UNIVERSITY OF COPENHAGEN
Department of Computer Science

PhD Thesis

Jon Eklöf

AI transformation in the Manufacturing Industry

Supervisor: Thomas Hamelryck

Name of department:	Department of Computer Science
Author:	Jon Eklöf
Title and subtitle:	AI transformation in the manufacturing industry
Topic description:	This thesis investigates, from a multidisciplinary point of view, what is required from a manufacturing company leader support widespread introduction of AI.
Supervisor:	Thomas Hamelryck
Submitted on:	18 October 2023
Grade:	
Number of characters:	128 859 (excluding appendices)

The cover artwork “Re-birth, learning by creating” is created by the artist Jonas Lundberg who uses generative AI models to create art. In this case, he has leveraged an AI model called Midjourney (Midjourney 2023) (created by the research lab with the same name) that generates art from text.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

If there is one thing that I have discovered throughout this process that is that certain issues can only be solved together with others. Therefore, I want to express my sincere gratitude to all who have made this research possible. Your support has been invaluable to me. In particular, I would like to mention:

My principal supervisor Thomas Hamelryck, for sharing your deep knowledge, your perseverance, your unwavering committment and your courage to embark on this journey together with me.

Professor Ulrika Lundh Snis, for your invaluable insights, your enthusiasm, clear guidance and calming words.

Co-authors Alexander Grima, Cadell Last, and Ola Rønning who have contributed to this work. Thank you for your time, effort, and expertise. Without you, none of this would have materialized.

My colleagues at GKN Aerospace Amanda Dalstam, and Joakim Andersson for your respective invaluable contribution and for believing in me.

My wife Sara, my children Isak and Eline and my dear friend Jonas, and the rest of my closest family and friends. Your unwavering love, support, and encouragement have been invaluable to me throughout my journey, from celebrating my successes to standing by me during the hardships. I am truly fortunate to have each and every one of you in my life.

ABSTRACT

The manufacturing industry is becoming increasingly complex, dynamic, and connected. As a result, companies are facing challenges in managing highly nonlinear and stochastic activities due to the many uncertainties and interdependencies they face. In recent years, the development of artificial intelligence (AI) has shown potential for transforming the manufacturing domain through the use of advanced analytics tools for processing large amounts of manufacturing data. There are many examples of research investigating how AI can be used to optimize performance in manufacturing companies. However, studies show that only a small percentage of firms across industries engage in widespread adoption of AI. Most companies only run ad hoc pilots or apply AI to a single business process. Some researchers claim that AI has the potential to disrupt the manufacturing industry as we know it, arguing that companies that only partially commit to an AI transformation will be outcompeted by those who can offer new data-driven services. Others, however, question the current direction and potential of AI.

This thesis investigates, from a multidisciplinary point of view, what is required from a manufacturing company leader to efficiently support a widespread introduction of AI. It combines quantitative and qualitative methods combined with real-life examples of introducing AI in an aerospace manufacturing company. While the term AI is commonly used in various contexts, there is some confusion surrounding the concept, and a vast spectrum of definitions have emerged. Therefore, the thesis begins with describing essential concepts within the field of AI as well as their history.

Leaders of manufacturing companies play an important role in achieving widespread implementation. However, there is limited research on how leaders best can contribute to widespread implementation. The first step of this research was therefore to create a capability framework for leaders of manufacturing companies that wish to introduce AI on a wide scale. In this work, we identified a willingness to learn about AI as one of the most important capabilities leaders could benefit from when supporting the widespread adoption of AI. However, this is challenging. The field grows as technologies emerge that could fit under the AI umbrella. Additionally, researchers and practitioners lack a coherent definition of AI, which has led to a mystification of the term.

To gain an understanding of the leadership capabilities required for an effective AI implementation, it is necessary to consider the nature of AI technology itself and its use. Technological advancements within AI have enabled non-experts to develop AI applications that previously required in-depth computer science, statistics, and mathematics knowledge. To develop the understanding of these technological advancements, we conducted a second study that specifically examined the role of abstraction in AI. Our aim was to quantify the level of abstraction in deep learning by investigating the number of lines of code utilized in deep learning projects. We saw a dramatic decrease in the number of lines of code used, indicating an increase in abstraction. Thereafter, we proceeded to investigate implications of this abstraction increase, particularly with respect to mimesis. While these developments contribute to the democratization of AI, promoting collaboration and allowing us to do more with less, they also come with several drawbacks such as convergence on suboptimal solutions, quality assurance problems and scarcity of talent that most organizations are not yet strategically prepared to handle.

Much of the research on AI in manufacturing has been conducted outside the daily operations of manufacturing companies. Further research has been requested on introducing it into daily operations. Against this background, in order to investigate implications, considerations, and trade-offs that need to be made when introducing AI into daily operations I conducted an action research study using a case study of a global manufacturing company deploying AI to develop capabilities and enhance decision-making.

This research offers a multidisciplinary investigation of some of the challenges and considerations involved in implementing AI in the manufacturing industry and highlights the role of manufacturing company leaders in facilitating the adoption of AI in the industry. While AI can be a useful tool for solving problems, it may not always be the best solution, and companies should carefully consider where and how to use it. The research clearly shows the importance of multidisciplinary approach, collaboration, as well as the combination of competencies to succeed in AI transformation.

1. INTRODUCTION

The focus of this thesis is on AI transformation in the manufacturing industry, and the associated leadership challenges and implications. In this thesis, I frequently use the term "leader" to refer to individuals who possess the ability or have been appointed to inspire, influence, build trust, and create a vision for the future within their organization. This definition is intentionally broad, encompassing individuals at all levels and areas of responsibility within the company, whether formal or informal. I have chosen to use a term that is inclusive, rather than limiting itself to specific levels or areas of responsibility within the organization (Saran 2018).

In this first chapter, I outline the content of the thesis.
Chapter 2 provides a technological framework for the thesis. It covers the concept of AI, its history, and introduces essential AI-related concepts that the reader will encounter throughout the thesis and in my research studies. The chapter contextualizes subsequent chapters and serves as a foundation for an in-depth exploration of the research topic.
Chapter 3 is focused on AI in the manufacturing industry. It covers the relationship between AI and manufacturing, potential areas for AI adoption and the limited adoption of AI and barriers to implementing AI. The chapter also discusses the leadership implications of these barriers and sets out the research questions aimed at addressing the issues and supporting the implementation of AI in the manufacturing industry.
Chapter 4 sets out the overall research design. In this chapter, I discuss practice-based and action research and elaborate on the role of the researcher as the reflective practitioner. I conclude the chapter with a short description of the manufacturing context through which I have studied the subject, namely aerospace component manufacturing, and introduce the company that has funded the research.
Chapter 5 describes the results and contributions of the thesis and outlines the three papers generated as a result of this research.
In Chapter 6, I conclude the thesis by setting out my conclusions throughout my research and providing suggestions for further research.

2. TECHNOLOGICAL FRAMEWORK

This chapter provides an overview of the technological framework that forms the basis of my research on the leadership aspects of introducing AI into the manufacturing industry. This framework serves as the technological lens through which I have studied the research topic and contains descriptions of essential concepts within the field of AI that the reader will encounter throughout the thesis and in my research studies.

Through this chapter, readers can understand the context and significance of the research and the technological foundations that inform the subsequent chapters. In my studies, it has been important to have an understanding of AI and some fundamental concepts related to it to understand both the research problem domain of AI transformation in the manufacturing industry on a more general level and the technologies that I have investigated and applied in my research studies in particular.

I will begin by discussing the concept of AI and its history where I also address the potential confusion surrounding the definition of AI. Furthermore, I introduce certain fundamental concepts related to AI. By providing an understanding of the technological framework of the research, this chapter aims to contextualize the subsequent chapters and lay the foundation for an in-depth exploration of the leadership aspects of introducing AI into the manufacturing industry.

2.1. Artificial Intelligence

Even though the term AI is frequently used, there is yet to be a generally accepted definition (Monett, Lewis and Thórisson 2020). Researchers such as Johnson and Verdiccihio that have written about the way AI is discussed and presented posit that there is a confusion around what AI is (Johnson and Verdicchio 2017). Since its first use by the computer and cognitive scientist John McCarthy, in 1955, when he defined it as the science and engineering of intelligent machines that mimic the cognitive functions that we associate with the human mind, i.e., the ability to sense, reason, act, learn and adapt (McCarthy 2007), a vast spectrum of definitions have emerged.

