to value (+)	Lit. Int
	Several cheap cloud resources and services can become costly in the long-term (-)	Int
	Trade-off between time to value and data silos and maintenance costs (-)	Int
Compatibility	Network latency (-)	Lit
	Lack of responsiveness of resources located outside Europe (-)	Int
	Challenging integration process of the existing infrastructure/legacy technologies and the cloud environment (-)	Lit, Int
	Difficulties accessing workers with the required IT expertise and AI services specialists with domain expertise (-)	Lit, Int
	Sophisticated AI systems require massive amounts of data (-)	Lit, Int
Complexity	Challenging configuration processes and set-up, management, and maintenance of the AI services system (-)	Lit, Int
	GDPR compliance challenges related to the location of data and use of collected data for new AI solutions (-)	Int
	Uncertain governance and regulation compliance (-)	Lit
	Difficulties in assigning legal responsibilities and liabilities (-)	Lit, Int
Trialability	Opportunities to prevent vendor lock-in (+)	Int
	Subscription- and test of technologies for the time required (+)	Lit, Int

The identified drivers and inhibitors of AI services adoption are further elaborated in the next sections in addition to a presentation and discussion of the CSFs for harnessing AI capabilities through cloud services.

5.2 Critical success factors for harnessing AI capabilities though cloud services

The empirical research findings emphasize three main pillars for AI services success. Figure 15 illustrates the synergy between organizational, technical, and cloud management factors and includes fundamental concepts and prerequisites that enable AI services success and maximized organizational value. The model shows that the organizational factors influence both cloud management and technical factors because an anchored communication, innovativeness, knowledge, and team autonomy influence operation and management of the cloud and the organization´s ability to scale, automate, and customize AI solutions. Cloud management influences technical factors because processes within data management ensure that data is accurate, available, and accessible, and infrastructure management ensure that the cloud environment is autonomous and business-specific to exploit the compute infrastructure to develop custom, scalable, and automated AI solutions. These three pillars subsequently affect the technology commercialization process. The technology commercialization process describes what factors are important in the stages from idea to production of an AI solution in a cloud environment and has naturally a significant impact on success or failure. The following sections will describe each unit that serves as a basis for the final list of CSFs.

Figure 15 – Dependencies and interactions between affinity groups and sub-affinity groups at the operational level



5.2.1 Organizational factors

An organization can either facilitate or hamper the technology commercialization process due to several internal characteristics, including its resources, personnel, and patterns of communication and decision-making. Some of these characteristics are formal and represent how an organization divides its labor into distinct tasks, while others are informal through behavioral patterns.

Organizational structure

Organizational structures are central to a successful team. Schultz et al. (2013) propose that an organizational structure based on formal control may increase innovative performance by enabling coordination among different functional units, increasing cost-effectiveness, decreasing uncertainty, and minimizing mistakes. However, evidence from this research signifies that a low degree of formalization, no standardization of products or processes, and decentralized decision-making for inter-and intra-team activities are fundamental to facilitating innovation. These findings are aligned with Mintzberg's (2003) research. Mintzberg proposes that an organic structure made up of ad-hoc project teams and mutual adjustment between teams, without the need for formal coordination of roles, is strongly connected to providing innovation. Surprisingly, a prerequisite for success proposed by all target case informants is that the data scientists and the data science department is separated from the other departments. Building the data science department on the side of other departments has enabled autonomy, adhocracy, and trust. For instance, "it is inherent in the data science department that they can try new things, experiment, and fail, and they do not have anyone telling them what to do" (i05). Lifshitz-Assaf et al. (2018) also argue that innovation requires a high level of work autonomy. Autonomy has encouraged curiosity, enabled independent thinking, and provided an environment where employees can experiment and test new problem-solving approaches with minimal fear of failure. Data science is iterative in nature, and projects cannot be ordered externally. It is essential to provide data scientists with room for experimentation. "Separating the departments and working with AI as a program has resulted in an increased focus on the AI projects, which has been important for success" (i01). The separation of the Data Science department has also made it easier to reach common goals and simultaneously avoid inhibiting old processes and internal politics because everyone is working against the same goal, "which is to succeed with AI and get it into production" (i02).

Professional resources from different units of the case organization work together for the same goal – data scientists, data engineers, data analysts, full-stack developers, IT architects, and business developers are constructed into interdisciplinary, cross-functional teams. By working interdisciplinary, "the teams can understand each other's tasks, needs, and progress towards the end-goal" (i05). Data engineers understand the origin of data, how the data can be used, what the data can be used for, what processes the data is formed in, and the data quality. They make sure that the data is transferred to the cloud and can be accessed by the relevant people. Data analysts structure the data and understand the physical business and how this information can be used for business purposes. Data scientists and statistical experts develop customized AI solutions and manage the developed ML models. Love and Roper (2009) stated that cross-functional teams play an essential part in the innovation process, enabling knowledge sharing, developing trust, and overcoming spatial and organizational barriers. Some scholars argue that organizational structures that encourage experience-based learning, knowledge sharing, and interaction – such as project teams, problem-solving groups, and task rotation can positively contribute to the performance of innovative activities (Jensen et al. 2007; Gloet and Terziovski 2004). Similarly, an informant states, "it is important to vary employees' tasks and provide new challenges to preserve employee's motivation and avoid solely maintenance of the produced solutions" (i01). Continuous rotation and replacement of tasks are shown to augment knowledge and skillsets and enable more people to perform various tasks. Task rotation requires a certain increase- or transfer of competence, but at the same time, "it enables collaboration, and workers can quality assure the work of each other" (i02). As a result, interdisciplinary teams enhance the flexibility of responsibilities.

Table 4 – Organizational structure-related success factors

Success factors	Indications found in the case organization	Interview source	Supporting literature
Separate the data scientists and a data science department from other departments	Separating the data scientist from the other departments and building a separate data science department has enabled autonomy, trust, room for experimentation, increased focus on the AI projects, low degree of formalization, decentralized decision-making, and adhocracy	(i01-07)	
Build cross-functional teams	Cross-functional teams consisting of full-stack developers, data analysts, data engineers, IT architects, business developers, and data scientists to aid understanding of each other's tasks, needs, and progress towards the end-goal	(i01-i09)	Love and Roper (2009)

Culture and internal communication

Organizational culture has been attracting more attention in the last few decades due to its potential role in improving an organization's future (e. g., Hutchison et al. 2019). AI will force companies to change their business vision, leading to a change in the workplace and organizational culture (Lopes et al. 2019). The rapid pace of technological progress in IT projects makes it difficult for expertise in a particular technique to become mature and established. Creating an environment where unproven tools and solutions are acceptable are thus necessary. The transition to developing AI through cloud computing has been a maturing journey for many in the case organization. For instance, "people working in IT may find it daunting to put Azure machine learning as a lid on top the existing IT infrastructure" (i01), and one of the biggest obstacles to start developing AI through the cloud was "skepticism internally" (i07). To foster adoption, AI services need to have high observability by the organizations' employees and/or their customers. An informant argues that cloud computing enables the organization to mature faster compared to development on a traditional technology stack. In particular, cloud computing allows the teams to start simple and gradually build the complexity, aligned with the increase of experience and knowledge. Previously, managers did not necessarily need to know what the data scientists were doing. Now, data scientists have stepped into the forefront, and data science has thus become a "team sport" that requires collaboration and a shared understanding of the concepts related to the field. The importance of data science work is now being embedded into the fabric of the company. "If data scientists are isolated when developing algorithms, no one will understand what they have done and what problems the algorithms can solve" (i04). Gustafson et al. (2003) suggest that an innovation that fits with the organization's existing values, norms, strategies, goals, skill mix, supporting technologies, and ways of working is more likely to be adopted. It is clearly emphasized in the case organization's organizational strategy that the company aims to be data-driven, and this strategy is rooted in the entire organization. An informant states that the journey towards becoming a data-driven organization is not just a question of information systems and organizational processes but the means of business anchoring, competence sharing, and teamwork. Furthermore, a corporate culture that supports knowledge sharing can lead to more effective achievement because instilling a culture of standardizing and maintaining information is critical to achievement (Lai and Lee 2007). For instance, "encouraging effective communication across teams and bringing people together with different perspectives and work tasks can improve problem-solving and lead to smarter, more efficient decision making» (i09). Mazzei (2010) found that intangible resources, such as knowledge and positive employee attitudes, contribute to a company's success. Similarly, an informant suggests that to boost employee morale, managers should share good news and individual or department "wins" (i09).

Effective inter-organizational communication across structural (e.g., departmental) boundaries within the organization enhances the success of implementation and the chances of routinization (Greenhalgh et al. 2004). Effective internal communication can reduce uncertainty and serve as a catalyst for change. Today, business development is driven by knowledge. It is crucial to enhance all employees' understanding of AI and cloud technology concepts and create awareness of what characterizes good machine learning use cases. For instance, in the early stages of the adoption, several employees functioned as interpreters. The purpose of the role was to describe and covey the AI initiative to people that was not a formalized part of the program. The interpreters aimed to bridge IT and business and anchor effective communication between the two. "What data scientists do is not well understood by most people. If the communication is not sufficiently anchored, the lack of understanding can inhibit the synergy effects between business and IT" (i07). Furthermore, data scientists of the case company try not to use domain jargon or heavily loaded technical words and expressions in communication with others. "The ML model is only as good as others understand it" (i05). Hence, although an ML model has 99% accuracy, it is only as good as people perceive it. More importantly, managers must see the value in the solutions to be willing to finance the projects. It is also essential to connect analytical people who understand the industry in which the business operates with versatile technical professionals. AI, ML, and the cloud are abstract terms and often linked to buzzwords. It takes effort to build a bridge between technology and business and communicate the capabilities of these technologies combined. For instance, mutual knowledge between functional groups provides a potential bridge to organizational productivity (Krauss and Fussel 1990). The case company's data scientists arrange workshops to educate employees at a conceptual level of ML and AI. "If everyone in the organization has knowledge on a conceptual level, everyone knows how the technology can be used" (i04). Further, ideas must be driven by business needs. Therefore, business developers must specify needs while IT solves the needs. "When business people are involved in the development process, there is a greater chance that the project will succeed and create value because the solutions are useful in both a business and customer perspective" (i06). As the business environment becomes more turbulent and time-dependent, organizational productivity often depends on an in-depth knowledge of technologies, processes, and people – both in and across diverse functional areas (Nelson and Cooprider 1996). An informant states that one of their goals is to have an industry-leading environment within BI and advanced analytics, with a sufficient combination of business domain- and technology understanding. A majority of the informants emphasize the necessity of having people with sufficient industry domain knowledge and insight into the business processes and market potential to understand what AI services will satisfy customers. Ideally, AI services specialists should have domain expertise in their organization's industry (Pandl et al. 2021). Conventional wisdom is that managerial communication is important. However, the sharing of knowledge is a different process than managerial communication. Shared knowledge goes beyond the basic informational level. A first step in going beyond the informative briefing stage is to build a common language. The organizational structure has considerably impacted the corporate culture and internal communication. The Data Science department belongs to a larger department called Digital Innovation, which reinforces a culture that supports the whole organization to try new things. As a department and group, the case company's mission is to challenge the entire organization and test innovative ideas. It has allowed them to pursue new projects and dare to challenge established truths. "To succeed, one must dare to succeed. It is better to ask for forgiveness than permission" (i02). As a process in any organization, innovation requires a cultural climate and innovative behavior that enhances creativity and effort in new initiatives (Büschgens et al. 2013).

The cloud enables experimentation due to the on-demand access to computing resources, allowing developers to explore and discover conditions in the company's datasets and explain how AI can solve various business problems for decision-makers. According to i03, exploration enables to create interest and understanding of AI. The cloud allows for subscription- and test technologies for the time required without buying software and installing it. Trialability are educational for business decision-makers and have had a significant impact on the shared understanding of ML and AI. While experimentation encourages innovation, it can also be time and resource-draining (Cespedes and Hoyne 2022). Mitigating this challenge, in 2018, the company developed an innovation methodology that is incorporated into the organizational strategy. The methodology, hereafter referred to as "Felix", is a company-wide method based on known theories and working methods and is a framework for ensuring rapid development of services that customers and employees desire. Rather than deciding on a specific way to innovate, Felix puts together a sequence of three phases – exploration, creation, and implementation, that help teams build insights. The Felix methodology sets guidelines for innovation in the company and serves as a foundation for how people in the organization work with innovative solutions. This working method significantly impacts the culture, collaboration, and problem-solving approaches. An informant states that the Felix methodology is not directly linked to success. However, the teams use the methods' principles and ideas, which have provided a common language for how to think, explore, and make decisions. Companies can gain actionable insights by conducting simple business experiments. Experimentation enables to augment the insights from the data and plan business processes on a test-and-learn basis. An informant suggests that the process of experimentation can provide massive opportunities for businesses. In particular, "they will find themselves in a better position to either quit processes that do not drive value or continue with a given strategy if it proves to be successful (i09)". Flexibility and experimentation afford employees autonomy, freedom, and balance, which translates into higher productivity, loyalty, and engaged employees. The teams use the Felix methodology as a foundation and notion for project execution, which "fits very well with the agile project methodology and how the Data Science department works" (i04).

Table 5 – Culture and inter-organizational communication-related success factors

Success factors	Indications found in the case organization	Interview source	Supporting literature
Facilitate for experimentation	Experimentation allow developers to discover conditions in the company's datasets and explain how AI can solve various business problems for decision-makers. Experimentation afford employees autonomy, freedom, and balance and enables to create interest and understanding of AI.	(i01-i09)	
Implement a company-wide innovation methodology	An innovation method sets guidelines for innovation because principles and ideas provide a common language for how to think, explore, and make decisions. The innovation methodology should challenge the entire organization to test innovative ideas minimizes fear of failure and allow them to pursue new projects and challenge established truths	(i01; i02; ;05; i07; i09)	
Anchor effective communication between IT and business	Investing in interpreters, avoiding heavily loaded technical descriptions, communicating the capabilities of ML, and educating employees at a conceptual level of ML enhance employees' understanding of AI and cloud technology concepts and create awareness of what characterizes good machine learning use cases	(i01; i02; i04; i07; i08; i09)	Nelson and Cooprider (1996)

Agile experimentation

Project methodologies are an integral part of the core make-up of a project. Methodology plays a vital role in supporting the project team throughout the project life cycle to achieve the project's goals. Research indicates a connection between project success and project environment (Joslin and Müller 2016). Findings from the empirical data emphasize an agile methodology for efficient AI development that provides organizational value. As discussed, data science is a team sport. It is crucial to ensure a collaborative environment with many different competencies on the teams. For instance, the company's Cloud Platform department is not a formalized department in the AI imitative. However, they operate and are responsible for the cloud platform and are also part of the teams, making intra- and inter-team collaboration more efficient. The cloud fosters a collaborative environment, culture, and practice because data, information, and applications are available on-demand and accessible from almost any device. Both agile methodologies and cloud computing facilitates flexibility in AI projects. In particular, agile project management has been a vital part of developing AI through the cloud because it promotes an experimental and iterative approach that focuses on continuous releases while incorporating customer feedback with every iteration. An agile perspective is centered around learning through experimentation and introspection, constantly reframing the problem and its solution (Hahn and te Brömmelstroet 2021). "Experimentation is a cornerstone of agile software delivery, and it is also part of the iterative processes and the team member's tasks" (i06). Experimentation in machine learning is essential because data scientists spend much time performing steps, such as algorithm searches, model architecture searches, and hyperparameter searches, to name a few. It can be time-consuming, but it's necessary to arrive at the best-performing model for the problem at hand. An essential element is that the teams and business developers must agree on the quality requirements of an algorithm before embarking on an AI project. When the developers have built an algorithm that satisfies the quality parameters agreed upon before the start of the project, they can inform business decision-makers that the ML model is ready to be deployed in production. Because the acceptance criteria have already been agreed upon, it decreases the time to production and ensures the expected value.