Within AI research, over 28 definitions have developed in the last decade, and subsequent attempts to systematize the approaches are still being discussed (Monett, Lewis and Thórisson 2020). Many subdomains are developing within AI, and the concept can be classified at various levels of generality. Moreover, its meaning will probably change as available technologies continue to evolve (Pandl, et al. 2020, Fortuna and Gorbaniuk 2022). This confusion about AI also appears in research areas, such as general management. Here, researchers such as Davenport, Govindarajan, Weill, Bassellier and Persaud speak of the subject without defining what they mean by it (Davenport and Foutty 2018, Govindarajan and Immelt 2019, Bassellier, Benbasat and Reich 2003, Persaud 2021).

Below I present the most commonly known definitions of AI apart from McCarthy's definition described above. For a more comprehensive discussion about the definition of AI, please see (Russell and Norvig 2009, Legg and Hutter 2007).

Alan Turing, a British mathematician and computer scientist who is famous for his work within theoretical computer science and AI, presented a famous AI definition in his paper Computing Machinery and Intelligence (Turing 1950). Turing, however, did not refer to it as "artificial intelligence" but as "computing machinery and intelligence". Turing based his AI definition on a test he called the "imitation game". Based on this test, "Artificial intelligence" means any computer that passes the "Turing test". The Turing test is a game played with a human, a computer, and a human judge. The human judge is separated from the other two participants, who can only communicate via text. The Turing test is passed if the human judge cannot effectively discriminate between the human and the computer.

Another definition common among AI researchers, such as Stuart Russell and Peter Norvig, is that of AI as intelligent agents. For example, Russell and Norvig use the following definition: "Artificial intelligence" means an intelligent agent. "Agent" means a software system that perceives its environment through sensors and acts upon that environment through actuators. "Intelligence" means the ability to select an action that is expected to maximize a performance measure (Russell and Norvig 2009). This definition of AI as intelligent agents has also been used by, for example, Yann LeCun in the later years when discussing how machines could learn as efficiently as humans and animals (LeCun 2022). This definition of AI as an agent, where algorithms are used that can select an action that is expected to maximize a performance measure, is also what I commonly refer to in this thesis.

2.2. The history of AI - from myth to probabilistic deep learning

After introducing the concept of AI, I will provide a brief history of the idea of intelligent machines, tracing back the ancient myths of crafted intelligent beings, via the emergence of AI research in the mid-1900s until the advent of probabilistic deep learning in more recent years. I believe this historical account is helpful in contextualizing the evolution of the field of AI, including the challenges and setbacks faced along the way.

2.2.1. Antique myths of crafted intelligent beings

The idea of intelligent machines began already in ancient times, with myths of crafted intelligent beings. In Greek mythology, for example, around 700 B.C., stories spoke of Talos, a giant constructed of bronze who guarded the island of Crete (Hunter 2015). Throughout history, artisans have created realistic humanoid automata designed to give observers the impression that they are operating under their own power. One famous example of such automata is the late 18th century chess-playing “mechanical Turk” (a fraud with a human chess master hiding inside to operate the machine) (Schaffer 1999, McCorduck 2004).

Fig 1: Image of the mechanical Turk from Joseph Friedrich Freiherr von Racknitz’s book that tried to explain the illusions behind the chess-playing automaton (Humboldt-Universität 2022).

2.2.2. The emergence of AI research

In the mid-1900s, neurology research showed that the brain was an electrical network of neurons that fired in all-or-nothing pulses. Accordingly, in the 1940s and 50s, scientists from mathematics, psychology, engineering, economics, and political science began to discuss creating an artificial brain (Russell and Norvig 2009). A fundamental step towards creation of AI was taken when Alan Turing discussed that humans use available information and reason to solve problems and make decisions and that it should be possible to build intelligent machines (Turing 1950).

However, it was not until 1956, when computer scientist John McCarthy hosted a conference at Dartmouth College that the term AI was coined. It is also considered that it was at this conference that the research field of AI was born. During the conference the participants posited, "*every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it*" (Crevier 1993, Kaplan 2022).

2.2.3. Symbolic and connectionist AI

In the 1950s, the field of AI began to take shape as two distinct visions for achieving machine intelligence emerged. The first of these visions, known as symbolic AI, proposed using computers to create a symbolic representation of the world and systems that could reason about it. This approach was championed by notable figures such as Allen Newell, Herbert A. Simon, and Marvin Minsky. The second vision, known as the connectionist approach, aimed to achieve intelligence through learning (Manyika 2022). This approach emphasized the use of neural networks and machine learning algorithms to enable machines to learn from experience.

A famous proponent of the connectionist approach was Frank Rosenblatt. Sparking off the ideas of an artificial brain, in 1957, Rosenblatt introduced the perceptron - one of the most important building blocks for what later became artificial neural networks (ANNs). ANNs are algorithms loosely inspired by the structure of biological neural networks such as the human brain. Rosenblatt's perceptron was a single-layer ANN model that contained an input value, a weight and bias, a net sum, and an activation function. It weighed inputs and applied an activation function that resulted in an output. Rosenblatt predicted that the *"perceptron may eventually be able to learn, make decisions, and translate languages"*.

Subsequent research exposed limitations to what perceptrons could do and that Rosenblatt's predictions had been exaggerated (Russell and Norvig 2009). However, this did not mean that the idea of the perceptron went extinct. Today, the perceptron serves as the foundation of ANNs – which essentially are multi-layer perceptrons.

Both Symbolic AI and connectionist approaches had their own strengths and weaknesses, and the debate between the two approaches continues to shape the field of AI today. Symbolic AI was criticized for its inability to handle the complexity and uncertainty of real-world problems, while the connectionist approach was criticized for its lack of transparency and interpretability.

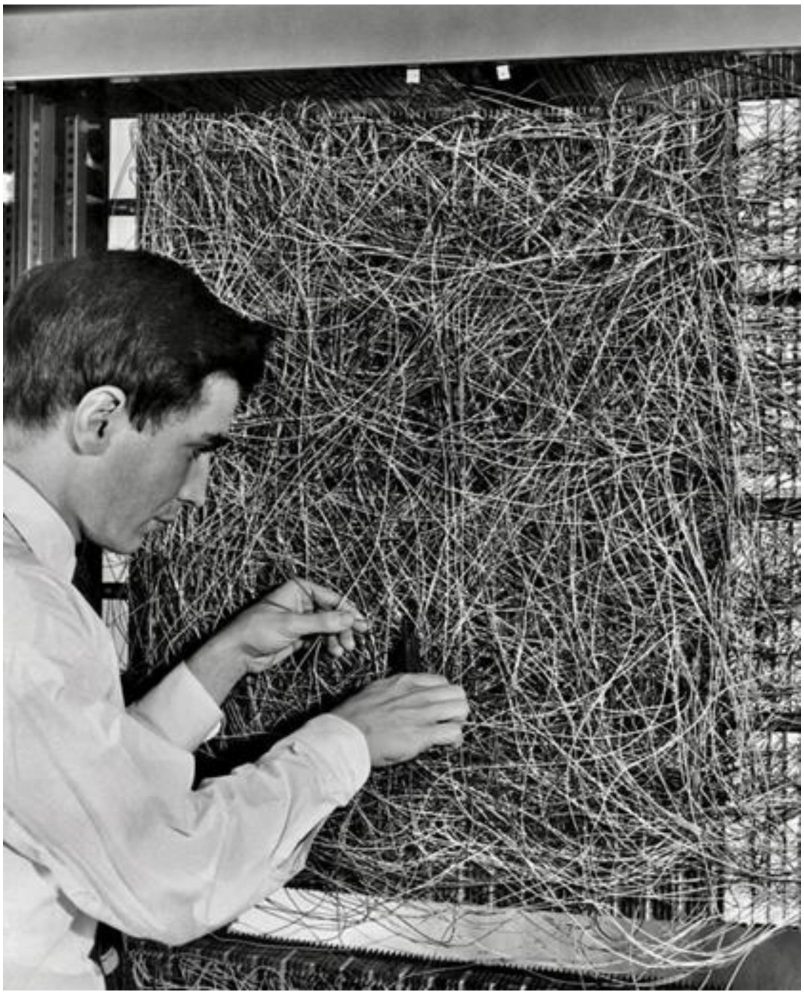

Fig 2: Frank Rosenblatt in 1960 with a Mark I Perceptron (Santa Barbara Museum of Art 2022).

2.2.4. The first AI winter

In the 1960s and 1970s, researchers within the field of AI were convinced that symbolic approaches would ultimately lead to the creation of a machine with artificial general intelligence (Newquist 2020). Some experts, such as Herbert Simon, predicted that machines would be capable of performing any task that a human could do within the next 20 years. Similarly, Minsky believed that the problem of creating AI would be substantially solved within a generation. However, these researchers failed to fully grasp the complexity and difficulty of the remaining tasks (Crevier 1993). While AI research received much support and funding during the 1950s and 1960s, in the 1970s it became clear that researchers had underestimated the challenges associated with creating AI, and the field experienced major critiques and financial setbacks (Russell and Norvig 2009). In 1974, AI research was defunded by the U.S. and British Governments, leading to a difficult period for AI research known as an "AI winter", during which obtaining funding for AI projects became increasingly difficult (Russell and Norvig 2009).

2.2.5. Expert systems

This “AI winter” continued until the 1980s when a new form of AI technology referred to as “expert systems” emerged and became the new focus of AI research (Newquist 2020). There are several types of expert systems, including rule-based (that uses rules as the knowledge representation for knowledge coded into the system) (Grosan and Abraham 2011) and frame-based (that processes problem-specific information in the working memory, with a set of frames contained in the knowledge base) (Rattanaprateep and Chittayasothorn 2006). Expert systems rely on deductive inference, restrict themselves to small domains of specific knowledge and use logical rules derived from the knowledge of experts to solve problems within these specific domains. The expert systems proved for the first time that AI could be useful. However, expert systems proved expensive and difficult to administer (Russell and Norvig 2009, Cowan 2001). Each small domain required its own expert system, and the systems required manual updates. After these limitations became apparent in the 1990s and early 2000, the research field of AI faced additional setbacks.

2.2.6. Emergence of deep learning

The next fundamental shift within the field of AI occurred around 2010 when deep learning models emerged. While in 2010, ANNs were nothing new (Russell and Norvig 2009), the access to large amounts of data and cheaper and faster computers combined with for example groundbreaking research by Alex Krizhevsky et al. in creating AlexNet (Krizhevsky, Sutskever and Hinton 2012), allowed researchers to develop powerful AI models for image and video processing, text analysis, and speech recognition (LeCun, Bengio and Hinton, Deep learning 2015). In many aspects, the advent of deep learning transformed how we process vision and speech data. Given enough data, deep learning systems proved powerful enough to identify patterns between a given set of inputs and a set of corresponding outputs (Marcus 2018).

However, even though many AI applications today use deep learning, soon after its emergence, its limitations started to show. François Chollet (author of the deep learning software framework Keras and one of the main contributors to Google's deep learning software framework TensorFlow) said, already at the end of 2017 (only five years after Krizhevsky et al. published their major paper), that:

"For most problems where deep learning has enabled transformationally better solutions (vision, speech), we've entered diminishing returns territory in 2016-2017." (Chollet 2017).

2.2.7. The advent of probabilistic deep learning

Deep learning works best when there are many, many training examples. However, it risks falling short when training examples are few or very complex. In 2014, Kingma and Welling proposed a way of dealing with these limitations of regular deep learning by combining traditional statistical methods with deep learning – so-called probabilistic deep learning (Kingma and Welling 2014). This approach has been gaining popularity in recent years as a way to improve the performance and robustness of deep learning models (Tran 2020).

The main advantage of probabilistic deep learning is that it accounts for uncertainty, both model uncertainty and data uncertainty (Tran 2020, Chang 2021).

2.3. Fundamental concepts of AI

In this section, I provide an introduction to several key concepts within the field of AI. While this section is not intended to be an exhaustive list of all concepts within the field, it is essential to understand these common concepts to follow the discussions and analyses presented throughout this thesis and my research studies. Please note that this section serves as an introductory guide and is intended to provide a basic understanding of the key concepts. Further reading may be necessary for a more in-depth understanding of these concepts and their practical applications.

2.3.1. Machine learning and deep learning

As mentioned above, what I often refer to in this research is the definition of AI as an intelligent agent, where algorithms are used that can select an action that is expected to maximize a performance measure. The technology that is often used to achieve this is referred to as machine learning. Machine learning has been defined in various ways, and is often divided into two categories: traditional machine learning and deep learning, although this division is sometimes debated (Yang, et al. 2018, Pin, En and Peng 2021). Traditional machine learning relies on established statistical methods like linear regression and support vector machines, while deep learning uses ANNs with multiple layers to make predictions (Marcus 2018). It is important to note that all deep learning methods, including probabilistic deep learning methods, fall under the umbrella of machine learning, but not vice versa.

2.3.2. Supervised, unsupervised & semi-supervised learning

A common division within machine learning, apart from that between traditional machine learning and deep learning, is between supervised, unsupervised, and semi-supervised learning. Supervised learning (or supervised models) makes use of labeled datasets to train or “supervise” the algorithms to predict outcomes accurately. Labeled inputs and outputs allow the model to measure its accuracy and improve. Supervised learning, rely on techniques such as logistic regression and decision trees and is commonly used for tasks such as email spam filtering (Dada, et al. 2019, Sivakumar, et al. 2018).
Unsupervised learning models use techniques such as k-means to analyze unlabeled data sets. These algorithms detect hidden patterns in data without the need for human

intervention (hence, they are “unsupervised”) and are used in areas such as image segmentation and handwriting recognition (Ahmed, Seraj and Islam 2020).
Semi-supervised learning models use training datasets with both labeled and unlabeled data to solve tasks such as text and image classification (Ouali, Hudelot and Tami 2020). It relies on techniques such as generative models (such as the semi-supervised variational autoencoder) or semi supervised support vector machines (Hady and Schwenker 2013). Semi-supervised learning is interesting because it can use unlabeled data to improve supervised learning tasks when there are low amounts of labeled data (Xiaojin and Andrew 2009).

2.3.3. Discriminative vs. generative models

Yet another major division in machine learning is between generative and discriminative modeling. In discriminative modeling, the aim is to learn a predictor based on observed data, such as classifying emails as spam or not spam using previous training examples. In contrast, generative modeling aims to learn a joint probability distribution over all variables, allowing for the generation of new data (Kingma and Welling 2014). This capability makes generative modeling useful in a range of applications, including image and language generation, as well as protein design.

2.3.4. Deduction, induction, abduction

In the context of AI, inference refers to the process of using a model or system to make predictions or decisions based on input data. In other words, the process of drawing conclusions from observed data (Russell and Norvig 2009). AI models typically rely on two types of inference. Deductive inference and inductive inference.
Deductive inference is a type of reasoning where if the premises (A) are true, then the conclusion (B) is necessarily true (Johnson-Laird 2010). A classic example of deductive inference is the syllogism: "If all men are mortal, and Socrates is a man, then Socrates is mortal." In computer science, deductive reasoning is often used in automated theorem proving, where a computer program derives logical consequences from a set of given axioms or premises. The expert systems of the 1980s often relied on deductive inference. Today, most machine learning algorithms are inductive inference engines, i.e. the model is able to make predictions based on patterns observed in data. Inductive inference is based

on statistical data and observed frequencies and is a type of reasoning where if the premises (A) are true, then the conclusion (B) is likely to be true (Angluin and Smith 1983). For example: if (A) most dogs are friendly, then knowing that an animal is friendly, (B) increases the probability that the animal is a dog.

There is also a third type of inference called abductive inference. Abductive inference is a type of reasoning where if the premises (A) are true, then the best explanation for the conclusion (B) is inferred from the available evidence (Kettner 1991). Humans often rely on abductive inference, but then refer to it as “common sense”. In machine learning, abductive inference can be seen as the process of constructing a range of models worthy of consideration given the data.

The concepts that I have explored above refer to AI on a more general level. Throughout my research, I have focused much on deep learning and probabilistic deep learning. Therefore, I will spend some time explaining important concepts within these specific areas.

2.4. Artificial Neural Networks and Deep Learning

Loosely inspired by the structure of biological neural networks, ANNs are computational processing systems consisting of interconnected computational nodes that work collectively to learn from the input to optimize its final output. Deep learning networks are simply ANNs with many hidden layers stacked upon each other (Goodfellow, Bengio and Courville 2016).