Cloud computing enables experimentation because it removes the need to access capital to experiment while also allowing the workers to focus on what matters the most – "the cloud gets rid of much of the heavy lifting associated with enterprise IT» (i09). In addition, a cloud development platform supports fast development cycles and eliminates major distribution requests, which can downgrade agile development. The iterative approach to agile methodologies is aligned to a changing business environment, changing customer- and market demands, and changing data. Furthermore, demands and solutions evolve through the collaborative effort of self-organizing and cross-functional teams and their customers. The interdisciplinary and cross-functional teams are responsible for the operation, management, and end-to-end life cycle of the AI solutions. Using an agile method when developing AI through cloud computing helps convince non-technical stakeholders supporting black box visibility. This generally results in better communication and understanding of the overall project. Instead of buying months to build an AI product or service, working on the cloud can quickly release a minimal viable product and demonstrate business value immediately. Treating the model as a product and practicing agile methods in AI projects allow the data science team to focus on outcomes. As a "product," it can be updated with new features and improvements. Features and enhancements can also be documented and recorded using cloud computing tools for the best transparency.

Table 6 – Agile experimentation-related success factors

Success factors	Indications found in the case organization	Interview source	Supporting literature
Apply an agile methodology	Agile project management facilitates an iterative approach that focuses on continuous releases while incorporating customer feedback with every iteration	(i01-i09)	
Include experimentation in the iterative processes	Experimentation in machine learning is essential because data scientists spend much time performing steps, such as algorithm searches, model architecture searches, and hyperparameter searches	(i01-i07)	Hahn and te Brömmelstroet (2021)
Agree on quality parameters of algorithms before embarking on an AI project	When acceptance criteria are agreed upon before the project start, the ML model can be deployed in production faster	(i01; i02; i04; i05; i06)	

5.2.2 Cloud management factors

Cloud management refers to managing data across cloud platforms, either with or instead of on-premises storage, and managing the cloud infrastructure resources and services. A well-designed cloud management strategy can help IT professionals control the dynamic and scalable computing environments. Cloud management ensures the implementation of cloud data policies and procedures that give organizations control of their data, both in the cloud and in set-ups where data is stored or sourced in a combination of on-premises and cloud applications. Sufficient cloud management enables consistent deployment, operations, and optimization of applications, resources, and tools in the cloud infrastructure, which substantially impact the AI technology commercialization process.

Data management

In cloud computing, IT infrastructure comprises virtual resources that support the flow, storage, processing, and data analysis. Data management enacts data governance policies and procedures and is an integral part of processing, storing, and organizing data for cloud AI projects. According to the DAMA Internationals Data Management Body of Knowledge 2.0 (DMBoK2), data management is: "the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles." The maturity of an organization's data management determines how a company can strategically implement new business ideas or models. Some informants mention the importance of adhering to a data strategy. When implementing a data strategy, "there is so much more opportunity than just doing what you do better" (i09). Entirely new business models can be created through well-managed data and a solid data strategy. A data strategy aims to ensure that data is managed and used as an asset. However, "it is impossible to adhere to a strategy unless all employees are aligned, and there are implemented processes for managing the data" (i07). The data strategy is also about capturing the correct data, extracting the value from the data, and using it to develop new services. A data strategy enables superior data management and analytics capabilities that support managerial decision-making and ultimately enhance financial performance (DalleMule and Davenport 2017).

Unstructured data, data quality, and GDPR are challenges when developing AI solutions for new services because they require resources to access and organize the new data. "The quality of an algorithm is never better than the data quality" (i05). The probability of unstructured data and poor data quality increases when several information systems and business processes generate data. Additionally, there are human factors that come into play. For instance, when carriers do not scan the packages at the right time, the data quality decreases, subsequently affecting the algorithms' quality. "A common misconception I notice in many companies is that they believe artificial intelligence can do more for them than what is actually achievable. This is often due to the data quality" (i09). The quality of data cleansing directly impacts the accuracy of the derived models and conclusions.

A conventional information system process data, while machine learning uses the data. According to the GDPR, collected data cannot be used for any purpose other than that for which the data was initially collected without approval. Therefore, a challenge is approval of using the data for new services. "It is difficult to justify using the collected data for a purpose other than for which it was originally collected" (i08). Incorporating a cloud strategy is necessary to reap the benefits of the cloud and ensure data security measures (i02; i03; i07). Additionally, an overall cloud strategy "closely linked to business requirements is necessary" (i09). Compliance and accountability of GDPR lay the foundation of any data process. GDPR has a considerable effect on cloud data privacy and compliance for both the cloud service vendors and the data consumers (e. g., company/users). GDPR mandates that organizations collecting and processing information related to EU citizens adhere to its articles irrespective of where they are located or where the data is stored (Elluri and Joshi 2018). This regulation is currently available only in the textual format and requires significant manual effort to ensure compliance. In the target case, GDPR is highly preserved and integrated into the company's data- and process rules – "the stored data can be used for business purposes " (i05). On the other hand, cloud computing introduces challenges related to the geographical location of data storage. On an operational level, stability, connection, latency, and responsiveness of data sources are a concern. As implied in the literature, AI services may not work well with systems that need to infer decisions in real-time due to potentially high latency. Additionally, "resources are more responsive if data is located in Norway than outside Europe" (i08). There are also critical legal issues around where data is located. The ability for cloud services to quickly return to an operational state after a failure or data loss depends on standby locations. In case of server failure, data can be automatically transferred to a data center outside Norway, depending on the cloud computing contract and agreement with the cloud vendor. However, there are requirements for Norwegian organizations in where data is stored, and companies have obligations when transferring personal data out of the EU. In July 2020, the European Court of Justice handed down Schrems II's verdicts, R-01/2020 and R-02/2020. The verdict provides regulations for transferring personal data to countries outside the EU. The European Court of Justice emphasized that organizations in EU member states must always examine whether conditions in the third country will undermine the level of protection required. The verdict thus affects the use of cloud services and how the underlying IT infrastructure is organized. A significant challenge for the companies that use cloud services to develop AI is ensuring that the agreement with the cloud provider follows Norwegian legislation and GDPR and complies with data privacy laws and corporate policies. The server locations should therefore be based on these considerations. A company might need to have all data under the same jurisdiction (e, g., in Europe). In other cases, Safe Harbor principles, in which US companies comply with EU laws, can be good enough (Serrano et al. 2015). Furthermore, many external servers are located outside Norway's borders. The entire resource directory provided by cloud vendors is not always available in the organization's respective region (e.g., Norway). Hence, when choosing tools and resources, it is crucial to ensure that the resource hosts the data in Norway "so that actors outside the EU do not demand access to the organization's data, which may impede the privacy of individuals" (i08). For instance, recital (101) of the GDPR suggests that,

"…. when personal data are transferred from the Union to controllers, processors or other recipients in third countries or to international organizations, the level of protection of natural persons ensured in the Union by this Regulation should not be undermined..."

However, GDPR privacy provisions are not binding in third countries, and third-country recipients- or international organizations' laws may have precedence over GDPR privacy provisions. Therefore, the transfer of personal data to countries outside the EU, so-called third countries, requires a legal basis for processing personal data, a so-called "transfer basis." Until the summer of 2020, the most commonly used transfer basis has been the Privacy Shield and the EU's standard agreements ("SCCs"). The European Court of Justice concluded that the Privacy Shield was an invalid transfer basis. The reason was that the USA's monitoring practices and regulations conflicted with several of the rules in the GDPR and the EU Charter of Fundamental Rights. Consequently, all transfers based on Privacy Shield have been illegal since Privacy Shield was adopted in 2016. The Schrems II verdict has thus set a new standard: The right to transfer personal data to third countries will fall on in-depth and complex risk assessments. All companies must have an overview of what personal information is collected, used, and shared by virtue of being responsible for data processing. Hence, if a company has a cloud-based solution that either transfers or hosts data in third countries, they are affected by the Schrems II decision. The responsibility for complying with the rules for transferring personal data to third countries (and the GDPR in general) lies with the data controller.

Table 7 – Data management-related factors

Success factors	Indications found in the case organization	Interview source	Supporting literature
Implement a data strategy and processes for how to manage the data	A data strategy aims to ensure that data is managed and used as an asset and aims to control, protect, and enhance the value of data and information assets. A data strategy enables to capture the correct data, extracting the value from the data, and using it to develop new services	(i07; i09)	DalleMule and Davenport (2017)
Incorporate a cloud strategy	A cloud strategy is necessary to realize the benefits of the cloud and to ensure data security measures	(i02; i03; i07; i09)	
Ensure that the agreement with the cloud provider follows Norwegian data legislation and GDPR	It is crucial to ensure that a cloud resource hosts the data in Norway and, in case of a server failure, that data is transferred to standby locations in Europe	(i08)	Serrano et al. (2015)

Infrastructure management

Cloud infrastructure management refers to the set-up, configuration, monitoring, and optimization of the cloud environment and components and resources in the cloud. Cloud infrastructure management thus enables organizations to consolidate and scale IT resources. Literature and industry often argue that cloud computing aims to democratize AI by abstracting some of the complexity associated with developing AI solutions through cloud computing. "When going from proof of concept to scaling ML models, having a business-specific AI environment and infrastructure on a cloud platform is critical "(i08). Most AI service providers offer a graphical user interface (e.g., dashboards) to simplify operation. Offered GUIs can be used to select, tune, and deploy an appropriate machine learning algorithm or monitor KPIs and visualize analytics performed over the data. However, findings from both the literature review and empirical research signify that AI delivered through infrastructure- and platform services require overcoming significant complexity associated with the configuration of computing resources and the operation and management of components in the cloud environment. Additionally, integrating it into existing information systems can be a demanding process. Infrastructure as a service can, in many ways, be compared to the traditional way of developing AI. The administrator is responsible for setting up the VM, operating system, disks, and network infrastructure. An advantage of such responsibility is that "you have to work more closely with the technology and understand the set-up, enabling you to see opportunities more easily and experiment to get where you want" (i05). Cloud platforms are customizable with an extendable architecture. Users can integrate third-party services, connect with various cloud-based AI infrastructures, and configure these infrastructures to meet their needs. An informant advocates the necessity to build the cloud infrastructure explicitly towards the company's datasets, data structures, and ETL processes. "It is essential to pre-determine ETL processes, datasets and data structures before building the infrastructure and developing the platform" (i07).

AI and ML are developing rapidly. There are frequently new updates, code, and data that businesses seek to acquire. A traditional technology stack requires management and maintenance of hardware and servers, and it is costly and time-consuming. On the other hand, the cloud provides access to the latest developments and the freedom to choose resources, enabling organizations to stay competitive. Access to state-of-the-art tools and updated resources have been a major driver for the company to adopt AI through the cloud. However, updates may, for example, affect how the platform reads files from the data warehouse, which requires competence to configure correctly. For instance, "if a security update prevents someone from accessing the APIs, someone must enable the access manually" (i02). A significant pitfall when adopting AI services is the "belief that you are buying an AI machine in the cloud" (i02). The reality is that businesses need people with the right competence to set up, operate, and maintain the cloud environment. As a result, there is much work associated with AI in the cloud, and it requires specialized skill sets and professional resources. The compatibility characteristic of the DOI theory addresses the fit with existing infrastructure and knowledge of the workers or potential adopters. Interviewees indicate that there may be a lag between cutting-edge research on AI services and AI services offerings in practice. Additionally, it may be a lag between the release of new, innovative AI services offerings and the skills and awareness of users. Adoption intentions might change once the maturity level and knowledge of AI services increase.

The time-consuming processes of setting up and maintaining a distributed computing cluster and choosing a suitable cluster configuration (Doug and Sang 2018) suggest the need for IT experts. An informant advocates 5-6 data engineers per data scientist (i05), and another informant recommends 8-10 data engineers per data scientist (i08) to set up, configure, and manage the cloud infrastructure. Furthermore, full-stack developers are needed to build APIs to publish solutions and develop user interfaces that allow access to the data. For instance, the finished products of predictive ML models that estimate the delivery time of parcels are published as API web services in e-commerce checkouts for online shoppers. The APIs provide online shoppers with a time window for when the package is expected will arrive at their local pick-up point.

The cloud also provides access to SaaS tools, such as Databricks. Databricks is a cloud-based data engineering tool used to process and transform massive quantities of data and explore the data through machine learning models. However, Databricks does not work isolated. It is dependent on having a place to retrieve data and a place to store data. Therefore, it is necessary to continuously ensure that SaaS tools and PaaS products are connected and work together. A challenge highlighted is the pitfall of underestimating security requirements and how systems and components are integrated and managed. Therefore, IT architects are necessary to ensure that processes and systems are connected. A considerable advantage mentioned by several informants is the collaboration with the cloud vendor, Microsoft. Through developing the AI environment, Microsoft has assisted and supported the company through the whole process. "It has been vital to work closely with Microsoft because they know the products. Microsoft has supported installing and configuring resources in the cloud environment" (i02). Similarly, numerous research has shown that the degree of vendor support affects cloud computing service adoption on the organizational level (e. g., Oliveira et al. 2014; Shu 2016). Cloud configuration is the process of setting hardware and software details for elements of a cloud environment to ensure interoperability. A cloud provider must enable service for clients using various hardware and software and ensure that the service is reliable, performance is acceptable, and communications are secure.

The combination of cloud computing and AI leads to unique advantages over in-house AI solutions. AI infrastructure services are flexible, and machine learning requires flexibility. For instance, "you may need 100 GPUs for a project, for a week only" (i08). On-demand access to specialized hardware is unique to AI in the cloud. For instance, Azure provides valuable flexibility for data scientists to set up a machine learning logic instead of a much more time-consuming set-up on on-premises infrastructure. Cloud solutions are typically built on top of existing IT infrastructure, removing the need to renew the infrastructure completely. What's more, building the Azure environment on top of the existing infrastructure removed the time-consuming ordering processes from external suppliers while facilitating experimentation and rapid value creation. For instance, "it is important to get started quickly and deliver value to make it clear that the solutions can be used" (i01). Additionally, building AI and cloud initiatives alongside the existing infrastructure removes the need to reuse legacy technology stacks (i03). Data is the primary source of any AI system. "As a data consumer, the data must be accessible and good quality" (i04). Another informant advocates three essentials for a successful ML model: 1) access to the required data, 2) access to computing power and resources that enable retrieving and manipulating the data, and 3) scalability (i05). Access to the required data necessitates that the data is structured and well documented, and everyone must know what data exists and how to access it. Hence, ensuring that the cloud environment is not entirely separated from the existing infrastructure is crucial to prevent data silos and maintenance costs. For instance, "separating the cloud environment from the existing infrastructure risks creating a large maintenance machinery" (i01). However, integrating the existing infrastructure and the cloud environment to enable streamlined data access is reported as some of the biggest challenges by the interviewees. Legacy systems may not support the connection to or integration of cloud technologies (Pandl et al. 2021). For instance, an informant states that building a bridge between existing technologies and the cloud is challenging. "It requires communication channels that retrieve data from other central computer engines, which requires a lot of adjustments and configurations to ensure substantial security" (i05). In particular, new technology should replace legacy technologies to increase speed or reduce maintenance costs. There are often many technologies with many hooks to old systems in a large and old company – the technologies depend on other old systems to work. "Migrating systems and technologies to the cloud and setting up the same hooks to relevant sources is time-consuming, considering the company's size" (i05).