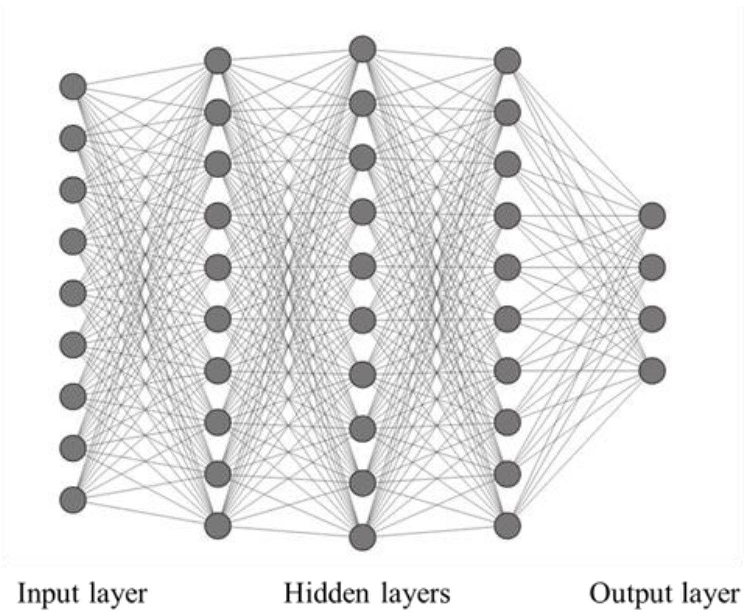

Fig 3: Illustration of an ANN showing the relationship between the layers and how the network nodes are interconnected.

A multidimensional vector is loaded into the input layer that transforms it into the hidden layers. Each node in the hidden layers has its own weights, bias, and activation function. For each node, the input values $\boldsymbol{x}$ are multiplied by their corresponding weight $\boldsymbol{W}$, added together with a bias term $\boldsymbol{b}$, and an activation function is applied to provide an output.

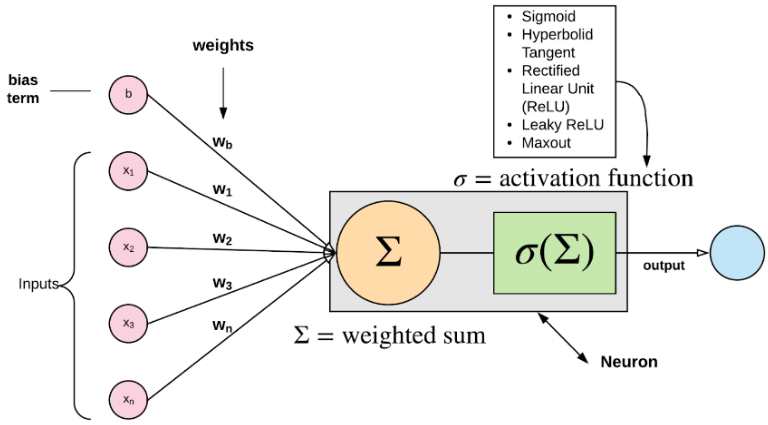

Fig 4: Illustration of how each node of an ANN has its inputs, weights, and activation function. The node takes an input ***x*** and scales it by the weights ***W*** and biases ***b***. Subsequently, a nonlinear activation function , such as Sigmoid, ReLU etc., is applied, which renders an output. Image taken from (Bisong 2023).

2.4.1. Activation Functions

ANNs use activation functions within the hidden layers to ensure a nonlinear transformation of the input vector and in the output layer to ensure that the neural network provides the desired output. The activation function provides for if a neuron should be activated and transforms the summed weighted input at a node into an output value to be fed to the next hidden layer or as final output (Nwankpa, et al. 2018). There are different activation functions with different areas of use – often divided based on their ranges or shapes of their curves. For example, the curves of the activation functions Sigmoid and Tanh, typically used for binary classification, both have an S-shape. The Sigmoid function outputs a value between 0 and 1 while the Tanh ranges from -1 to 1. Since the range of Tanh is between -1 to 1 the mean will be close to 0 in the Tanh graph (Nwankpa, et al. 2018).

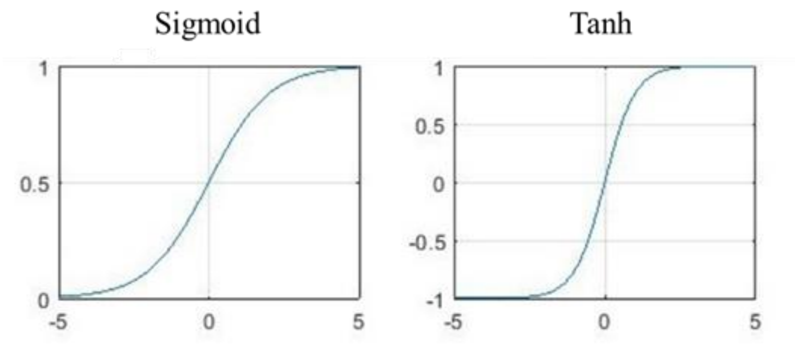

Fig 5: Graphs of Sigmoid and Tanh activation functions compared, image taken from (Nwankpa, et al. 2018).

The Softmax function another type of sigmoid function, often used for classification of more than two classes (Nwankpa, et al. 2018). ReLU is one of the most commonly used activation functions within deep learning i.a. since it is less computationally expensive than Sigmoid and Tanh because it involves simpler mathematical operations that enable better training of deeper networks. As of 2017, was is the most popular activation function for deep neural networks (Glorot, Bordes and Bengio. 2011, Nwankpa, et al. 2018, Ramachandran, Zoph and Le 2017). ReLU has a range of 0 to infinity. It gives an output of x if x is positive and otherwise 0.

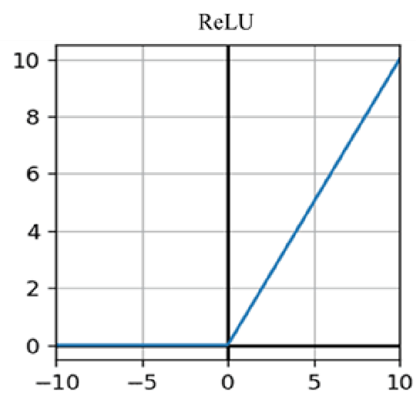

Fig 6: The graph of a ReLU activation function.

2.4.2. Loss function and gradient descent

The hidden layers of ANNs perform nonlinear transformations of the inputs entered into the network from the previous layer. Once the model has produced an output, it compares the output against the given target output using a loss function. A loss function (sometimes also referred to as a cost function) is a function that compares the target and predicted output values and aims to minimize loss between the predicted and target outputs - usually using gradient descent. Gradient descent allows for the optimization of weights and biases to minimize the average loss. This adjustment process is referred to as “learning”.

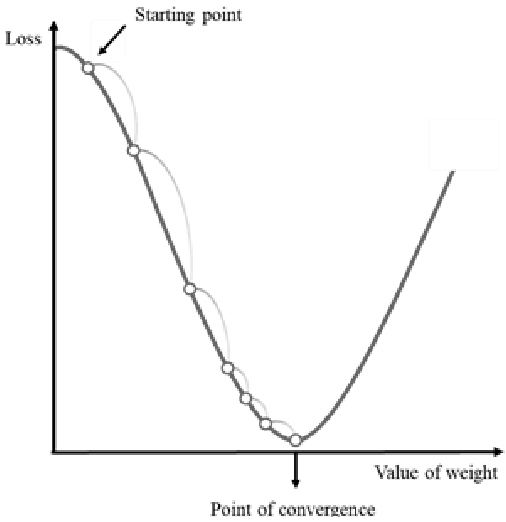

Fig 7: Illustration of gradient descent. The gradient descent starts at an arbitrary point to evaluate the performance. From there, the algorithm will gradually inform the update of the weights and bias until the slope reaches the lowest point on the curve, known as the point of convergence. (Image taken from (JavaPoint 2023).

The aim of using gradient descent is to find the parameters of the ANN that incurs the lowest loss. Automatic differentiation is used to get the gradient, and the gradient descent starts at an arbitrary point to evaluate the loss. From there, the algorithm will gradually (in what is known as learning steps) find the slope and can thereafter use a tangent line to observe the steepness of the slope. The slope will inform the update of the weights and bias until the slope gradually reaches the lowest point on the curve, known as the point of convergence (Bottou 2010).

2.4.3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a type of ANN that was first introduced in 1998 in a paper by LeCun, Bottou, Bengio and Haffner (LeCun, Bottou, et al. 1998). CNNs are, due to their ability to assign importance to spatial relationships within the data, primarily used in the field of pattern recognition within images (O'Shea and Nash 2015). This ability allows encoding image-specific features into the architecture, making the network more suited for image-focused tasks while reducing the parameters required to set up the model.

Apart from the input layer, CNNs usually have three types of layers. These are:

(i) Convolutional layers, where kernels move across the input image to detect local features at different positions. To produce the feature map, the kernel convolves with the input image by computing the dot product between the input image and the kernel (O'Shea and Nash 2015).

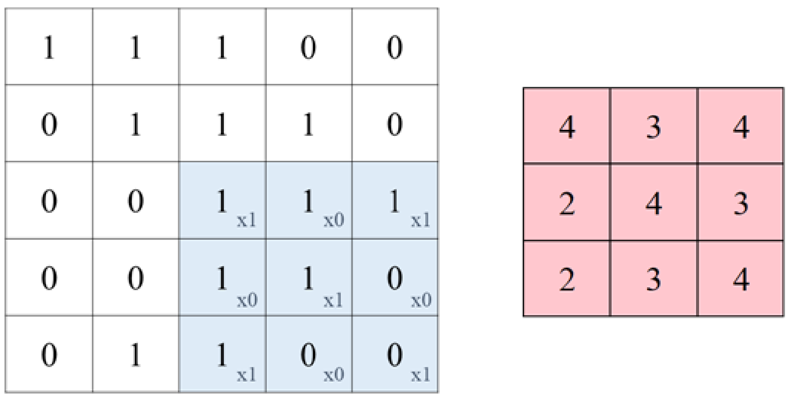

Fig 8: An illustration of a convolutional layer where the white section represents a 5x5x1 pixel input image. The kernel (a 3x3x1 matrix represented in the color blue) moves across the input image and convolves with the input image by computing the dot product between the input image and kernel to produce the feature map (the 3x3x1 red matrix).

(ii) Pooling layers, that reduce the dimensions by combining each group of the outputs of the convolutional layer(s) into a single neuron in the next layer - thereby reducing the computational complexity of the next set of layers. There are two common variations of pooling operations: average pooling and max pooling. An average pooling layer averages its input values by taking their mean, while max pooling takes the biggest value.

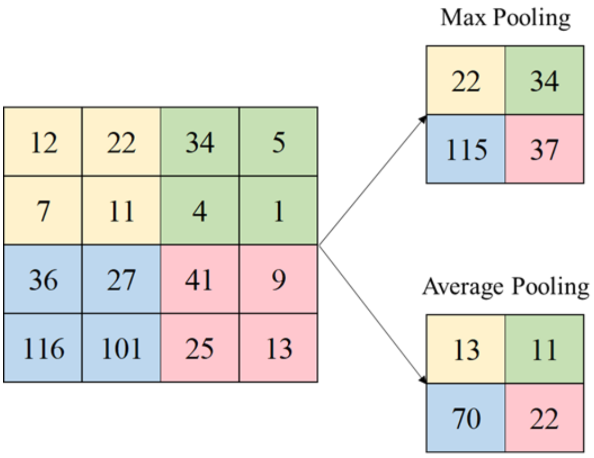

Fig 9: An illustration of max pooling and average pooling.

(iii) Fully connected layers, that attempt to produce class scores from the activations to be used for classification.

A CNN architecture is formed when input, convolutional, pooling, and fully connected layers are stacked (O'Shea and Nash 2015).

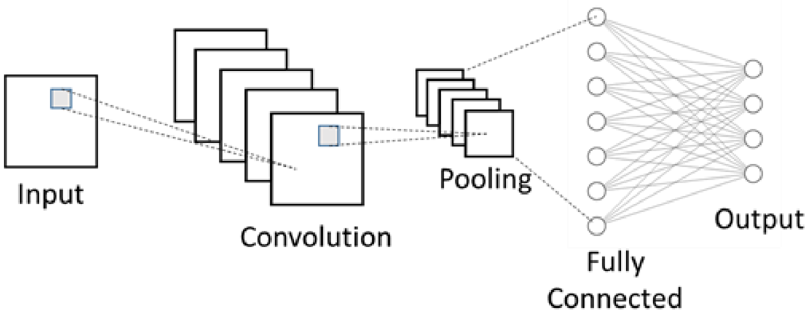

Fig 10: Illustration of the constituents of a CNN with input, convolution, pooling, fully connected and output layers.

CNNs improve the ability to compute complex datasets by identifying and reducing the number of relevant features. However, one of the most significant limitations of traditional

ANNs is that they tend to struggle with the computational complexity required to compute image data.

Machine learning benchmarking datasets such as the MNIST are suitable for most forms of ANNs due to its relatively small image dimensionality. With MNIST, a single neuron in the first hidden layer will contain 784 weights (28×28×1 since MNIST is normalized to black and white values), which is manageable for most forms of ANNs. However, considering a more substantial image, such as the X-ray images we processed in one of our studies with an image size of 256x256 pixels (even when normalized to black and white values), the number of weights on just a single neuron of the first layer could increase significantly. When we consider that to deal with this input scale, the network will also need to be much larger than one used to classify color-normalized MNIST digits, the benefits of CNNs become even more apparent.

2.4.4. Autoencoders, variational autoencoders and semi-supervised variational autoencoders

AutoEncoders

The variational autoencoder (VAE) builds on the autoencoder (AE) architecture, an ANN architecture that contains an encoding function, which maps an input to a compressed latent space representation, and a decoding function, which maps from the latent space back into the original space. The idea originated in the 1980s and was later promoted by Hinton & Salakhutdinov (Hinton and Salakhutdinov 2006). The word “auto” indicates that its learning is unsupervised, and the word “encoder” means that it learns encodings of data (Maheshwari, Mitra and Sharma 2022). The simplest architecture for constructing an AE is to limit the number of nodes of the network’s hidden layer(s), restricting the amount of information that can flow through the network. That way, the architecture of an AE includes a bottleneck that forces a compressed representation of the original input.

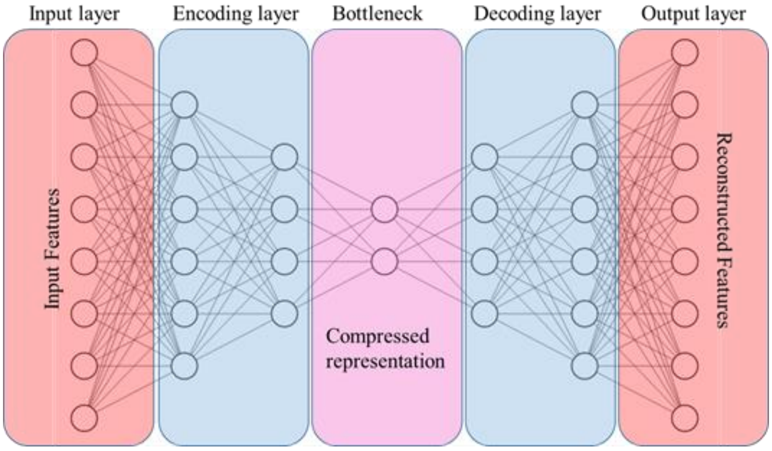

Fig 11: An illustration of the network architecture of an AE where the hidden layer works as a bottleneck that forces a compressed representation of the original input. The decoding layer maps from the bottleneck to the output layer.

The AE makes it possible to take an unlabeled dataset and task the network with outputting **y**, a reconstruction of the original input **x**. The AE is trained by minimizing the reconstruction error, which measures the differences between our original input and the reconstruction.

VAEs

VAEs maps output of the encoder model into parameters of a probability distribution over latent space. In other words, where the ordinary AE maps the input to a single latent representation vector, the VAE maps the input to a distribution over latent space vectors. Before diving deeper into the description of the VAE, I will briefly describe Bayes' theorem and probabilistic reasoning and how they relate to VAEs.

2.4.5. Bayes' theorem and graphical models

The idea of VAEs is deeply rooted in Bayesian statistics and graphical models, making use of variational Bayes methods for parameter inference (Kingma and Welling 2014). I start with explaining Bayesian statistics. At the heart of Bayesian statistics lies Bayes' theorem. Bayes theorem states that if there are two events A and B that each have probabilities P(A) and P(B) of occurring, then: $P(\mathrm{B}|\mathrm{A}) = \frac{P(A|\mathrm{B})\mathrm{P(B)}}{P(A)}$. Bayesian statistics makes use of Bayes' theorem to construct a probability distribution over the parameters of a model given data.

In Bayesian statistics, Bayes's theorem relates four factors to each other: "prior", the "likelihood", the "posterior", and the"evidence".

$$Posterior = \frac{likelihood * prior}{evidence}$$

Written in terms of latent variables **z** and data **x** this becomes:

$$P(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{z})\pi(\boldsymbol{z})}{p(\boldsymbol{x})}$$

I explain each of these terms in turn.