Table 8 – Infrastructure management-related success factors

Success factors	Indications found in the case organization	Interview source	Supporting literature
Have 5-10 data engineers per data scientist	Data engineers have the competence to set up, configure, manage, and maintain the cloud infrastructure	(i05; i08)	Doug and Sang (2018)
Collaborate with the cloud vendor	The cloud vendor knows the products and can support the installing and configuring of resources in the cloud environment	(i01; i02; i03; i05)	Oliveira et al (2014); Shu (2016)
Integrate on-premises infrastructure and the cloud environment	Developing the cloud infrastructure on top of existing infrastructure increase time to value and reduce the need to reuse legacy technology. However, integrating the existing infrastructure with the cloud environment prevents data silos and maintenance costs and allow AI solutions to scale	(i01; i04; i05; i07)	

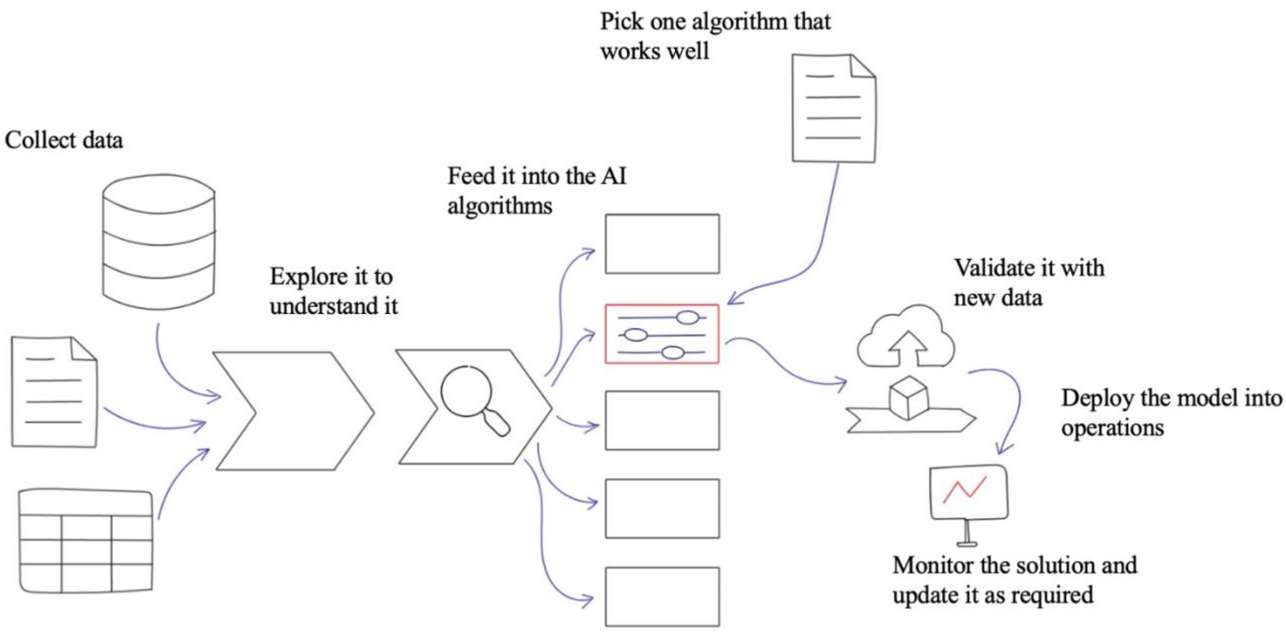
5.2.3 Technological factors

Improvement in productivity and efficiency depends not only on technology but also on how it is integrated into the organization. Interviewees frequently highlight the importance of AI services' ability to automate, scale, and customize AI solutions. AI automation is vital in the technology commercialization process because it increases speed and efficiency, saves time, and allows scalability. AI solutions can lead to automation of business processes, and business processes can lead to automation in AI development. Automation in AI development has proven to deliver benefits such as improved performance drivers, faster processing time, and increased flexibility. ML models must be integrated into operational technology systems, organizational processes, and staff workflows to ensure automation of business processes. According to Gartner, scalability is a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands. Scalability thus refers to both a scalable cloud platform and scalable AI solutions. Scalable algorithms and infrastructure enable to apply the power of AI to critical business processes. A scalable cloud platform enables organizations to scale the AI solutions to solve emerging business problems and respond to increases and decreases in demand. Finally, customizability affects the technology commercialization process because control over the data and ML model enables users to build higher quality, business-specific models. Customizability refers to both custom cloud platform development and custom AI solutions.

Automation

Contrary to prior research suggesting that automation removes the need for AI engineers (e. g., Lins et al. 2021), the empirical evidence of this study signifies that AI developer- and infrastructure services require complex management processes for automating the development of AI solutions on the cloud. The interviewees highlight two drivers for automated AI development, 1) pipeline management and 2) automation of algorithms. Pipeline management focuses on machine learning tasks, such as data preparation, training configuration, training and validation, and repeatable deployments. Pipelines must be reusable and streamlined through the entire process to production to ensure automation. Besides, ML pipelines ensure the solution's value after it is deployed in production without requiring too much maintenance. Data scientists and ML engineers need information on concepts and tools to accelerate, reuse, manage, and deploy AI developments. Hence, standardization of machine learning pipelines is vital. Fortunately, the Azure platform makes pipeline management easier (i06). An Azure Machine Learning pipeline is an independently executable workflow of a complete machine learning task. Subtasks are encapsulated as a series of steps within the pipeline (Esposito 2020). Pipelines require a significant investment and they must be monitored regularly. A pipeline can either be monitored manually or with alarms alerting developers when model accuracy decreases. Furthermore, it is essential to determine which metrics to measure the model on when deployed in production. Frequent testing of the ML model will ensure its validity. However, “it is challenging to avoid manual testing with automation” (i05). AI development requires the high-level steps visualized in figure 16.

Figure 16 – AI development steps



A typical model building lifecycle involves the process of collecting, preparing, analyzing, and infusing insights into data iteratively until the wanted process efficiency is achieved. Automating these model-building tasks help developers simplify their AI lifecycle management. An ML model will change and often perform poorly when exposed to new data or user patterns. Therefore, it is crucial to automate processes that reveal information on how the model behaves to identify any inconsistencies and then re-train the model to increase accuracy. To do so, the case organization´s data scientists write code and logic so that an ML model can run itself. As mentioned, the quality parameters and accuracy requirements are defined in advance. Then, the algorithms check themselves to determine whether the quality parameters are sufficient. If the accuracy is below the requirement, the model is re-trained and deployed again by itself. These processes are not dependent on many human resources due to the automation of algorithms (i05). Furthermore, to update and include new data in the solutions, the data flow is described and monitored, and when something does not work, the developers are automatically notified.

Scalability

Cloud providers can elastically provision and release hardware resources available to the platform and thus scale horizontally per the user-defined configurations and requirements. However, scaling AI initiatives from proof-of-concept to the enterprise level present many challenges, such as data management, trial and error, and sufficient interdisciplinary expertise. Interviewees suggest that enterprise scalability of AI development and deployment includes 1) establishing Machine Learning Operations (MLOps), which enables to streamline the ML lifecycle and create a framework that facilitates communication and collaboration between data scientists, DevOps engineers, and IT, and 2) establishing extensible system architectures to add components, data, and functionality to the solutions without impacting the internal structure or data flow. MLOps aims to simplify the management process and automate the deployment of ML models in large-scale production environments. An informant states that, “not adapting to MLOps as a practice can result in insufficient scale that is vital to meet the needs of the business” (i08). Hence, MLOps are proved essential for large-scale projects.

Both the cloud platform and the elements in the cloud environment need to interoperate and communicate to scale AI solutions. "A critical success factor for deploying AI solutions into production is having a scalable platform in the background from day one" (i02). The benefits of using AI services must be easily quantifiable and observable. Workers in an organization typically demand proof of concept demonstrating the value of AI services before they are convinced of the technology's benefits. An informant states that, at the beginning of the cloud AI venture, it was necessary to develop the Azure environment on the side of the existing infrastructure to decrease time to production, which aimed to "increase internal interest in the technology and point to some success stories" (i01). What's more, "companies that adopt AI incrementally generate value faster compared to companies that attempt organization-wide AI adoption» (i09). However, as discussed, to scale the platform and AI solutions over time, the cloud and the on-premises infrastructure must be connected to avoid data silos and additional maintenance costs. It is also essential to maintain continuity in the AI deployment approach to prevent the creation of siloed use-case-based AI solutions that can't perform holistically. Correspondingly, instead of spending a long time building the platform, an informant advocates the importance of starting with a small business problem that can be solved with the available data. Starting small and gradually building complexity enables gradually building a platform and an AI environment that can be scaled to the next problem, which "will prevent silos for a specific isolated problem and is necessary, in an enterprise architecture perspective" (i02). Another informant also advocates starting with a small project to get started – "don't try to boil the ocean, but choose a pain point to solve – where you can show demonstrable progress" (i09). As a result, to ensure scalable cloud-based AI projects, it is necessary to find a small problem that can be solved with the available data and scale the solution aligned with the platform development, use case development, and the increase of experience. "There is a big difference between academically cutting-edge AI and what is useful" (i04). It is thus essential to focus on simplicity rather than starting with all the fancy tools the cloud has to offer.

Customizability

Automating and scaling AI solutions requires companies to build complex, use-case-driven systems that address unique business requirements and needs. Customizability refers to developing business-specific AI solutions from scratch by taking advantage of cloud technology's tools, frameworks, and computing power. Developing bespoke AI products requires cloud-based services capable of handling AI workloads. Every organization uses specific data to achieve specific goals according to a specific business problem. For an AI solution to provide a perfect fit for the data, usability, and business needs, these specifications need to be translated and built into the ML model. Developing custom AI tools involves building custom algorithms and proprietary APIs.

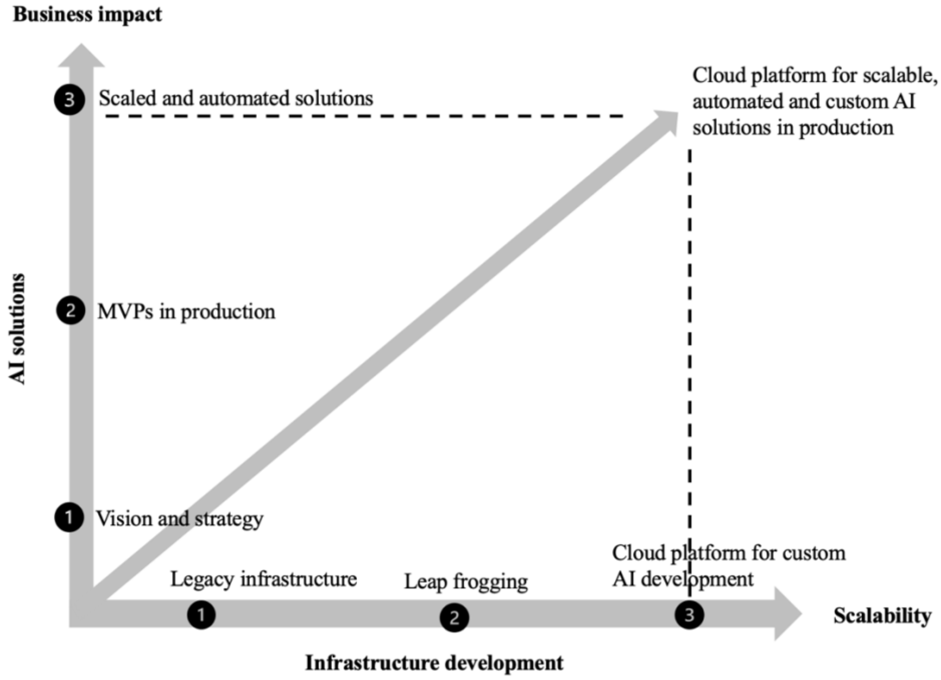
Most cloud providers describe and provide educational recourses to users about their offerings while also offering a trial period. In cases where the information system uses commoditized prediction APIs that are easy to integrate, adopters can choose one provider out of many possible providers and try out their solution while still having the flexibility to switch to another provider later on. On the contrary, for users of more complex offerings such as machine learning as a service, trying the services can impose challenges. The user may need to learn about the specific MLaaS offering, integrate it with a complex process into their information system, and upload large amounts of data into the cloud storage. In such cases, AI services may result in provider lock-in effects. An approach to custom platform development is to build nimble architectures using best-of-breed components. An informant advocates that building a platform using different technologies for different components will prevent vendor lock-in (i09). Furthermore, technology typically develops faster than businesses are able to follow up. "Building a platform of different technology components prevents betting it all on a single vendor or technology" (i08). It thus becomes easier to swap technologies and benefit from multiple innovation sources. An informant states that "customizability is not an issue; the cloud environment is highly configurable" (i03). The cloud offers an ecosystem that provides seamless integration. Combining cloud-native and open source technologies enables IT engineers to stitch together different platform components. Besides, custom AI solutions may leverage existing software libraries. On the other hand, according to Yao et al. (2017), more user control and customizability over the platform and solutions suggest higher performance and greater risk because higher configurability leads to the risk of producing poorly performing models. Additionally, users must obtain high-quality and adequate data because feeding low-quality input data into an AI model's training process will decrease accuracy and explainability (e. g., Thiebes et al. 2021). Therefore, custom AI development requires significant expertise, a steep learning curve, and domain-specific know-how to sustain a well-managed cloud environment that facilitates custom AI development. However, custom AI solutions may be better when considering integrations with existing software. The cloud offers tools and technologies that connect various applications, repositories, and IT environments to exchange data and processes.

Table 9 – Technical success factors

Success factors	**Indications found in the case organization**	**Interview source**	**Supporting literature**
Standardize pipelines and automate algorithms	Standardized pipelines expose information on concepts and tools to accelerate, reuse, manage, and deploy AI developments. Pipelines automate model-building tasks and simplify the AI lifecycle management. Automated algorithms expose information on how the model behaves without much human intervention	(i05; i06; i07; i08; i09)	
Establish MLOps and extensible system architectures	Establishing MLOps enables to streamline the ML lifecycle and automate the deployment of ML models in large-scale production environments. Extensible system architectures enables to add components, data, and functionality to the solutions without impacting the internal structure or data flow	(i01; i02; i05; i09)	
Build nimble architectures using best-of-breed components and build a cloud platform that uses different technologies for different components	An approach to custom platform development is to build nimble architectures using best-of-breed components. Building a platform using different technologies for different components can prevent vendor lock-in and make it easier to swap technologies to benefit from multiple innovation sources for customized AI solutions	(i02; i03; i08; i09)	

The implications drawn from the issues related to automation, scalability, and customization can be visualized in figure 17.

Figure 17 – Platform development for automated, scalable, and customizable AI solutions



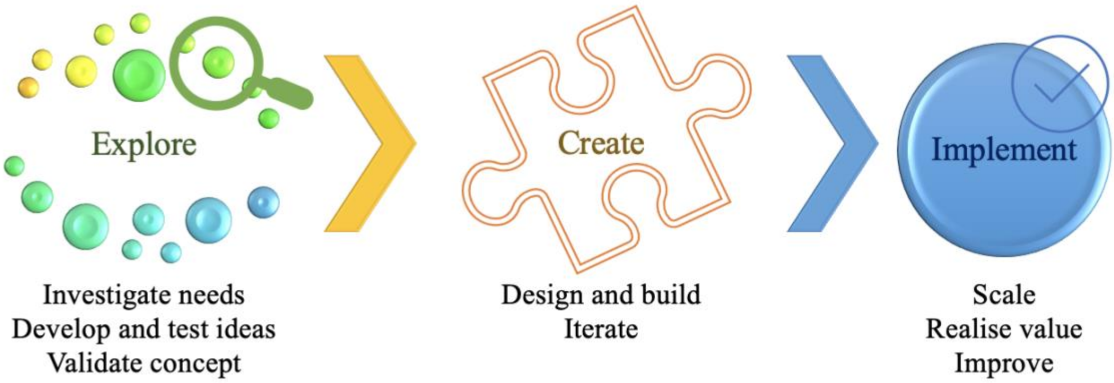
The most promising AI solutions go from ideation to automated and customized solutions at scale. Adopting AI and improving its success rate requires investment in the processes from idea to production of an AI solution. The AI vision and strategy need to suggest concrete, market-specific, and customized AI solution designs. As discussed, the first AI solutions pose requirements to the platform. Then, the platform can be developed by applying agile development principles and leveraging the requirements from use cases to innovate the platform. As legacy infrastructure is characterized by massive, vendor-based systems, developing the cloud environment over the existing infrastructure decreases time to value and enables observable results. As a result, it is a matter of managing the trade-off between fast time-to-market (including building the cloud infrastructure and platform on top of existing infrastructure) and building high platform standards (including portability, migration, and interoperability across multiple platforms to avoid data silos and maintenance costs and underpin security and data protection). The objective is to avoid building point solutions that are hard to scale or that IT teams to build a platform that is not used due to insufficiency. The technology commercialization process elaborated in the following sections adds further context to the model.

5.2.4 Technology commercialization process

Technology commercialization is the process of bringing an idea to market and creating financial and customer value. The Felix methodology serves as a basis for the technology commercialization process. The method introduces a test-and-learn approach for developing digital offerings. The three phases, explore/experiment, create, and implement aims to increase efficiency and reduce time to market. Similarly, the cloud infrastructure accelerates the speed at which an idea is deployed in production due to the ability to operate autonomously and scale resources according to demand.

The Felix method is a company-wide way of working. It aims to start small, break down the problem and see if it can be solved before scaling it to a larger project. The method can be associated with leapfrogging. Leapfrogging refers to small and incremental innovations that lead a dominant firm to stay ahead. An informant states that it's not necessarily the sophistication of the ML model at the beginning. It's about creating a better experience for customers. Today, companies do not compete against their closest competitor; they're competing against the best customer experience in the market, even if that's in an entirely different sector. The Felix method provides a framework for exploring, experimenting, and implementing the right solutions – "the method has made it easier to identify a minimum viable product, which has been very important" (i07). The informant further implies that the method reduces failure's negative consequences and reduces the risk of making wrong decisions. There are iterative processes within the three phases of the method – explore, create, and implement. These phases aim to understand customer needs, create a prototype, test it, get feedback, and deploy it in production. However, it is important to note that the method only applies to customer projects and not to internal AI projects.

Figure 18 – Innovation model[1]



Explore and experiment

The explore phase is concerned with identifying market needs and trends. Ideas are developed and tested, and the market potential for new ideas is validated. AI's nature is about exploration and experimentation. The cloud enables to experiment with many different ML models and tests various possibilities. It is easier to turn around due to the flexibility offered by cloud computing. "Exploring provides business value because you do a piece of work, which is not necessarily put into production, but it can enhance experience and knowledge of the workers without having to invest too much time or resources" (i07). If a solution does not work as expected or appears irrelevant to the market, it is easy to move on to the next project. Investing time, money, and human resources in AI development on an internal technology stack can make projects harder to discontinue and thus pressure the company to deploy the developed solutions in production that may have no real business value. Cloud computing makes it possible to use the service capacity when needed. If there is a certain period of time that the teams need a lot of service capacity to try new things and experiment, they can scale the service capacity as necessary. The pay-per-use model of cloud computing is valuable considering that AI requires trial, error, and experimentation to succeed. To succeed with AI, volume is essential. "I encourage clients to do 100 AI pilots. Half of them won't work, but half that work can really pay off" (i09). Experimenting in the cloud makes it easier to decide whether to pursue the pilot or not. On the other hand, data quality is a challenge when experimenting with- and developing solutions for novel services. Additional resources are required to access and organize the new data needed to create new services. To explore and

[1] The model is adopted from the target case and given a fictitious name and a new design to anonymize the organization studied

experiment with new solutions, the company performs Hackathons where they try new things on a small scale to see if they can provide value and then bring it to the creation phase.

Experimentation in machine learning is essential because data scientists spend much time performing steps, such as algorithm searches, model architecture searches, and hyperparameter searches. Scaling the cloud resources to expand business capabilities aligned with the experimentation of ML models and algorithms searches, model architecture searches, and hyperparameter searches is thus necessary. Extensive experimentation can enable promising results. However, machine learning experiment management is needed to avoid a lack of control and to understand precisely what model performed best. ML experiment management advocates using the same environments to run the models, track feature versions and record the parameters. Thus, experimentation management entails tracking and monitoring metadata like code versions, data versions, hyperparameters, environment, and metrics, organizing them in a meaningful way, and making them available to access and collaborate within the organization. Tracking refers to collecting all the metainformation about the ML experiments that are needed to share the results and insights with the rest of the team, replicate the results of the ML experiments, and safely keep the results that take a long time to generate.

Create

The creation phase aims to test customer value and market potential by building and testing a proposed solution. A stable and functioning architecture is crucial. "The IT architecture must be continuously based on the company's objectives and how customers want to access the information they are looking for" (i02). Therefore, it is essential to constantly involve customers and business developers during the creation of the solutions.

AI requires a vision and strategy that combines quick wins with a route to scale. For instance, "the solution must be aligned with strategic and tactical goals." Decision-makers must ask, "Is this the path the organization wants to pursue?" (i07). Hence, technology alignment management aims to ensure that the company's overall strategy drives the IT strategic plans. Business and IT alignment suggest the combination of processes, structures, plans, and decisions that either bring successes or failures to organizations.

During the creation and development execution of AI solutions in a cloud environment, an informant suggests that data scientists need rich experience in tuning the ML models to find out what works (i05), which is aligned with findings from Lins et al. (2021) and Yao et al. (2017). Compared to traditional software development, managing ML models adds additional complexity, including model architecture, code, hyperparameters, and features. In addition to the required skills to optimize feature, model, and parameter selection (e.g., Wang et al. 2018; Duong and Sang 2018), model management is also a crucial element in the long run. In particular, “tracking the model performance ensures consistency and satisfaction of business requirements and customer needs” (i08). It is also essential to have control over data and ML models. "It is difficult to know how the ML model arrives at predictions if you do not have control over the data and the model" (i08). Sufficient data- and model control enable troubleshooting of any inconsistencies and, more importantly, adhere to GDPR legislation. Further, "developing traditional applications and deploying them in production introduce major pitfalls, but machine learning add even more dimensions to pitfalls" (i08). Such pitfalls are often related to unstructured data, quality and availability, and GDPR. An informant advocates the importance of having thorough processes to be coordinated internally and ensuring that the data reflects the business processes.

Managing technology alignment can start with the organizational structure and be boiled down to communication between technical experts and business developers. An informant advocates "a good dialogue and to simply speak in plain [Norwegian], even when trying to explain complex technical issues" (i09). Enabling alignment between IT and business requires committing to insightful discussions. For instance, "explaining the concepts behind developing AI through the cloud allows business leaders to understand the building blocks for AI, and the IT leader can develop a strategy that satisfies the business leaders' requirements" (i09). Furthermore, instead of directing the discussion towards IT and related technologies, it is found to be more beneficial to ask business leaders about their jobs, including market trends and the critical business challenges the organization is currently facing. Then, the IT experts' job is to evaluate the technology solutions that could serve those needs (i07; i08; i09). Metrics and key performance indicators (KPIs) reflect the business outcomes of critical systems and processes. They are valuable tools to help IT professionals and business leaders determine how well IT operations align with business needs. An informant suggests that "business outcome-driven KPIs include tracking whether the programs delivered provided business benefits" (i09).

With AI, the challenge is often getting an ML model from a closed environment to production. Model results in production are usually weaker than during model training. "It is dangerous to underestimate the complexity of how robust a model must be" (i09). The development of AI on the cloud often goes fast forward from exploration to implementation. However, it is essential to work thoroughly with an idea and test it with a customer. "You have to put a lot of time and effort into a pilot project to see if it provides real value" (i01). The creation phase must be given enough time so that a large cloud maintenance machinery is not built next to the existing infrastructure but that it is connected to ensure interoperability. The problem is often that the setup of a pilot does not fit with the implementation. In a large organization, there is often a complex cloud architecture and structure and how data is related, but in a pilot, there are shortcuts. Developing a minimum viable product (MVP) enables one to understand the customers' interest in the solution without fully developing the product. MVP is a development technique that involves introducing a new product in the market with basic features. The final product is released only after receiving sufficient feedback from the initial users of the solutions. An informant suggests that to cover the MVP, one must understand how a pilot project can be scaled and how to mitigate the challenges related to deploying a solution.

Implement

In the implementation phase, the solution is scaled up and deployed. It is continuously improved until it has reached its full potential. In the implementation phase, one must ensure that opportunity and market match. It is not always the right time, and not all solutions will suit the company, the market, or the customers. In the implementation phase, it is vital to ensure that the projects during the creation phase are feasible and that the solutions work for a wide range of customers.

An ML model often works well in a closed environment with internal testing. However, in a production environment, an AI solution must cope with extreme data scenarios, including noisy, unstructured, or small data sets. AI solutions deployed in production will either prove value or remain flop. «Enjoy the comfort of the data science sandbox, but prepare for the cold, harsh world of production» (i09). Therefore, it is essential to have a framework that supports a smooth transition from testing to production. An AI solution should achieve and maintain robustness and stability. Integrated technologies and AI expertise are vital to achieving robustness and stability. Achieving production and commercialization of AI solutions can be particularly difficult, but it is only the first half of the battle. An AI solution cannot be abandoned after it has been deployed. Therefore, the second half of the battle is maintaining the AI solution in production to ensure that the model does not go off the rails as data changes. Maintaining AI in production requires model version control, updates, new data, human machine-interaction optimization, monitoring of the model's robustness and generalization, and ongoing input noise detection and correlation. Ongoing maintenance can thus be challenging and expensive. Consequently, to ensure that the AI project provides real value, the case company is restrictive with which projects are implemented. "There must be enough resources available, and one must have reasonable faith that it will be a success" (i01). Hence, once a solution has been deployed into production, it is crucial to have the right tools and sufficient time and resources to monitor and adjust the solution over time.

After the machine learning model is trained and deployed, this model's predictions may behave like a black box. ML models that are thought of as black boxes are impossible to interpret. In Europe, GDPR legislation requires organizations to be able to explain algorithmic decisions if they impact the customer (Article 4 (4) and Article 22 and Recitals (71) and (72) of the GDPR). Additionally, research has devoted much attention to trustworthy AI (e. g., Thiebes et al. 2021). The authors argue that organizations will only be able to realize the full potential of AI if trust is established in its development, deployment, and use. Therefore, reverse engineering would make a model more reliable because it enables to find explanations for why a certain prediction was made. Accordingly, machine learning explainability is crucial for an organization to fully understand the AI decision-making processes with model monitoring and accountability of AI. Machine learning explainability can help data scientists and developers better understand and interpret an ML model's behavior, such as determining which features the model considers necessary and the relationship between the feature values and the target predictions. Furthermore, explainability helps to debug and improve the quality of the model and allows users of ML algorithms to understand why the model made a certain prediction.

Cloud computing is associated with the difficulties of assigning legal responsibilities and liabilities in the complex processing environments and the general implications of AI working to determine and drive customer applications' functionality (e. g., Cobbe and Singh 2021).

Fortunately, a cloud platform can provide tools and frameworks with capabilities for explaining, validating, monitoring, and mitigating bias in AI models as part of the end-to-end AI lifecycle. Therefore, developing a cloud platform explicitly designed to instill user trust is essential in a market where customers are the center of business processes and decision-making. On the contrary, an informant state that “AI solutions' lifecycles are challenging to predict because the lifecycle relies not only on the data but also on the dynamics of competition and requirements for sustainability” (i07). Fortunately, it is relatively easy to make rapid changes due to the flexibility of the cloud platform. In addition, the agile working methodology is a prerequisite for rapid change.

Table 10 – Technology commercialization process success factors

Success factors	Indications found in the case organization	Interview source	Supporting literature
Manage AI Experiments	Machine learning experiment management is needed to avoid a lack of control and to understand precisely what model performed best	(i01; i02; i07; i09)	
Manage technology alignment	Technology alignment management aims to ensure that the company's overall business strategy drives the IT strategic plans	(i05; i07; i08; i09)	
Ensure machine learning explainability	Machine learning explainability is crucial for an organization to fully understand the AI decision-making processes with model monitoring and accountability of AI. ML explainability can help data scientists and developers better understand and interpret an ML model's behavior	(i04; i05; i08)	Thiebes et al. (2021)

5.3 Organizational objectives and success measures

Organizations should have specific metrics for organizational performance and productivity measures that they are seeking to improve. When success criteria are formally defined and then measured, IT project outcomes are improved, and project resources are better utilized (Thomas and Fernández 2008). As the CSF method included eliciting measures from the interviewee, the informants were asked about their individual and the organization's success measures. A primary organizational goal of the company is to become more data-driven, and the AI initiative is a subset of a larger venture to become more data-driven. In addition, the goal of the AI projects is to be the customers' first choice, which is aligned with the entire enterprise's vision. For example, a major AI project that estimates the exact delivery time of a parcel aims to provide better customer satisfaction. As a result, craftsmen can plan their workday more efficiently because they know when their package of tools arrives.

The AI projects' value is typically measured based on OKRs (Objectives and Key Results). OKRs are a management tool and a collaborative goal-setting methodology used by teams and individuals to set goals with measurable results. OKRs also enable tracking progress and create alignment. However, an informant states, "defining OKRs for an AI project is challenging because AI success is measured on several axes" (i02). Projects are often measured on commercial profitability and budget (e. g., Wateridge 1998). An inhibitor to AI services adoption is lack of ability to estimate the organizational and financial it will provide.

While the cloud computing paradigm shifts the location of the infrastructure to the network to reduce the costs associated with the management of hardware and software resources, several of the informants mention the difficulty of estimating ROI and thus the lack of proof of economic and commercial benefits of the AI projects. The reason for the difficulty is the absence of financial ratios. For IaaS and PaaS environments, challenges are often lack of cost control of the development and maintenance of ML models. An informant states that the solutions might be more expensive over time due to the lack of cost control. Additionally, "resources in the cloud may seem cheap, but many, seemingly cheap resources add up and can quickly become expensive" (i08). Besides, hosting costs are often overlooked when developing custom AI products. AI tools require a significant processing capacity, necessitating investment in cloud-based services capable of handling AI workloads. «It is difficult to get precise estimates on the costs before a project commences, but accurate estimates may arise moving ahead with the first deployment» (i09). Fortunately, cloud computing provides much flexibility – quickly. If a company invests in specialized hardware and expensive GPUs to run machine learning or deep learning algorithms in-house, they risk leaving the hardware in the dust when the projects are finished. On the other hand, cloud computing provides business value concerning resource utilization. AI is a technology that is in great change. It is risky to invest in something that is outdated tomorrow. An informant implies that although the hourly rate in the cloud is higher compared to the hourly rate for in-house resources, the cloud may be cheaper in the long term because the resources are scaled up and down when needed. A common denominator referred to by the informants is the fast time to market due to easy access to resources and tools, which may save time and costs related to internal expertise. Therefore, "it is reasonable to believe that AI services may be cheaper over time" (i02). Success measures elicited from the interviews are highlighted in the table below.

Table 11 – Success measures

Success measures		
Organizational success measures	Technical success measures	Customer value
Measure the efficiency of AI development – the time spent from idea to production	Monitor prediction and forecasting accuracy	Monitor the increase/decrease of API calls
Measure the degree of data scientist's satisfaction with their own work	Measure quality parameters agreed on in advance	Measure volume and turnover of the services for which you have launched AI solutions
Measure employees' level of conceptual knowledge of AI services	Measure the operability and number of algorithm errors	Conduct market analysis to measure the perceived innovativeness in the market
Monitor increase/decrease in technology interest and internal requests for new AI use cases using the cloud	Measure the level of automation in the solutions	Conduct market analysis and determine why customers choose/do not choose your services

If the quality of data is high, users will be able to produce better outputs. Quality data increases reliable outputs, improves the decision-making process, and lowers the risk of inaccurate outcomes. The quality of an algorithm will never be better than the quality of the data. The case company measures the quality of the algorithms – that they deliver within the quality parameters they have agreed on with the business managers. As discussed in the literature review, the trade-off between the service's result accuracy and responsiveness suggests increasing the service response time to provide better results or increasing responsiveness to achieve, for example, better customer engagement. Different models may result in different performances in terms of latency and accuracy, so selecting the proper model for a given task is essential. To ensure that the customers use the AI solutions and applications, it is possible to monitor the number of API calls. By measuring volume and turnover of the launched AI solutions and measuring traffic to the applications and how the solutions are used, it is possible to determine why the traffic goes up and down.