Prior - $\varphi(\mathbf{z})$

The Bayesian prior is the probability distribution of a parameter before any data is observed. This can be considered as a way of including "prior knowledge" in the model, hence the name. A prior can be chosen based on precise quantitative reasoning related to the nature of the parameter or empirical domain knowledge.

Likelihood - p(**x**|**z**)

The likelihood describes the probability of observing the data given the parameters of the model.

Posterior - p(**z**|**x**)

The posterior probability is the object of interest in Bayesian inference. It is the probability distribution of the parameters of the model given the data. The posterior can be considered as the prior updated with new information (brought in by the likelihood).

Evidence - p(**x**)

The probability of observing the data is referred to as the evidence. The evidence is the marginal probability of the data.

Using the Bayesian terminology above, the VAE can be illustrated as follows:

Using the Bayesian terminology above, prior = $p\theta(\mathbf{z})$, likelihood = $p\theta(\mathbf{x}|\mathbf{z})$, posterior = $q\varphi(\mathbf{z}|\mathbf{x})$ (which works as the approximation of the Bayesian posterior), evidence = $p(\mathbf{x})$ (not included in the figure above).

2.4.6. Variational Bayes

Variational Bayes is a method for approximating complex probability distributions using a simpler distribution known as a variational distribution. In the context of a VAE, variational Bayes is used to estimate the parameters of the VAE. In a VAE, the latent variable **z** is indirectly observable, leading to a scenario where both the marginal probability and the posterior distribution of **z** are computationally intractable. This forms an obstacle for the direct optimization of these parameters using gradient descent methodologies. To overcome this challenge, it is possible to use an auxiliary distribution, known as a variational distribution, **q**, to approximate the true posterior distribution over the latent variable. By introducing this auxiliary distribution, we can integrate ("sum out") **z** from the likelihood function, making it possible to use gradient descent to estimate the parameters of the VAE. The specific technique used to accomplish this is known as the reparameterization trick. The reparameterization trick involves re-parameterizing the sampling process of **z** in a way that allows the gradients to be computed with respect to the parameters of the VAE model. In practice, a quantity called the Evidence Lower Bound (ELBO) is maximized, which minimizes the difference between the posterior and the variational approximation in terms of the Kullback-Leibler (KL) divergence. The KL divergence is a standard way to evaluate the similarity of two probability distributions or densities.

It is important to note that while variational Bayes is "Bayesian" about the latent variable **z** (as **z** has a distribution), it is not Bayesian about the parameters of the neural network. For the parameters of the neural network, the VAE finds a single best value using gradient descent.

2.4.7. Variational AutoEncoders

VAEs consist of two interdependent but independently parameterized ANNs:

The decoder with its own parameters (weights and biases, θ), denoted by p(**x**|**z**), reconstructs data **x** based on the prior **z** and p(**x**|**z**)

The encoder, with parameters , denoted as q(**z**|**x**). The encoder works as an approximation that is needed to estimate the parameters of the decoder. The encoder takes a data point D_i as input and encodes a latent representation of this data to the latent space **z**. The encoder maps D_i to parameters of a distribution over and shares variational parameters across data points – a process referred to as amortization (Kingma and Welling 2014).

The optimization objective of the VAE is the evidence lower bound, abbreviated as ELBO. Maximizing the ELBO corresponds to maximizing the evidence, p(**x**).

As it happens, minimizing the distance between these two distributions in terms of KL divergence, also means maximizing the ELBO - something that Kingma and Welling explore in detail (Kingma and Welling 2014).

To provide a concrete example of how the VAE works, I will again use MNIST as an example. **x** is a 28x28-pixel image of a handwritten number. The encoder encodes the data of 784 dimensions into the parameters of a probability distribution p(**z**|**x**) over the latent space . The latent variable or latent representation has a much lower dimensionality than 784 (say, 50-dimensional). Typically, a Gaussian distribution is used for p(**z**|**x**). The images in MNIST are black and white. Each pixel can thus be represented as 0 or 1, and the probability distribution of a single pixel can be represented using a Bernoulli distribution. The decoder gets as input the latent representation of a digit **z**. The decoder then decodes the latent representation into 784 real numbers between 0 and 1 (the parameters of 784 Bernoulli distributions).

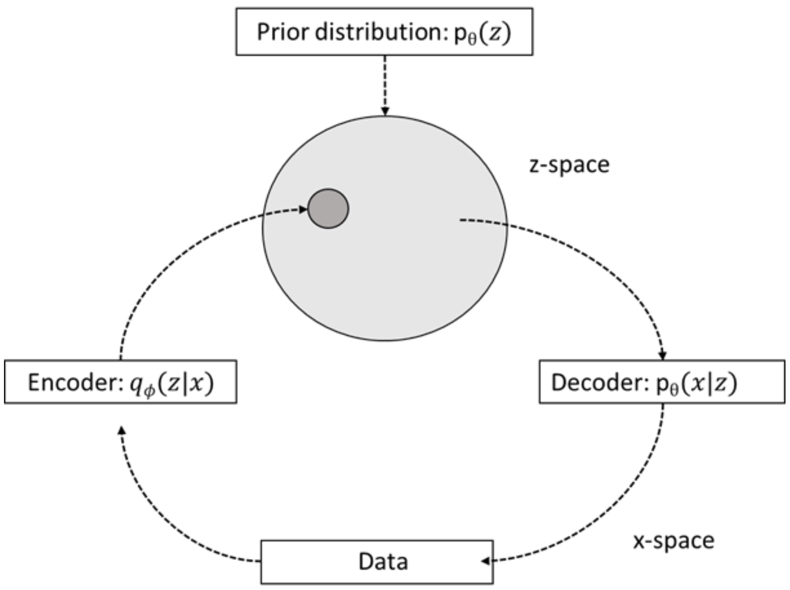

Fig 12: A graphical representation of a VAE. **z**=latent space, **x**=data, θ=parameters of the model, φ=parameters of guide.

2.4.8. SS-VAEs

In cases where the label of the input data is missing, it is possible to use the architecture of the VAE to estimate the label by using a so-called SS-VAE. This concept, first explored by Kingma and Welling, uses the encoding model to improve classification (Kingma and Welling 2014). If the label is missing, the decoder samples it from a partially observed latent variable ***y***. In this instance, the classifier $q\varphi(\mathbf{y}|\mathbf{x})$ serves as both the inference module for the supervised task and as an approximate posterior (and encoder) for the **y** variable (Kingma and Welling 2014).

2.4.9. MNIST

In this thesis, I sometimes refer to the MNIST (Modified National Institute of Standards and Technology) database (MNIST). MNIST is a dataset containing 70 000, 28x28 pixel black and white images of handwritten digits. It has historically been and continues to be, widely used for training and testing in the field of machine learning (Yann, Corinna and Bruges 2022).

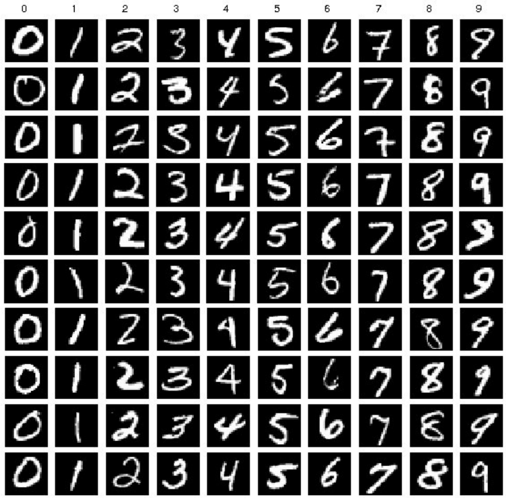

Fig 13: Sample images from MNIST test dataset of handwritten digits, image taken from (Lim, Young and Patton 2016).

3. RESEARCH PROBLEM DOMAIN

In this chapter, the focus is on the research problem domain and the research questions related to AI transformation in the manufacturing industry. The chapter begins by establishing the relationship between AI and manufacturing companies, exploring how AI can be utilized in various manufacturing processes and the potential benefits associated with its adoption. The chapter also investigates the barriers that could prevent an AI transformation of the manufacturing industry, including technical, human and organizational, and societal barriers. A discussion of the leadership implications of these barriers to introducing AI follows. Finally, the chapter sets out the research questions aimed at effectively supporting a widespread implementation of AI in the manufacturing industry. This chapter aims to further contextualize the subsequent chapters and continues the in-depth exploration of the leadership aspects of introducing AI into the manufacturing industry.