Solutions deployed in production are up and running every day. Therefore, the number of algorithm errors can be measured to address the robustness and stability of the ML models. In some situations, improving a model may not have anything to do with the data, techniques, model training, or hyper-parameters. Instead, it may be that the wrong question is being asked. "Consider looking at the problem from different angles and refine the problem definition that the ML model is supposed to solve" (i08). Leveraging the data to extract latent indicators and hidden relationships enables to refine the question.

An important measure is the efficiency of the development – the time from idea to production. An informant suggests that the cloud accelerates time to production due to an isolated environment on the Azure cloud. The autonomous environment makes it easier to achieve autonomous processes. When a solution is deployed in production, it needs to be full-fledged. An informant states, "data scientists should not have to go back to maintain, update, and be a janitor" (i05). However, data scientists can revisit the solutions later to make technology improvements or provide more training data samples to the algorithms that improve the solution. Furthermore, it is important that data scientists are satisfied with their execution of their own work. For instance, "success for me is knowing that the solution is close to the roof of what is possible to achieve, that one has solved the problem effectively, and provided a result that management is satisfied with" (i05). A similar response is, "tracking the downtime of employees due to IT issues may indicate the degree of development efficiency" (i09).

5.4 Summary of findings

Based on the literature review and the empirical data collection, this study has identified ten factors that drive- and 17 factors that inhibit the adoption of AI developer tools and infrastructure services. Inhibitors involve elements such as navigating lack of internal knowledge, lack of proof of economic benefits, legacy technology stacks, the responsibility for managing and maintaining the cloud environment, and assessing and complying with legal obligations and liabilities related to the location of data and the use of collected data for new AI applications. AI services inherit many of the same complexities as in-house AI development. However, the driving factors are mainly inherited in the affirmative characteristics of cloud computing. Hence, benefits include elements, such as on-demand access to resources and hardware, scalability and project flexibility.

Table 12 – Drivers and inhibitors of AI services adoption

DOI characteristics	Inhibitors (-) and drivers (+)	Source
Relative advantage	Customizability (+)	Lit, Int
	Control over data and AI solutions (+)	Lit, Int
	On-demand availability of computing resources and access to specialized hardware (+)	Lit, Int
	Direct link to cloud storage (+)	Lit
	Elastic provision of hardware resources and horizontal scalability (+)	Lit, Int

	Project and technology flexibility (+)	Lit, Int
	Cost-effective utilization of specialized hardware resources (+)	Lit, Int
	Data and privacy concerns (-)	Lit
Observability	Lack of measurable results (-)	Int
	Difficulty in estimating the total cost of a project (-)	Lit, Int
	Difficulty estimating ROI (-)	Lit, Int
	Lack of proof of economic and commercial benefits (-)	Int
	Lag between the release of new, innovative AI services offerings and the skills and awareness of users (-)	Int
	Fast time to value (+)	Lit. Int
	Several cheap cloud resources and services can become costly in the long-term (-)	Int
	Trade-off between time to value and data silos and maintenance costs (-)	Int
Compatibility	Network latency (-)	Lit
	Lack of responsiveness of resources located outside Europe (-)	Int
	Challenging integration process of the existing infrastructure/legacy technologies and the cloud environment (-)	Lit, Int
	Difficulties accessing workers with the required IT expertise and AI services specialists with domain expertise (-)	Lit, Int
	Sophisticated AI systems require massive amounts of data (-)	Lit, Int
Complexity	Challenging configuration processes and set-up, management, and maintenance of the AI services system (-)	Lit, Int
	GDPR compliance challenges related to the location of data and use of collected data for new AI solutions (-)	Int
	Uncertain governance and regulation compliance (-)	Lit
	Difficulties in assigning legal responsibilities and liabilities (-)	Lit, Int
Trialability	Opportunities to prevent vendor lock-in (+)	Int
	Subscription- and test of technologies for the time required (+)	Lit, Int

CSFs have been described as the characteristics, conditions, or variables that can significantly impact the success of an AI project when properly sustained, maintained, or managed (Milosevic and Patanakul 2005, p. 183). The critical success factors found in this study are summarized in table 10.

Table 13 – CSFs list

Affinity groups	Affinity sub-groups	Operational unit CSFs
Organizational factors	**Organizational structure**	Separate the data scientists and a data science department from other departments
		Build cross-functional teams
	Organizational culture and internal communication	Facilitate for experimentation
		Implement a company-wide innovation methodology
		Anchor effective communication between IT and business
	Agile experimentation	Apply an agile methodology
		Include experimentation in the iterative processes

		Agree on quality parameters of algorithms before embarking on an AI project
Cloud management factors	**Data management**	Implement a data strategy and processes for how to manage the data
		Incorporate a cloud strategy
		Ensure that the agreement with the cloud provider follows Norwegian data legislation and GDPR
	Infrastructure management	Have 5-10 data engineers per data scientist
		Collaborate with the cloud vendor
		Integrate on-premises infrastructure and the cloud environment
Technical factors	**Automation, scalability, and customizability**	Standardize pipelines and automate algorithms
		Establish MLOps and extensible system architectures
		Build a cloud platform that uses different technologies for different components
Technology commercialization process	**Exploration and experimentation**	Manage AI experiments
	Creation	Manage technology alignment
	Implementation	Ensure machine learning explainability

Given the lack of research on adopting and succeeding with AI developer tools and infrastructure services, this study identifies a promising field of research. By applying the CSF method and qualitative analysis of an organization that can point to successful adoption and organizational value from AI services based on a customized, autonomous cloud environment, this paper may provide important implications, which are elaborated in the following sections.

6. Implications

Implications aim to describe this study's impact on research and practice in the field of study. The principal findings of this study thus imposes implications for both researchers and practitioners.

6.1 Implications for research and practice

Extant literature emphasizes technical perspectives of AI services. Central elements include the set-up, the responsibility for maintenance, security, access controls, configuration, and maintenance of cloud environments, and requirements for expertise in configuration for training ML models and developing custom AI solutions. This study broadens the research paradigm by incorporating additional organizational mechanisms. While single case studies' generalizability is limited, they can provide an important insight into the direction of any future research. Previous research has presented CSFs in various formats, such as a CSF hierarchy, prioritization of CSFs, and creation of categorization schemes. This study suggests dependencies and interactions between the affinity groups revealed during the thematic analysis of the data. The conceptual model is formed after a conceptualization and generalization process, and it consists of fundamental components influencing a business' ability to harness the potential of AI. This study thus ventures into a research area that concerns interactions between contingencies of organizational factors and cloud-based AI technology. There is a lack of scholarly articles on CSF to harness AI capabilities through cloud computing services. While the CSF method is widely used in other research domains, such as ERP systems implementation (e. g., Saade and Nijher 2016; Mahraz et al. 2020), there are, to my knowledge, no CSFs studies explicitly conducted on AI developer and infrastructure services. Additionally, while papers on AI cloud services adoption exist, there is a lack of research on the phases from adoption to success of the technology.

Most machine learning models are conceived on a whiteboard, born on a laptop, and scaled on the cloud. AI is experimental in nature and requires autonomous, cross-functional teams, a flexible, customized cloud environment, and significant investment in culture, inter-organizational communication, and employee expertise to thrive. The findings in this study may impact practice in three critical areas of decision-making. 1) before the adoption of AI services and composition of a cloud environment, 2) during the development of AI products and the maintenance of the AI solutions and the cloud infrastructure, and 3) after AI solutions are deployed in production, including measures to ensure continuous business value. This study's contribution to practice aims to enable practitioners a foundation for decisions regarding the adoption and use of AI services. Depending on the organization's situation, this study can provide novice adopters prerequisites for successful adoption and organizations that have already implemented AI services recommendations for successful AI commercialization processes.

By comparing extant research with findings from the interviews, this study validates many elements identified in prior research. However, some previous research emphasizes complexity abstraction, and some even argue that AI services reduce the need for prior AI knowledge. A possible explanation for such propositions is that extant studies include AI software services and pre-built ML models, which can be accessed through simple API calls, thus reducing the need for data science and AI expertise. This study found that organizations need various professional resources constructed into cross-functional teams to harness AI capabilities efficiently. The identified CSFs may help business decision-makers and researchers pave the way for additional or more in-depth analysis of the factors.

The five characteristics related to the DOI theory have enabled to address the research questions due to its fit for practice. Although the interviews were based on the CSF method of conducting interviews and eliciting CSFs from the informants – themes, concepts, and elements revealed in the data are deemed relevant in the context of the DOI characteristics. As such, this study identified seven new inhibiting factors and one driving factor that, to my knowledge, is not discussed in extant literature. However, this study has limitations and gaps, which influence the relevance of the findings for each reader, but also provide opportunities for further research.

7. Limitations and future research avenues

As the implications section implies, this study provides a holistic picture of the studied phenomenon. The findings serve as a foundation for creating the conditions for success, which can be explored further through quantitative and qualitative data analysis. While the categorization and affinity grouping of CSFs can aid decision-makers in exploring the organization's current situation and adequately prepare decisions in a participatory way for future endeavors, this paper does not account for the weight of importance of the identified CSFs. This study is conducted in a large organization, and the industry in which the firm operates is based on a complex value chain, generating enormous amounts of data. Hence, the foremost limitation of this study is the limited generalizability. The findings may not benefit small organizations with less data and access to/finance for employee expertise. Additionally, organizations have different starting points, advantages, and disadvantages that influence the applicability of each success factor. Although the sub-objective is to identify the most complete and logical CSFs given the business conditions and use cases, it may not apply to all firms. Therefore, future research may benefit from conducting a quantitative research method to determine the weight of importance of each component related to various organizational circumstances and thus understand the statistical significance of each concept in relation to firm size, industry, or other organizational variables. In general, conducting a quantitative study may yield entirely different results compared to the findings of this study. Furthermore, this study is conducted in a Norwegian organization. Future research may need to replicate the study in another country to identify how the geographical location of an organization may affect the results.

Future qualitative research can form research questions based on one or more of the affinity groupings found in this study. For instance, data is the lifeblood of AI, and the cloud presents both challenges and opportunities for data management. As AI is scaled, it becomes increasingly challenging to manage, cleanse, maintain, and utilize data. It is nearly impossible to deploy AI at scale across an organization without proper data management methods and tools for managing the various aspects of a data lifecycle. Consequently, future research may benefit from a more in-depth study of data management practices on the cloud and how to mitigate GDPR compliance challenges related to the location of data and the use of collected data for new AI solutions.

Future research may bridge the gap in this study by conducting a qualitative analysis of cost control related to AI services. IT success criteria definitions (e. g., Wateridge 1998) suggest that commercial profitability is central for organizations when assessing project success. There is strong evidence that AI services lack observable financial benefits and commercial profitability. Future research may need to study cost control of AI services, including the entire process from adopting the technology, to the ideation of a solution and the cost of continuous improvements and maintenance of both AI solutions and the cloud environment.

The findings related to the DOI theory suggest factors that drive AI services adoption, which also include trade-offs or negative consequences. For example, customizability and control are reported by both literature and interviewees as driving factors. However, customizability requires AI services specialists with domain knowledge. This combination is rare and can be expensive to hire or train. Further, control over data and ML models enables monitoring how the ML model arrives at predictions, troubleshooting any inconsistencies, and adhering to GDPR legislation. Controversy, if users lack the required expertise or have performed bad configuration settings, the models may perform poorly, affecting accuracy and business outcomes. Additionally, the direct link to cloud storage facilitates easy access to the data on the cloud, which decreases the effort required to use AI services. However, since large-scale data is commonly distributed, provision of necessary computing power, storage capacity, and orchestration of various infrastructure components at different places have to be managed, which requires adequate technological or infrastructure knowledge. Finally, building the cloud infrastructure and AI environment on top of existing infrastructure enable fast time to value and, thus, immediate observable benefits. However, separating the cloud environment may also result in data silos and additional maintenance costs. So, what are the net benefits? How can business leaders make trade-off decisions? Future research may need to account for organizational circumstances, existing IT infrastructures, and business processes to make strategic decisions regarding the most effective ways to harness AI capabilities.

It can be challenging to parse and transform data into activity statements. Often, informants don't provide complete information and the responses may be ambiguous. Although the data analysis method aims to reduce bias, it may not be completely unavoidable due to extant knowledge, ideas, and experiences. There are many activity statements that are extracted from the interview data and some must be interpreted. Interpretation may increase risk of bias. Additionally, in some instances during the interviews, informants points out what he/she believes to be a CSF. However, this may or may not be a CSF for the whole organization or operational unit. The researcher must thus be careful not to be biased by the informants' intentions or beliefs.

Another limitation may be overlooked or limited access to literature. Additionally, AI services can include a hybrid of cloud computing deployment models, meaning that organizations can build AI software services on top of AI developer services, which rely on AI infrastructure services. This is leading to intertwined cloud supply chains. An opportunity for future research may be to nuance the knowledge presentation of each AI services stack relating to the respective cloud deployment models. For instance, future research may need to study the optional dependency relationships among AI software, developer, and infrastructure services and the implications for potential adopters.

8. Conclusion

This study contributes to IS and advanced analytics research by building a foundation underpinning AI and cloud computing research, using previously separated research streams to aggregate knowledge. Literature and industry reports indicate that many organizations fail to adopt AI and harness its full potential. Prior research mainly covers AI software services and applications relating to the conventional SaaS cloud model. Other research papers take a technical perspective when studying AI delivered through cloud computing services. While extant research lacks sufficient evidence of adoption and prerequisites for the success of AI infrastructure and developer services, this study aimed to fill the research gap by building on the DOI theory to analyze the drivers and inhibitors for adopting AI infrastructure and platform services. The findings resulted in 10 driving factors and 17 barriers to adoption. Mitigating adoption barriers and subsequently aiding practitioners in succeeding with AI services, this study has presented and discussed critical success factors based on a case study of an organization that has adopted and succeeded with AI infrastructure and platform services. By applying the Critical Success Factor (CSF) method, 20 sets of CSFs are categorized into affinity groups and sub-affinity groups.

Being aware of the pitfalls that may encounter allows decision-makers to circumnavigate them and subsequently capitalize on customer-centric and business-specific AI solutions. Additionally, the recent advancements and innovativeness of analytics firms and cloud providers seem to further enrich forthcoming AI services and aid organizations with state-of-the-art tools and resources.

While computer scientists and technical experts are the main drivers of prevalent research on AI services, this study call for more research taking a socio-technical and organizational perspective on AI services to foster adoption, diffusion, and innovation.

9. References

Alshenqeeti, Hamza. 2014. “Interviewing as a Data Collection Method: A Critical Review.” *English Linguistics Research* 3 (1): p39. https://doi.org/10.5430/elr.v3n1p39.

Bao, Bing-Kun, Yangyang Xiang, Lusong Li, Shuen Lyu, Harsh Munshi, and Honghong Zhu. 2018. “Scalable Cloud Service For Multimedia Analysis Based on Deep Learning.” In *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 1–4. San Diego, CA: IEEE. https://doi.org/10.1109/ICMEW.2018.8551579.

Bauer, Ingrid, Liudmila Zavolokina, and Gerhard Schwabe. 2020. “Is There a Market for Trusted Car Data?” *Electronic Markets* 30 (2): 211–25. https://doi.org/10.1007/s12525-019-00368-5.

Bhattacharjee, B., S. Boag, C. Doshi, P. Dube, B. Herta, V. Ishakian, K. R. Jayaram, et al. 2017. “IBM Deep Learning Service.” *IBM Journal of Research and Development* 61 (4/5): 10:1-10:11. https://doi.org/10.1147/JRD.2017.2716578.

Birt, Linda, Suzanne Scott, Debbie Cavers, Christine Campbell, and Fiona Walter. 2016. “Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation?” *Qualitative Health Research* 26 (13): 1802–11. https://doi.org/10.1177/1049732316654870.

Boag, Scott, Parijat Dube, Kaoutar El Maghraoui, Benjamin Herta, Waldemar Hummer, K. R. Jayaram, Rania Khalaf, Vinod Muthusamy, Michael Kalantar, and Archit Verma. 2018. “Dependability in a Multi-Tenant Multi-Framework Deep Learning as-a-Service Platform.” In *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*, 43–46. Luxembourg: IEEE. https://doi.org/10.1109/DSN-W.2018.00022.

Böhm, Markus, Stefanie Leimeister, Christoph Riedl, and Helmut Krcmar. 2011. “Cloud Computing – Outsourcing 2.0 or a new Business Model for IT Provisioning?” In *Application Management*, edited by Frank Keuper, Christian Oecking, and Andreas Degenhardt, 31–56. Wiesbaden: Gabler. https://doi.org/10.1007/978-3-8349-6492-2_2.

Boynton, Andrew C., and Robert W. Zmud. 1984. “An Assessment of Critical Success Factors.” *Sloan Management Review*, 1984. https://d1wqtxts1xzle7.cloudfront.net/30387665/6-assessment_csf-boynton-zmud-with-cover-page-v2.pdf?Expires=1644922229&Signature=GGigpUDFyimKg33ep~RI2JFLpFy1RHqtewZ8B~OFTviDaPsQozb~67RIU67q0bcgICBH8bGaiRcCeiwU6aJ6tqXHgoLmMut6rrNhDuXJ1OioBLnTAIxheoM04S2PG0pJO1FDaSeTiwl-aXJhsTutIym7m3OK73mSV46aRYe~jDM24z3aUeSAu6rA3b4YlIGpuH-~ZkRttESymfzAHxRnAnx6iJJpe73oNjMjTATHdL07hwHQAalJ4glqKnNeGpM5gPFBylFCvUpmIdC5nrBIwOL5c~OuSlTUVvitu62jxMaz8l3YAvhjEWR4-or3ZRLE6RY9D01jCE9GmlsoFlyTXA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA.

Braun, Virginia, and Victoria Clarke. 2006. “Using Thematic Analysis in Psychology.” *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Caballer, Miguel, Sahdev Zala, Álvaro López García, Germán Moltó, Pablo Orviz Fernández, and Mathieu Velten. 2018. “Orchestrating Complex Application Architectures in Heterogeneous Clouds.” *Journal of Grid Computing* 16 (1): 3–18. https://doi.org/10.1007/s10723-017-9418-y.

Caralli, Richard. 2004. “The Critical Success Factor Method: Establishing a Foundation for Enterprise Security Management,” 1298497 Bytes. https://doi.org/10.1184/R1/6585107.V1.

Cavaye, A.L.M. 1996. “Case Study Research: A Multi-Faceted Research Approach for IS.” *Information Systems Journal* 6 (3): 227–42. https://doi.org/10.1111/j.1365-2575.1996.tb00015.x.

Christenson, Dale, and Derek H.T. Walker. 2004. “Understanding the Role of ‘Vision’ in Project Success.” *Project Management Journal* 35 (3): 39–52. https://doi.org/10.1177/875697280403500306.

Chui, Michael, and Sankalp Malhotra. 2018. “AI Adoption Advances, but Foundational Barriers Remain.” https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain.

Cobbe, Jennifer, and Jatinder Singh. 2021. “Artificial Intelligence as a Service: Legal Responsibilities, Liabilities, and Policy Challenges.” *Computer Law & Security Review* 42 (September): 105573. https://doi.org/10.1016/j.clsr.2021.105573.

DalleMule, Leandro, and Thomas H. Davenport. 2017. “What’s Your Data Strategy?” *Harvard Business Review*, 2017. http://globaltraining.com.ec/files/i1.pdf.

Darke, Peta, Graeme Shanks, and Marianne Broadbent. 1998. “Successfully Completing Case Study Research: Combining Rigour, Relevance and Pragmatism.” *Information Systems Journal* 8 (4): 273–89. https://doi.org/10.1046/j.1365-2575.1998.00040.x.

Davenport, Thomas H. 2018. *The AI Advantage: How to Put the Artificial Intelligence Revolution to Work*. MIT Press.

Davenport, Thomas H., and Rajeev Ronanki. 2018. “Artificial Intelligence for the Real World.” https://hbr.org/2018/01/artificial-intelligence-for-the-real-world.

Deloitte. 2018. “State of AI in the Enterprise, 2nd Edition Early Adopters Combine Bullish Enthusiasm with Strategic Investments.” https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf.

Duan, Yucong, Qiang Duan, Xiaobing Sun, Guohua Fu, Nanjangud C. Narendra, Nianjun Zhou, Bo Hu, and Zhangbing Zhou. 2016. “Everything as a Service (XaaS) on the Cloud: Origins, Current and Future Trends.” *Services Transactions on Cloud Computing* 4 (2): 32–45. https://doi.org/10.29268/stcc.2016.0006.

Duong, Ta Nguyen Binh, and Nguyen Quang Sang. 2018. “Distributed Machine Learning on IAAS Clouds.” In *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, 58–62. Nanjing, China: IEEE. https://doi.org/10.1109/CCIS.2018.8691150.

Dutta, Sourav. 2018. “An Overview on the Evolution and Adoption of Deep Learning Applications Used in the Industry.” *WIREs Data Mining and Knowledge Discovery* 8 (4). https://doi.org/10.1002/widm.1257.

Eisenhardt, Kathleen M. 1989. “Building Theories from Case Study Research.” *Academy of Management Review* 14 (4): 532–50. https://doi.org/10.5465/amr.1989.4308385.

Elluri, Lavanya, and Karuna Pande Joshi. 2018. “A Knowledge Representation of Cloud Data Controls for EU GDPR Compliance.” In *2018 IEEE World Congress on Services (SERVICES)*, 45–46. San Francisco, CA: IEEE. https://doi.org/10.1109/SERVICES.2018.00036.

Elshawi, Radwa, Sherif Sakr, Domenico Talia, and Paolo Trunfio. 2018. “Big Data Systems Meet Machine Learning Challenges: Towards Big Data Science as a Service.” *Big Data Research* 14 (December): 1–11. https://doi.org/10.1016/j.bdr.2018.04.004.

Esposito, Dino, and Francesco Esposito. 2020. “Introducing Machine Learning”. Microsoft Press. http://ptgmedia.pearsoncmg.com/images/9780135565667/samplepages/9780135565667_Sample.pdf.

Ferguson, Charles R., and Roger Dickinson. 1982. “Critical Success Factors for Directors in the Eighties.” *Business Horizons* 25 (3): 14–18. https://doi.org/10.1016/0007-6813(82)90123-9.

Fountaine, Tim, Brian McCarthy, and Tamim Saleh. 2019. “Building the AI-Powered Organization.” https://hbr.org/2019/07/building-the-ai-powered-organization.

Gartner, Inc. 2020. “Gartner Top 10 Trends in Data and Analytics for 2020.” 2020. https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020.

Girasa, Rosario. 2020. “AI as a Disruptive Technology.” In *Artificial Intelligence as a Disruptive Technology*, by Rosario Girasa, 3–21. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-35975-1_1.

Gloet, Marianne, and Milé Terziovski. 2004. “Exploring the Relationship between Knowledge Management Practices and Innovation Performance.” *Journal of Manufacturing Technology Management* 15 (5): 402–9. https://doi.org/10.1108/17410380410540390.

Godse, Manish, and Shrikant Mulik. 2009. “An Approach for Selecting Software-as-a-Service (SaaS) Product.” In *2009 IEEE International Conference on Cloud Computing*, 155–58. Bangalore, India: IEEE. https://doi.org/10.1109/CLOUD.2009.74.

Greenhalgh, Trisha, Glenn Robert, Fraser Macfarlane, Paul Bate, and Olivia Kyriakidou. 2004. “Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations.” *The Milbank Quarterly* 82 (4): 581–629. https://doi.org/10.1111/j.0887-378X.2004.00325.x.

Gu, Vicky Ching, James J. Hoffman, Qing Cao, and Marc J. Schniederjans. 2014. “The Effects of Organizational Culture and Environmental Pressures on IT Project Performance: A Moderation Perspective.” *International Journal of Project Management* 32 (7): 1170–81. https://doi.org/10.1016/j.ijproman.2013.12.003.

Gujarati, Arpan, Sameh Elnikety, Yuxiong He, Kathryn S. McKinley, and Björn B. Brandenburg. 2017. “Swayam: Distributed Autoscaling to Meet SLAs of Machine Learning Inference Services with Resource Efficiency.” In *Proceedings of the 18th ACM/IFIP/USENIX Middleware Conference*, 109–20. Las Vegas Nevada: ACM. https://doi.org/10.1145/3135974.3135993.

Haenlein, Michael, and Andreas Kaplan. 2019. “A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence.” *California Management Review* 61 (4): 5–14. https://doi.org/10.1177/0008125619864925.

Hahn, Trey, and Marco te Brömmelstroet. 2021. “Collaboration, Experimentation, Continuous Improvement: Exploring an Iterative Way of Working in the Municipality of Amsterdam’s Bicycle Program.” *Transportation Research Interdisciplinary Perspectives* 9 (March): 100289. https://doi.org/10.1016/j.trip.2020.100289.

Halpern, Matthew, Behzad Boroujerdian, Todd Mummert, Evelyn Duesterwald, and Vijay Janapa Reddi. 2019. “One Size Does Not Fit All: Quantifying and Exposing the Accuracy-Latency

Trade-Off in Machine Learning Cloud Service APIs via Tolerance Tiers.” In *2019 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 34–47. Madison, WI, USA: IEEE. https://doi.org/10.1109/ISPASS.2019.00012.

Hanzlik, Lucjan, Yang Zhang, Kathrin Grosse, Ahmed Salem, Maximilian Augustin, Michael Backes, and Mario Fritz. 2021. “MLCapsule: Guarded Offline Deployment of Machine Learning as a Service.” In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 3295–3304. Nashville, TN, USA: IEEE. https://doi.org/10.1109/CVPRW53098.2021.00368.

Hsu, Chin-Lung, and Judy Chuan-Chuan Lin. 2016. “Factors Affecting the Adoption of Cloud Services in Enterprises.” *Information Systems and E-Business Management* 14 (4): 791–822. https://doi.org/10.1007/s10257-015-0300-9.

Hyon, Matt, and Bernard Apolinario. 2019. “Custom AI Models with Azure Machine Learning Studio and ML.NET.” *Custom AI Models with Azure Machine Learning Studio and ML.NET* (blog). April 2019. https://devblogs.microsoft.com/premier-developer/custom-ai-models-with-azure-machine-learning-studio-and-ml-net/.

Jarrahi, Mohammad Hossein. 2018. “Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making.” *Business Horizons* 61 (4): 577–86. https://doi.org/10.1016/j.bushor.2018.03.007.

Javadi, Seyyed Ahmad, Richard Cloete, Jennifer Cobbe, Michelle Seng Ah Lee, and Jatinder Singh. 2020. “Monitoring Misuse for Accountable ‘Artificial Intelligence as a Service.’” In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 300–306. New York NY USA: ACM. https://doi.org/10.1145/3375627.3375873.

Javadi, Seyyed Ahmad, Chris Norval, Richard Cloete, and Jatinder Singh. 2021. “Monitoring AI Services for Misuse.” In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 597–607. Virtual Event USA: ACM. https://doi.org/10.1145/3461702.3462566.

Jensen, Morten Berg, Björn Johnson, Edward Lorenz, and Bengt Åke Lundvall. 2007. “Forms of Knowledge and Modes of Innovation.” *Research Policy* 36 (5): 680–93. https://doi.org/10.1016/j.respol.2007.01.006.

Jordan, M. I., and T. M. Mitchell. 2015. “Machine Learning: Trends, Perspectives, and Prospects.” *Science* 349 (6245): 255–60. https://doi.org/10.1126/science.aaa8415.

Joslin, Robert, and Ralf Müller. 2016. “The Impact of Project Methodologies on Project Success in Different Project Environments.” *International Journal of Managing Projects in Business* 9 (2): 364–88. https://doi.org/10.1108/IJMPB-03-2015-0025.

Kaplunovich, Alex, and Yelena Yesha. 2017. “Cloud Big Data Decision Support System for Machine Learning on AWS: Analytics of Analytics.” In *2017 IEEE International Conference on Big Data (Big Data)*, 3508–16. Boston, MA: IEEE. https://doi.org/10.1109/BigData.2017.8258340.

Karmaker, Shubhra Kanti, Md. Mahadi Hassan, Micah J. Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. 2022. “AutoML to Date and Beyond: Challenges and Opportunities.” *ACM Computing Surveys* 54 (8): 1–36. https://doi.org/10.1145/3470918.

Lai, Ming-Fong, and Gwo-Guang Lee. 2007. “Risk-avoiding Cultures toward Achievement of Knowledge Sharing.” *Business Process Management Journal* 13 (4): 522–37. https://doi.org/10.1108/14637150710763559.

Lifshitz-Assaf, Hila, Sarah Lebovitz, and Lior Zalmanson. 2018. "The Art of Balancing Autonomy and Control." *MIT Sloan Management Review*, 2018. https://www.hbs.edu/ris/Publication%20Files/The%20Art%20of%20Balancing%20Autonomy%20and%20Control_27216126-cf32-40e5-8907-fe109f0edafa.pdf.

Lins, Sebastian, Konstantin D. Pandl, Heiner Teigeler, Scott Thiebes, Calvin Bayer, and Ali Sunyaev. 2021. "Artificial Intelligence as a Service: Classification and Research Directions." *Business & Information Systems Engineering* 63 (4): 441–56. https://doi.org/10.1007/s12599-021-00708-w.

Lopes, Bernardo, Pedro Martins, José Domingues, and Manuel Au-Yong-Oliveira. 2019. "The Future Employee: The Rise of AI in Portuguese Altice Labs." In *New Knowledge in Information Systems and Technologies*, edited by Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis, and Sandra Costanzo, 932:336–47. Advances in Intelligent Systems and Computing. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-16187-3_33.

Lopez Garcia, Alvaro, Jesus Marco De Lucas, Marica Antonacci, Wolfgang Zu Castell, Mario David, Marcus Hardt, Lara Lloret Iglesias, et al. 2020. "A Cloud-Based Framework for Machine Learning Workloads and Applications." *IEEE Access* 8: 18681–92. https://doi.org/10.1109/ACCESS.2020.2964386.

Love, James H., and Stephen Roper. 2009. "Organizing Innovation: Complementarities between Cross-Functional Teams." *Technovation* 29 (3): 192–203. https://doi.org/10.1016/j.technovation.2008.07.008.

Magistretti, Stefano, Claudio Dell'Era, and Antonio Messeni Petruzzelli. 2019. "How Intelligent Is Watson? Enabling Digital Transformation through Artificial Intelligence." *Business Horizons* 62 (6): 819–29. https://doi.org/10.1016/j.bushor.2019.08.004.

Mahraz, Mohamed-Iliasse, Loubna Benabbou, and Abdelaziz Berrado. 2020. "A Compilation and Analysis of Critical Success Factors for the ERP Implementation:" *International Journal of Enterprise Information Systems* 16 (2): 107–33. https://doi.org/10.4018/IJEIS.2020040107.

Mazzei, Alessandra. 2010. "Promoting Active Communication Behaviours through Internal Communication." *Corporate Communications: An International Journal* 15 (3): 221–34. https://doi.org/10.1108/13563281011068096.