3.1. AI and how it relates to manufacturing companies

According to researchers such as Govindarajan and Immelt, Foutaine et al., and Ancona, AI will transform the manufacturing industry, bringing significant consequences for companies, workers, and consumers (Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019, Ancona 2019). AI is already widely used in medical image analysis, bioinformatics, drug discovery, recommendation systems, financial fraud detection, visual art processing, and the military (Kim et. al 2022).

On the one hand, AI's current direction has been questioned by some researchers. They argue that it cannot distinguish causation from correlation and cannot deal with the unexpected (Marcus 2018). They have also pointed out drawbacks associated with the interpretability of AI models (Kim et. al 2022), and that we are in a diminishing returns territory in relation to AI (Chollet 2017).

On the other hand, other researchers argue that AI has the potential to disrupt the manufacturing industry and that it is widely agreed that AI is transformative (Govindarajan and Immelt 2019, Davenport and Foutty 2018, Fountaine, McCarthy and Saleh 2019). They believe that it can improve operational performance and provide new business models and that those manufacturing companies that do not commit fully to an AI

transformation will be outcompeted by those who can offer new data-driven services (Govindarajan and Immelt 2019, Ancona 2019).
Many manufacturing companies are challenged in meeting throughput, quality, and cost objectives while ensuring a safe working environment. Meeting these goals can be increasingly difficult with the multitude of demands stemming from growing product and process complexity, higher variability in customer demand and preferences, and relentless competitive pressures (Arinez, et al. 2020, Govindarajan and Immelt 2019, Soldatos and Kyriazis 2020).
This leads to the question if computers can help meet these challenges. This is not a new question. In the 1980s, one of the most discussed topics in manufacturing was computer-integrated manufacturing (CIM). CIM was an effort to integrate activities in manufacturing through the medium of computers. Now it is widely accepted that CIM leads to widespread improvement in productivity (Adiga 1993). While researchers such as Adiga posits that computers can improve operational performance, the actual benefit of AI in manufacturing is still unclear. Even though researchers such as Arinez et al. and Govindarajan and Immelt and Deveraux claim that recent developments indicate that AI could be one of the most disruptive enablers of Industry 4.0 (a German strategic initiative, aimed at creating intelligent factories where manufacturing technologies are upgraded and transformed by cyber-physical systems), with the potential to transform the manufacturing domain through advanced analytics tools, they have not quantified the level of disruption or potential that AI can provide (Arinez, et al. 2020, Govindarajan and Immelt 2019, Deveraux 2019, Lee, Bagheri and Kao 2015). According to a recent report from McKinsey, among manufacturing companies that have implemented AI on a wide scale, it is not uncommon to see 30-50% reductions in machine downtime, 10-30% increases in throughput, 15-30% improvements in labor productivity, and 85% more accurate forecasting (McKinsey, Capturing the true value of Industry 4.0 2022). However, I have not seen these numbers verified in other research.

Irrespective of its potential, many manufacturing companies have invested in the digital transformation of their production processes to transition to Industry 4.0 (Palensky, et al. 2008, Pan 2016). In Cyber-Physical Systems (CPS), with backbone technologies like cloud computing, and internet of things (IoT), AI applications can be used to analyze data from embedded sensors and instrumentation or data generated from manufacturing processes (Govindarajan and Immelt 2019, Soldatos and Kyriazis 2020, Arinez, et al. 2020). Advances in computational hardware and sensing technology for collecting critical process/machine data have made the application of AI feasible in a practical sense and led to an interest in the capabilities and benefits they offer (Arinez, et al. 2020, Chien, et al. 2020). Furthermore, new advanced machine learning frameworks allow for novel algorithms suitable for large-scale problems in realistic settings (Lwakatare, et al. 2020). The increased industrial interest in AI is also reflected in the number of journal articles on AI in manufacturing, mainly driven by research from China and USA, (Zeba, et al. 2021).

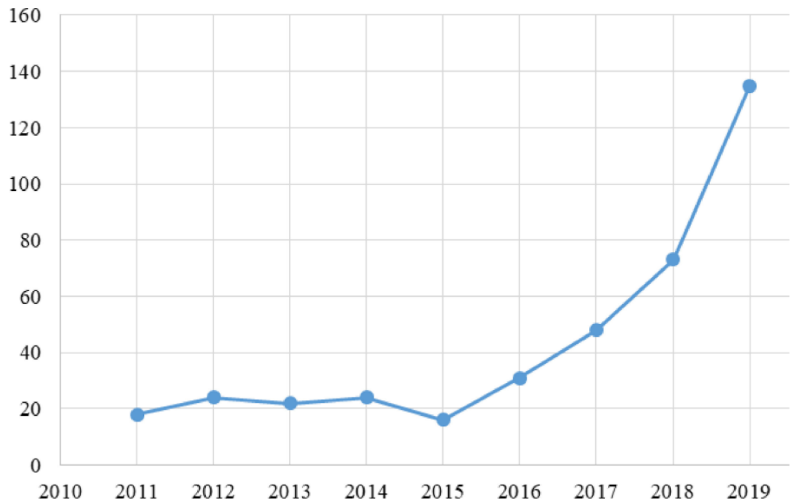

Fig 14: the number of journal articles by year during the period between 2011 and 2019 on the topic of AI in manufacturing, image taken from (Zeba, et al. 2021).

3.2. Related research on the use of AI in manufacturing

This thesis investigates leadership implications and challenges associated with the effective and widespread implementation of AI in manufacturing companies. To obtain an overview of relevant research on areas of AI applications in the manufacturing industry, I conducted a literature search using the following key concepts as inclusion criteria: "*AI and manufacturing,*" "*deep learning and manufacturing,*" and "*machine learning and manufacturing.*". However, obtaining an overview of research related to this study proved challenging due to the broad search criteria.

Therefore, I refined the search results and combined the key concepts with the keyword "*systematic review*". "*Systematic review*" was chosen to gain insights from comprehensive compilations of available research within the defined subject area. Systematic reviews are typically positioned at the top of evidence hierarchies as they adhere to strict methodologies and criteria, mitigate the risk of bias and ensure the inclusion of all relevant research. They allow the researcher to synthesize and critically appraise a number of studies in a specific context to provide evidence-based conclusions, (Okoli and Schabram 2010, Webster and Watson. 2002), which enables a more focused understanding of the existing research landscape. This approach allowed me to scan and read through the review articles to qualitatively select further articles for deeper insights on relevant studies and use of AI in manufacturing application areas.

3.2.1. Overview of the findings

The literature search identified 15 studies that met the search criteria. The search results revealed that extensive research has been conducted to explore the potential benefits of AI within the manufacturing industry. After excluding systematic review studies investigating unrelated topics such as comparisons of inherent differences of smart manufacturing and intelligent manufacturing (Wang et al. 2021), and the role of circular economy in manufacturing (Acerbi, Forterre and Taisch 2021), I ended up with eight systematic review articles offering systematic reviews of studies on application of AI in manufacturing. These reviews serve as a foundation for understanding the current state of applications in the field.

These articles identified multiple areas where AI could be applied in manufacturing companies. For example Arinez et al. concluded that AI could be used for job dispatching, scheduling and resource allocation (Arinez, et al. 2020) and Fahle et al. identified research that AI could be used for assistance and learning systems, assembly assistance and logistics (Fahle 2020). Based on the qualitative analysis of the articles, five recurring themes emerged regarding areas in manufacturing where AI could be introduced.

Specifically, the studies conducted by Fahle et al., Arinez et al., Hansen and Bøgh, Jamwal et al., Cioffi et al., and Kofi Nti et al. identified the significance of AI in predictive maintenance (where data from sensors and other sources are analyzed to detect patterns that can help predict when a machine is likely to fail or require maintenance) and quality control (where AI is used to detect patterns that may indicate a defect, allowing allows businesses to catch issues earlier in the production process), which could lead to improved manufacturing processes and reduced downtime. The same studies emphasized the potential of AI in enhancing human-robot collaboration, making robots more adept at sensing their surroundings, understanding human intent, and adapting to changing environments. This can help robots work more effectively with humans, leading to increased efficiency and safety in the manufacturing environment. Additionally, these studies highlighted the potential of AI in optimizing and controlling various manufacturing processes, which could improve throughput and potentially affect the production quality (Fahle 2020, Arinez, et al. 2020, Hansen and Bøgh 2021, Jamwal et al. 2022). Furthermore, research conducted by Fahle et al., Jamwal et al., and Cioffi et al. demonstrated that AI and machine learning had applications in streamlining supply chain activities enabling better demand forecasting, inventory management, and production planning (Fahle 2020, Jamwal et al. 2022, Cioffi et al. 2020). Finally, Cioffi et al. and Sung et al. highlighted the role of AI in managing energy consumption in manufacturing, which could lead to reduced costs and increased environmental sustainability (Cioffi et al. 2020, Sung, et al. 2022).