Mell, Peter, and Timothy Grance. 2011. "The NIST Definition of Cloud Computing." NIST Special Publication 800-145. National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

Milosevic, Dragan, and Peerasit Patanakul. 2005. "Standardized Project Management May Increase Development Projects Success." *International Journal of Project Management* 23 (3): 181–92. https://doi.org/10.1016/j.ijproman.2004.11.002.

Mir, Umar Bashir, Arpan K. Kar, M. P. Gupta, and R. S. Sharma. 2019. "Prioritizing Digital Identity Goals – The Case Study of Aadhaar in India." In *Digital Transformation for a Sustainable Society in the 21st Century*, edited by Ilias O. Pappas, Patrick Mikalef, Yogesh K. Dwivedi, Letizia Jaccheri, John Krogstie, and Matti Mäntymäki, 11701:489–501. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29374-1_40.

Munro, Malcolm C., and Basil R. Wheeler. 1980. "Planning, Critical Success Factors, and Management's Information Requirements." *MIS Quarterly* 4 (4): 27. https://doi.org/10.2307/248958.

Myers, Michael. 2015. "Ten Years of Qualitative Research in Organizations and Management: Some Reflections." *Qualitative Research in Organizations and Management: An International Journal* 10 (4): 337–39. https://doi.org/10.1108/QROM-07-2015-1307.

Nelson, Kay M., and Jay G. Cooprider. 1996. "The Contribution of Shared Knowledge to IS Group Performance." *MIS Quarterly* 20 (4): 409. https://doi.org/10.2307/249562.

Oates, Briony J. 2006. *Researching Information Systems and Computing*. London; Thousand Oaks, Calif.: SAGE Publications.

Oliveira, Tiago, Manoj Thomas, and Mariana Espadanal. 2014. "Assessing the Determinants of Cloud Computing Adoption: An Analysis of the Manufacturing and Services Sectors." *Information & Management* 51 (5): 497–510. https://doi.org/10.1016/j.im.2014.03.006.

Pandl, Konstantin D., Heiner Teigeler, Sebastian Lins, Scott Thiebes, and Ali Sunyaev. 2021. "Drivers and Inhibitors for Organizations' Intention to Adopt Artificial Intelligence as a Service." In . https://doi.org/10.24251/HICSS.2021.215.

Parker, Charlie, Sam Scott, and Alistair Geddes. 2020. "Snowball Sampling." In *SAGE Research Methods Foundations*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781526421036831710.

Powers, Richard F., and Gary W. Dickson. 1973. "MisProject Management: Myths, Opinions, and Reality." *California Management Review* 15 (3): 147–56. https://doi.org/10.2307/41164448.

Press, Gil. 2020. "AI Stats News: Only 14.6% Of Firms Have Deployed AI Capabilities In Production." January 13, 2020. https://www.forbes.com/sites/gilpress/2020/01/13/ai-stats-news-only-146-of-firms-have-deployed-ai-capabilities-in-production/?sh=651a47972650.

Rogers, E.M., Diffusion of innovations, 4th edn., Free Press, New York, 1995.

Rai, Arun, Panos Constantinides, and Saonee Sarker. 2019. "Next-Generation Digital Platforms: Toward Human–AI Hybrids." *MIS Quarterly* 43 (1): iii–ix. https://misq.umn.edu/skin/frontend/default/misq/pdf/V43I1/EdComments_V43_I1.pdf.

Rayome, Alison DeNisco. 2019. "Why 85% of AI Projects Fail," 2019. https://www.techrepublic.com/article/why-85-of-ai-projects-fail/.

Ribeiro, Mauro, Katarina Grolinger, and Miriam A.M. Capretz. 2015. "MLaaS: Machine Learning as a Service." In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, 896–902. Miami, FL, USA: IEEE. https://doi.org/10.1109/ICMLA.2015.152.

Rockart, John F. 1979. "Chief Executives Define Their Own Data Needs." *Harvard Business Review*, 1979. https://hbr.org/1979/03/chief-executives-define-their-own-data-needs.

Rockart, John F., and Christine V. Bullen. 1981. "A Primer on Critical Success Factors." *Center for Information Systems Research, Sloan School of Management, M.I.T*, 1981. https://dspace.mit.edu/bitstream/handle/1721.1/1988/SWP-1220-08368993-CISR-069.pdf?sequen.

Romero, Francisco, Qian Li, Neeraja J. Yadwadkar, and Christos Kozyrakis. 2019. "INFaaS: A Model-Less and Managed Inference Serving System." https://arxiv.org/pdf/1905.13348.pdf.

Rouhani, Bita Darvish, Siam Umar Hussain, Kristin Lauter, and Farinaz Koushanfar. 2018. "ReDCrypt: Real-Time Privacy-Preserving Deep Learning Inference in Clouds Using FPGAs." *ACM Transactions on Reconfigurable Technology and Systems* 11 (3): 1–21. https://doi.org/10.1145/3242899.

Rubin, Herbert J., and Irene S. Rubin. 2011. *Qualitative Interviewing: The Art of Hearing Data*. SAGE Publications, Incorporated.

Sakr, Sherif, Anna Liu, Daniel M. Batista, and Mohammad Alomari. 2011. "A Survey of Large Scale Data Management Approaches in Cloud Environments." *IEEE Communications Surveys & Tutorials* 13 (3): 311–36. https://doi.org/10.1109/SURV.2011.032211.00087.

Sallehudin, Hasimi, Razli Che Razak, Mohammad Ismail, Ahmad Firdause Md Fadzil, and Rogis Baker. 2018. "Cloud Computing Implementation in The Public Sector: Factors and Impact." *Asia-Pacific Journal of Information Technology and Multimedia* 07 (02(02)): 27–42. https://doi.org/10.17576/apjitm-2018-0702(02)-03.

Salomoni, D., I. Campos, L. Gaido, J. Marco de Lucas, P. Solagna, J. Gomes, L. Matyska, et al. 2018. "INDIGO-DataCloud: A Platform to Facilitate Seamless Access to E-Infrastructures." *Journal of Grid Computing* 16 (3): 381–408. https://doi.org/10.1007/s10723-018-9453-3.

Serrano, Nicolas, Gorka Gallardo, and Josune Hernantes. 2015. "Infrastructure as a Service and Cloud Technologies." *IEEE Software* 32 (2): 30–36. https://doi.org/10.1109/MS.2015.43.

Serrano, Nicolas, Josune Hernantes, and Gorka Gallardo. 2014. "Service-Oriented Architecture and Legacy Systems." *IEEE Software* 31 (5): 15–19. https://doi.org/10.1109/MS.2014.125.

Shaukat, Zeeshan, Faheem Akhtar, Juan Fang, Saqib Ali, and Muhammad Azeem. 2018. "Cloud Based Face Recognition for Google Glass." In *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence - ICCAI 2018*, 104–11. Chengdu, China: ACM Press. https://doi.org/10.1145/3194452.3194479.

Stanciu, Alexandru, Aurel Mihail Titu, and Dorin Vasile Deac-Suteu. 2021. "Driving Digital Transformation Of Knowledge-Based Organizations Through Artificial Intelligence Enabled Data Centric, Consumption Based, As-A-Service Models." In *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 1–8. Pitesti, Romania: IEEE. https://doi.org/10.1109/ECAI52376.2021.9515172.

Thiebes, Scott, Sebastian Lins, and Ali Sunyaev. 2021. "Trustworthy Artificial Intelligence." *Electronic Markets* 31 (2): 447–64. https://doi.org/10.1007/s12525-020-00441-4.

Thomas, Gary. 2011. *How to Do Your Case Study: A Guide For Students And Researchers*. London: Sage Publications Limited.

Thomas, Graeme, and Walter Fernández. 2008. "Success in IT Projects: A Matter of Definition?" *International Journal of Project Management* 26 (7): 733–42. https://doi.org/10.1016/j.ijproman.2008.06.003.

Tramèr, Florian, Fan Zhang, Ari Juels, Michael K. Reiter, and Ristenpart Ristenpart. 2016. "Stealing Machine Learning Models via Prediction APIs." In *Proceedings of the 25th USENIX Conference on Security Symposium*, 601–18. USENIX Association.

Truex, Stacey, Ling Liu, Mehmet Emre Gursoy, Lei Yu, and Wenqi Wei. 2019. "Demystifying Membership Inference Attacks in Machine Learning as a Service." *IEEE Transactions on Services Computing*, 1–1. https://doi.org/10.1109/TSC.2019.2897554.

Visma. 2021. *Hvordan Lykkes Med AI, Optimering Og Automatisering*. Oslo. https://www.visma.no/consulting/ai-optimering-automatisering-opptak/?utm_source=Eloqua&utm_medium=email&utm_content=NO_CN%20211110%20AI%20optimering%20autmatisering_alle_NY%20mal&utm_campaign=NO_CN%20211110%20AI%20optimering%20automatisering_alle&optin=.

Wang, Wei, Jinyang Gao, Meihui Zhang, Sheng Wang, Gang Chen, Teck Khim Ng, Beng Chin Ooi, Jie Shao, and Moaz Reyad. 2018. "Rafiki: Machine Learning as an Analytics Service System." *Proceedings of the VLDB Endowment* 12 (2): 128–40. https://doi.org/10.14778/3282495.3282499.

Wang, Weiyu, and Keng Siau. 2019. "Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda." *Journal of Database Management* 30 (1): 61–79. https://doi.org/10.4018/JDM.2019010104.

Wateridge, John. 1998. "How Can IS/IT Projects Be Measured for Success?" *International Journal of Project Management* 16 (1): 59–63. https://doi.org/10.1016/S0263-7863(97)00022-7.

Webster, Jane, and Richard T. Watson. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review." *MIS Quarterly* 26 (2): 13–23.

Yang, Haibo, and Mary Tate. 2012. "A Descriptive Literature Review and Classification of Cloud Computing Research." *Communications of the Association for Information Systems* 31. https://doi.org/10.17705/1CAIS.03102.

Yao, Yuanshun, Zhujun Xiao, Bolun Wang, Bimal Viswanath, Haitao Zheng, and Ben Y. Zhao. 2017. "Complexity vs. Performance: Empirical Analysis of Machine Learning as a Service." In *Proceedings of the 2017 Internet Measurement Conference*, 384–97. London United Kingdom: ACM. https://doi.org/10.1145/3131365.3131372.

Yin, Robert K. 1994. "Discovering the Future of the Case Study Method in Evaluation Research." *Evaluation Practice* 15 (3): 283–90. https://doi.org/10.1016/0886-1633(94)90023-X.

Zapadka, Patryk, André Hanelt, Sebastian Firk, and Jana Oehmichen. n.d. "Leveraging '"AI-as-a-Service"' – Antecedents and Consequences of Using Artificial Intelligence Boundary Resources." *Proceedings of the 41st International Conference on Information Systems*, 1–14.

10. Appendices

Appendix A – Informed consent

Forespørsel om deltakelse i forskningsprosjektet:

«Harnessing Artificial Intelligence Capabilities Through Cloud Services – A Case Study of Inhibitors and Success Factors»

Forskningsprosjektet er del av sluttfasen i masterstudiet "Master of Science in Information Systems – Digital Business Systems"

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å identifisere kritiske suksessfaktorer for å lykkes med adopsjon av kunstig intelligens skytjenester og utvikling av kunstig intelligens i skyen. Dette skrivet informerer om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål
Prosjektets formål er å øke graden av kunnskap om kunstig intelligens og skyteknologi. Prosjektet ønsker å identifisere hvilke muligheter og utforinger bedrifter kan støte på under adopsjon av teknologien, i tillegg til prosessene/fasene fra idé til produksjon av AI i et skymiljø. Målet med prosjektet er å kartlegge kritiske suksessfaktorer bedrifter må legge til grunn for å lykkes med adopsjon og utvikling av AI ved hjelp av skytjenester, samt hvordan bedrifter kan sørge for en kontinuerlig verdiskapning. Motivasjonen mak temaet er knyttet til fordelene ved økt digitalisering og konkurransekraft, samt effektivisere måten bedrifter jobber på.

Hvem er ansvarlig for forskningsprosjektet?
Høyskolen Kristiania, Kirkegata 24-26, 0153 Oslo er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?
For å oppnå økt kunnskap om AI og skyteknologi, samt forstå hvordan bedrifter kan lykkes for å skape forretningsverdi, er det hensiktsmessig å rekruttere fagfolk med erfaring og forståelse for faget og tematikken. I tillegg er det formålstjenlig å rekruttere deltakere som allerede har vært med på å skape forretningsverdi og suksess med utvikling og bruk av AI gjennom skytjenester. Denne henvendelsen vil derfor gå til fagpersoner innenfor data science, BI og datavarehus, samt personer innen IT og prosjektledelse.

Hva innebærer det for deg å delta?
Prosjektet baseres på en kombinasjon av sekundær data (forskningslitteratur) og empiri. Empirisk datainnsamling baseres på en kvalitativ case studie og vil foregå ved intervju(er). Intervjuer vil foregå etter avtalt dato, tid og sted med deltakerne. Intervjutiden vil vare i henhold til intervjuguiden*, estimert tid vil være ca. 1 time. Datainnsamling vil starte i januar

2022 og avsluttes da tilstrekkelig data er samlet inn og evaluert. Antall intervjuer og antall deltakere vil bli avtalt med respondentene.

Intervjuene har som formål å kartlegge kritiske suksessfaktorer for å lykkes med AI i skyen, og temaene innebærer blant annet utvikling og bruk av AI i skyen, prosjektledelse, utfordringer, og fasene fra idé til produksjon.

Det vil bli tatt lydopptak av intervjuer med mindre deltakeren ikke ønsker dette. Opptak av stemmer regnes som personopplysning, derfor er prosjektet meldt inn til NSD. Lydopptak vil bli transkribert og oversatt til engelsk. Intervjudataen vil være konfidensiell. Opplysninger om deltakerne blir behandlet basert på deres samtykke.

**Intervjuguiden beskriver temaer som vil ble tatt opp i intervjuet, samt hvilke spørsmål som vil bli stilt. Deltakere kan få tildelt intervjuguiden i forkant av intervjuet om det er ønskelig*

Deltakerens rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- Å be om innsyn, retting, sletting, begrensning, dataportabilitet og retten til å klage til Datatilsynet
- Å få utlevert en kopi av opplysningene
- Å trekke seg uten begrunnelse (lovhjemlet i personopplysningslovens artikkel 7, del 3 og artikkel 17)
- Å gi informert samtykke (lovhjemlet i Personvernforordningens artikkel 4 (del 11), artikkel 6 (del 1A) og artikkel 7)
- Anonymitet
- Konfidensialitet
- Behandling av personopplysninger er lovhjemlet i personopplysningslovens artikkel 4, del 1.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil kun bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

Jeg vil herunder erklære at:
1. Studenten og prosjektansvarlig (studentens veileder) har taushetsplikt.
2. Det vil bli tatt lydopptak av intervjuer med mindre deltakeren ikke ønsker dette.

3. Lydopptak vil bli transkribert og oversatt til engelsk.
4. Intervjudataen vil være konfidensiell.
5. Intervjudataen vil aldri bli delt innenfor organisasjonen (høyskolen), med konkurrenter eller i noen annen kommersiell form.
6. Intervjudata vil bli anonymisert og lagret elektronisk uten deltakernes personopplysninger eller deres tilknytning. (Navnet og kontaktopplysningene dine vil bli erstattet med en kode som lagres på egen navneliste adskilt fra øvrige data)
7. Dataene vil bli lagret kryptert på forskerens datamaskin for å forhindre ulovlig bruk i tilfelle tyveri eller inntrenging.
8. Navn og tilknytninger vil ikke bli publisert i forskningspublikasjonen (med mindre intervjuobjektene/bedriftene oppgir noe annet).
9. Intervjudataen vil danne grunnlag for det akademiske prosjektet.
10. Kun student og veileder ved behandlingsansvarlig institusjon vil ha tilgang på dataen
11. All data vil bli sendt til relevante intervjuobjekter før publisering, for å bekrefte at synspunktene deres er tolket riktig.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?
Opplysningene anonymiseres når prosjektet avsluttes/oppgaven er godkjent, som er 25.05.2022. Personopplysninger og eventuelle opptak vil bli slettet ved prosjektslutt.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.
På oppdrag fra Høyskolen Kristiania, Kirkegata 24-26, 0153 Oslo har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Prosjektansvarlig (studentens veileder), Moutaz Haddara, Professor i Informasjonssystemer ved Høyskolen Kristiania: E-post: moutaz.haddara@kristiania.no
- Vårt personvernombud: Kontaktinformasjon: personvernombud@kristiania.no.

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:
- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 53 21 15 00.

Med vennlig hilsen,

Prosjektansvarlig
[redacted]

Student
[redacted]

Dato Dato

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «Harnessing Artificial Intelligence Capabilities Through Cloud Services – A Case Study of Inhibitors and Success Factors», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i intervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

Prosjektdeltaker Dato

Appendix B – NSD project approval

Vurdering

Referansenummer
217650

Prosjekttittel
Kunstig intelligens i skyen

Behandlingsansvarlig institusjon
Høyskolen Kristiania – Ernst G. Mortensens Stiftelse / School of Economics, Innovation, and Technology / institutt for teknologi

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)
[redacted]

Type prosjekt
Studentprosjekt, masterstudium

Kontaktinformasjon, student
[redacted]

Prosjektperiode
03.01.2022 - 25.05.2022

Vurdering (1)

06.01.2022 - Vurdert
Det er vår vurdering at behandlingen vil være i samsvar med personvernlovgivningen, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet 06.01.2022 med vedlegg, samt i meldingsdialogen mellom innmelder og Personverntjenester. Behandlingen kan starte.

TYPE OPPLYSNINGER OG VARIGHET Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 25.05.2022.

LOVLIG GRUNNLAG Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER Vi vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om: - lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen - formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål - dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet - lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet DE

REGISTRERTES RETTIGHETER Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), og dataportabilitet (art. 20). Vi vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13. Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER Vi legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32). Ved bruk av databehandler (spørreskjemaleverandør, skylagring eller videosamtale) må behandlingen oppfylle kravene til bruk av databehandler, jf. art 28 og 29. Bruk leverandører som din institusjon har avtale med. For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til oss ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema Du må vente på svar fra oss før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET Vi vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Appendix C – Interview guide

Interview questions and topics are modified according to the informant's position, function and/or affiliation. This interview guide incorporates all the asked questions based on the informants' function, role, and/or affiliation.

Informantens rolle/stilling:
Avdeling:
Informantens rolle innenfor AI prosjekter

1. Hvilken funksjon eller rolle har du knyttet til deres AI prosjekter?
2. Hva ønsker du/din funksjon å oppnå med prosjektet/prosjektene?
3. Hva er viktig for at du skal oppnå dette målet?
4. Har du jobbet med kunstig intelligens på en tradisjonell teknologistack/selvdreven infrastruktur? Hvis ja, hva mener du er den største forskjellen mellom det kontra det å jobbe med AI i skyen?

Generell innledning

1. Hvilken skyleverandør bruker dere?
2. Hvorfor valgte dere denne leverandøren?
3. Hvilken skytjenestemodell bruker dere for å utvikle AI?
4. Hvorfor valgte dere dette abstraksjonsnivået?
5. Når startet dere å utvikle AI/ML via skyen?
6. Hvorfor startet dere med å utvikle AI gjennom skyteknologi?
7. Hvor mange er involvert i AI prosjektene?
8. Hvilke forretningsprosesser bruker dere AI løsningene til?

Mål og måleenheter

9. Hva er bedriften overordnede mål med AI prosjektene?
10. Hvilke fordeler eller forretningsverdi har dere fått ut av å drive med AI i skyen?
11. Det finnes ingen felles måleenhet for suksess, hvordan vil du/dere definere suksess – når vil du/dere si at et AI-prosjekt har lykkes?
12. Hva må til for at dere skal nå målene?
13. Hvordan måler dere ROI med et AI-prosjekt?

Utfordringer/muligheter

14. Mener du skytjenester har akselerert utviklingen av AI i deres bedrift? Hvis ja/nei, hvorfor?
15. Hvilke problem(er) ønsker dere å løse?
16. Hvordan har deres suksess med AI påvirket organisasjonen?
17. Hvilke utfordringer kan være knyttet til å jobbe med AI skytjenester?
18. Er det utfordringer knyttet til AI og skyteknologi som ikke ville vært der om dere hadde utviklet AI på selvdreven infrastruktur/tradisjonell teknologistack?
19. Kan du fortelle om noen utfordringer dere har støtt på?
20. Er det økonomiske fordeler eller ulemper ved å utvikle AI gjennom skyteknologi?
21. Hvilke risikoer er knyttet til AI skytjenester?
22. Kan du fortelle om mulighetene knyttet til å bruke AI skytjenester?
23. Hva mener du er det viktigste for at dere skal lykkes med et AI-prosjekt?

Ide til produksjon

24. Hvilke faser går dere igjennom fra idé til produksjon?

25. Er disse fasene/prosessene påvirket av at AI utvikles gjennom skytjenester?
26. Hvor stor del av utviklingsprosjekter/ideer blir sendt til produksjon?
27. Hva mener du skal til for at en idé blir sendt til produksjon?
28. Hva er den viktigste kunnskapen man må ha for å lykkes med AI i skyen?
29. Hva skal til for at en AI-løsning fortsetter å skape forretningsverdi etter at den er satt i produksjon?
30. Er det spesielle karakteristikker for AI og skyteknologi som gjør at dere må jobbe annerledes for å utvikle løsninger? F.eks., sammenlignet med AI på en selvdreven infrastruktur.

Annet

31. Er det noe du ønsker å legge til som du mener kan bidra til å lykkes med AI i skyen?

Appendix D – Activity statements, affinity grouping, and supporting themes based on the raw data

Affinity group: Organizational factors	
Activity statements	Supporting themes
Provide data scientists with room for experimentation and exploration of data and AI solutions	Ensure low degree of formalization and standardization
Enable a flexible work environment	
Vary employee's´ tasks	
Work with AI projects as a program	
Data analysts and data engineers capture the right data and make it accessible for the data scientists to explore it	
Test new ideas and challenge established truths	Decentralize decision-making
Demystify decision-making processes	
Set project goals and technology goals in advance of project execution	
The IT department should manage the cloud infrastructure and make it accessible for the data scientists	
Business managers do not tell IT how they should execute projects	
Preserve employee's motivation	Enable teams to work autonomously
Minimize fear of failure	
Create an environment where unproven tools and solutions are acceptable	
Work interdisciplinary	
Separate data scientists from other departments	
Separate a data science department from other departments	
Avoid inhibiting old processes and internal politics	
Reinforce a shared understanding of AI services concepts	Create a shared understanding of technical concepts and a common language for innovation
Set guidelines for how innovation is managed and maintained	
Communicate the capabilities of AI and cloud on a conceptual level	
Encourage regular communication between businesspeople and IT people	
Avoid heavily loaded technical terminologies	
Develop trust and overcome spatial and organizational barriers	
Explore conditions in the organization´s datasets to inform managers on business conditions	
Arrange data science workshops for employees that is not a formalized part of the AI initiatives	
Communication must be engaging, explanatory, and selling	

Affinity group: Cloud management factors	
Activity statements	Supporting themes
Preserve GDPR and integrate the regulations into the company's data- and process rules	Ensure compliance and accountability of GDPR
Ensure that data is managed and used as an asset	
Ensure data security and data privacy measures	
Make sure data is accessible, structured, and well documented	
Align employees in the processes of managing data	
Account for national requirements related to the location of data	
Store data in your respective region to avoid latency and poor responsiveness	

Establish a data security strategy	
Hire and train employees with IT expertise	Hire people with sufficient IT, AI, cloud computing, and domain expertise
Hire and train cloud computing specialists	
Focus on data quality	
Develop an autonomous cloud environment	
Develop the cloud environment based on specific business needs	
Understand the process of how data is created to how it can be used for business purposes	
Pilot with cloud vendor	
Do not underestimate the work associated with AI services	
Start simple and gradually build complexity	Manage integration and compatibility of on-premises infrastructure and the cloud environment
Move the enterprise IT infrastructure to the cloud for easy data access	
Take advantage of the rapid innovations in AI tools and recourses provided by cloud vendors and third-parties	
Separate the cloud environment from existing, on-premises infrastructure for rapid time to value	
Ensure that the cloud environment is not entirely separated from the existing infrastructure to avoid data silos and maintenance costs	
Take advantage of cloud vendor support	
Build the cloud infrastructure explicitly towards the company's datasets, data structures, and ETL processes	
Develop a data platform for cquisition, storage, preparation, delivery, and governance of your data	

Affinity group: Technical factors	
Activity statements	Supporting themes
AI services should automate the development of AI solutions	Streamline the AI development cycles
Invest in pipeline management	
Pipelines must be reusable and streamlined	
Standardize pipelines	
Minimize human input with automation	
Invest in manual or automatic pipeline monitoring	
Test the ML model frequently	
Automate processes that reveal information on how an ML model behaves	
Write code and logic so that an ML model can run itself	
Describe and monitor the data flow	
Add discipline to the development and deployment of ML models	
Streamline the ML lifecycle	
Describe and monitor the data flow	
Add components, data, and functionality to the solutions without impacting the internal structure or data flow	Manage enterprise AI systems
Adopt AI incrementally	
Prevent the creation of siloed use-case-based AI solutions	
Start with a small business problem that can be solved with the available data	
Focus on simplicity rather than starting with all the fancy tools the cloud has to offer	
Integrate an automatic data flow to and from the cloud environment	
Business specifications need to be translated and built into the ML model	
Build agile architectures using best-of-breed components	
Use different technologies for different components	
Analyze and structure the business problem before designing algorithms	

Increase AI specialists' industry domain knowledge	Build use-case-driven systems that address unique business requirements and needs
Achieve resiliency in the developed AI solutions	
Hire and train AI- and statistical experts	

Affinity group: Technology commercialization process	
Activity statements	Supporting themes
Extensive experimentation is necessary to enable the best performing ML models	Balance flexibility, experimentation, and accountability
Experimentation must be managed and monitored	
Investigate needs, develop and test ideas	
Avoid resource intensive and solely experimental projects	
Control and monitor ML models and cloud environment	
Experiment with a substantial amount of pilot projects	
Discontinue projects that prove no value	
Base the IT architecture on the company's objectives	Align people, processes, IT investments
Work thoroughly with an idea and test it with a customer	
Constantly involve customers and business developers during the creation of the solutions	
The creation phase must be designated sufficient time to avoid creating a large maintenance machinery on top of existing infrastructure	
Align the solution with strategic and tactical goals	
Tracking the model performance to ensure consistency and satisfaction of business requirements and customer needs	
Establish internal coordination to ensure that the data reflects the business processes	
People, processes, and IT must adhere to the overall business strategy	
The AI solution must be explainable and reliable	Preserve customer and employee trust in the AI solutions
GDPR legislation requires organizations to be able to explain algorithmic decisions if they impact the customer	
Utilize tools and frameworks with capabilities for explaining, validating, monitoring, and mitigating bias in AI models	
Invest in ongoing maintenance and monitoring of the deployed solutions	
Avoid black box AI solutions	
Continuously improve accuracy of an ML model	
Understand the AI decision-making processes	
Determine the applicable metrics to monitor an ML model	

Affinity grouping of supporting themes

Supporting themes	Sub-affinity group
Ensure low degree of formalization and standardization	Organizational structure
Create a shared understanding of technical concepts and a common language for innovation	
Decentralize decision-making	
Enable teams to work autonomously	
Create a shared understanding of technical concepts and a common language for innovation	Culture and inter-organizational communication
Enable teams to work autonomously	
Decentralize decision-making	
Ensure low degree of formalization and standardization	Agile experimentation
Balance flexibility, experimentation, and accountability	
Decentralize decision-making	

Ensure compliance and accountability of GDPR	Data management
Hire people with sufficient IT, AI, cloud computing, and domain expertise	
Manage integration and compatibility of on-premises infrastructure and the cloud environment	Infrastructure management
Hire people with sufficient IT, AI, cloud computing, and domain expertise	
Build use-case-driven systems that address unique business requirements and needs	Customizability
Hire people with sufficient IT, AI, cloud computing, and domain expertise	
Streamline the AI development cycles	Automation
Manage integration and compatibility of on-premises infrastructure and the cloud environment	
Manage enterprise AI systems	Scalability
Manage integration and compatibility of on-premises infrastructure and the cloud environment	
Balance flexibility, experimentation, and accountability	Explore/Experiment
Enable teams to work autonomously	
Ensure low degree of formalization and standardization	
Align people, processes, IT investments	Create
Preserve customer and employee trust in the AI solutions	
Build use-case-driven systems that address unique business requirements and needs	
Manage enterprise AI systems	Implement
Preserve customer and employee trust in the AI solutions	

Appendix E - Process for performing affinity grouping of activity statements (adopted from Caralli 2004)

1. Mark the origin of each activity statement.

Later, it may be important to be able to trace back to the interview or document from which the activity statement was created. Thus, it is a good idea to tag each activity statement for origin. (One suggestion is to apply line numbers to interview notes and to use the line number and the participant's initials to identify the origin of the statement, such as "RAC22.") However, be careful not to make this origin tag so prominent a part of the activity statement that it becomes a primary criterion for categorizing activity statements. The core content, intention, or meaning should always be the primary driver for categorizing the activity statements.

2. Use only the activity statements when creating affinity groupings.

The activity statements should be the only initial source for this activity. If an activity statement is difficult to categorize with other statements, it may need further clarification. This can be accomplished by referring to the raw data from which the activity statement was created, if necessary. Be careful, however, not to derail the grouping process by delving too deeply into the source data. The essence of the affinity grouping activity is to make immediately recognizable connections between somewhat disparate pieces of data.

3. Work each activity statement individually.

The essence of the affinity grouping activity is to make immediately recognizable connections between similar data elements by using the core content, intention, or meaning portrayed by the activity statement as the primary decision criteria. Each activity statement should be considered individually and placed into an affinity group. It is sometimes helpful to have several people consider each statement together and agree on placement in a group. If a statement cannot be grouped after reasonable consideration, it may appropriately indicate the creation of a new group. If not, place the statement aside and continue with the affinity grouping process.

4. Stabilize the affinity groups.

As a final check, once all activity statements have been grouped, each group should be examined to determine if subgroups are emerging and should be extracted. For example, during an affinity grouping activity (particularly when considering a large number of ac-tivity statements) the team performing the grouping may lose sight of the meaning of the groups it has created or may inadvertently change the definition of a group as the process unfolds. Eventually, this can result in the creation of groups that, in actuality, contain more than one distinct group. If this is the case, additional distinct groups should be separated. (One caution: looking for additional groups within a group is not the same as defining emerging themes that underlie or support all of the activity statements in a group. This activity is referred to as "developing supporting themes," and is the final ab-straction required to create CSFs.) In addition, this is a good time to consider duplicate activity statements. Duplicate statements (particularly if they are from different participants) can serve to confirm a par-ticular affinity group, and later a CSF. However, duplicates can be eliminated if neces-sary. Also, consideration should be given to the traceability of the activity statements at this point. Moving forward, the origin of the statement should be removed (certainly be-fore presentation to the organization), but it is a good idea to keep a copy of the groups with origin information in case further analysis is required later in the process.

5. Address any left over activity statements or small groups of statements.

After the affinity grouping exercise, a set of activity statements may remain that cannot be placed into a group. Each statement that doesn't fit into a group must be re-examined. In some cases, several remaining statements will form a new group; however, there is a possibility that some will never fit into a group. For those that cannot be grouped, a decision must be made as to their value. Keep in mind that a single activity statement may be so compelling that it eventually defines its own CSF; conversely, an activity statement might be found to be extraneous, and a decision might be made to discard it.