3.2.2. Implications of the literature search

The studies mentioned above provide insights into the field of AI applications in manufacturing, with a majority of research being conducted in recent years. Notably, China has surpassed the United States as the leading producer of research in this domain (Zeba, et al. 2021). Initial studies primarily focused on topics such as "*flexible manufacturing systems*" and "*decision support*", while more recent articles have shifted their attention to "*cyber-physical*", "*deep machine learning and big data*" and "*neural networks*". This indicates an increased emphasis on investigations of the potential of deep learning to process large amounts of data to build cyber-physical manufacturing systems, as noted by Zeba et al (Zeba, et al. 2021). This trend is also supported by the increased use of neural networks and decision tree algorithms in recent years (Fahle 2020). Further, it is anticipated that emerging topics, such as cyber-physical integration manufacturing strategies, smart and sustainable manufacturing, computational sciences for smart factories, and the development and implementation of novel technologies, will garner increased research interest in the future (Zeba, et al. 2021). The growing adoption of AI technologies across small and medium-sized enterprises illustrates the broad applicability of these technologies in manufacturing companies of varying sizes and scopes (Hansen and Bøgh 2021).

The results of the literature study provide insights into potential AI applications in the manufacturing industry, focusing on the manufacturing process as such. To a lesser degree they investigate the how or why of their introduction. Moreover, I identified a gap concerning the organizational aspects of introduction and how it affects management practices and the research failed to examine the practical implications of effectively implementing AI in this context. While general management literature, such as the works of Govindarajan and Immelt and Fountaine et al. often referenced in this thesis, attempts to address the questions of how and why manufacturing companies should adopt AI, it does so without specifying particular areas of application or considering the practical challenges during implementation (Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019).

Furthermore, the research in the systematic literature reviews tends to overlook AI's potential to transform workflows, promote innovative business models, and support the adoption of new management practices, which enable AI-driven decision-making (Lee, et al. 2019). Zeba et al. affirm the need for research exploring these transformative effects (Lee, et al. 2019). While the impact of AI on business model innovation in other industries, as demonstrated by companies like Airbnb and Uber, is well-documented, its influence within the manufacturing industry remains uncertain (Lee, et al. 2019).

Nonetheless, the management consultancy firm Marketsandmarkets estimates that AI in manufacturing will be valued at USD 2.3 billion in 2022, with a projection of reaching 16.3 billion by 2023, growing at a CAGR of 47.9% from 2022 to 2027 (MarketsandMarkets 2022). This growth can be attributed not only to productivity enhancements but also to emerging trends such as product personalization, mass personalization, and products-as-a-service (Wang, et al. 2017). For instance, Nike has generated over $185 million in revenue from non-fungible tokens (NFTs) for digital sneakers and similar products (McKinsey, Value creation in the metaverse 2022), while GE has shifted from selling engines or engine spare parts to providing "engine availability," charging customers based on engine usage hours (Govindarajan and Immelt 2019, GE Aviation 2022).

3.3. Lack of adoption of AI

Although the literature study confirmed that there are multiple areas and applications where manufacturing companies could use AI, research indicates that the implementation of AI in non-tech industries remains limited. Few companies have integrated AI into their operations and business strategy in a way that fundamentally changes how the organization functions, adds value, and engages with stakeholders (Agrawal, Gans and Goldfarb 2018, Holmström 2022, Govindarajan and Immelt 2019, Davenport and Foutty 2018).
According to Arinez, most research on AI in manufacturing has been conducted outside the daily operations of manufacturing companies (Arinez, et al. 2020). Further, as indicated by Fountaine et al., companies often apply AI to a single business process (Fountaine, McCarthy and Saleh 2019). Consequently, even if companies have begun to adapt and implement AI into operations to improve production efficiency and flexibility and reduce cost, its spread seems to be limited (Arinez, et al. 2020). According to Fountaine et al., only 8% of firms across industries engage in widespread adoption. 92% either do nothing, run ad hoc pilots, or apply it to a single business process (Fountaine, McCarthy and Saleh 2019). A Danish study in which only three percent of Danish companies actively use AI in their operations supports this view (Humlum and Meyer 2020).
It is paradoxical that despite the urgency highlighted by experts like Govindarajan and Ancona, few manufacturing companies seem to have adopted AI on a wide scale. This raises the question of whether the lack of widespread implementation, although research points to the possibility of AI providing benefits for manufacturing companies, is due to a lack of urgency or if it is because implementing AI is more difficult than anticipated. Against this background, it is crucial to explore some of the reasons behind the limited adoption of AI in the manufacturing industry.

3.4. Barriers to an AI transformation of the manufacturing industry

Below, I will explore barriers that I have identified through my literature studies of state of the art literature within the research domain, throughout my research studies as well as through my practical work in implementing AI in the manufacturing industry. The section does not provide an exhaustive list of barriers, but is rather intended to bring the awareness of the reader to some of the obstacles to AI transformation that current research has identified.

In summary, technical challenges such as the dynamic and complex nature of manufacturing environments and the need for high-performing AI models, human and organizational factors such as lack of trust and awareness of AI, difficulty in integrating traditional and new ways of working, and scarcity of required talent and societal barriers such as a lack of successful examples to emulate and the challenge of keeping pace with rapid advancements in AI methodologies all play a contributing role.

3.4.1. Technical barriers

Sung et al. argues that applications of AI in manufacturing industries have been particularly challenging due to the demand for high level performance of models in a highly nonlinear environment and in a high-dimensional space (Sung, et al. 2022) In addition, AI technologies applied in manufacturing are naturally different from those applied in other fields. In a manufacturing environment, AI must help people, machines, and systems communicate with each other. In contrast, in other fields, AI technologies are mostly applied to assist people (Chen and Wang 2022).

Further, the development, implementation and maintenance of large and complex AI systems is difficult. The difficulty is accounted for by, among other things, the dynamic and complex context where different components interact with one another and the environment (Sculley, et al. 2015, Dahlmeier 2017). Other researchers have raised concerns in relation to adaptability, scalability, safety and privacy of AI solutions in manufacturing and highlight these as some of the main barriers to a widespread implementation (Lwakatare, et al. 2020).

Even though the above indicates that there are several technical barriers to overcome to succeed with an AI transformation other researchers posit that an AI transformation is primarily a cultural and organizational challenge (Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019, Davenport and Foutty 2018, Brock and Wangenheim 2019).

3.4.2. Lack of awareness, trust and complacency

One such challenge is, according to Sung et al, that there is still a reluctance among manufacturing companies to adopt AI in their operations (Sung, et al. 2022). Many companies need more awareness of where and how it should be incorporated, lack trust in the technology, and have workers or leaders who fear becoming obsolete (Sung, et al. 2022, Fountaine, McCarthy and Saleh 2019, Ancona 2019).

Fitzgerald et al., and Govindarajan and Immelt highlight complacency among leaders, often rooted in ignorance about AI, as one of the main obstacles to a successful transformation (Fitzgerald, et al. 2013, Govindarajan and Immelt 2019). They posit that manufacturing companies often have no history of working with data and base their business models on producing hardware. Since these old models often perform at their peak at this stage, the incentive for change is at its lowest. In addition, the returns on long-term AI investments usually take time to realize and often pay off long after the current leadership is gone (Govindarajan and Immelt 2019).

3.4.3. Ambidexterity challenge; combining traditional ways of working while pioneering new ways

Another challenge is the balancing act between keeping what functions well in the company against the need to change (Govindarajan and Immelt 2019). Govindarajan and Immelt and Teece point out the traditional ways of working as some of the main obstacles to succeeding with an AI transformation (Govindarajan and Immelt 2019, Teesce 2014).

Others point out that these traditional ways of working, the best practices and bureaucracies, often are what made companies so successful in the first place (Westerman, Soule and Eswaran 2019, Fitzgerald, et al. 2013). They mention that bringing too much change at the same time or changing things too quickly risks destroying the foundation on which they capitalize. Further, they argue that companies that push the transformation too hard and fail could risk destroying what works well, losing their credibility, and scaring off their best employees (Westerman, Soule and Eswaran 2019, Fitzgerald, et al. 2013).

Further, according to Fountaine et al., AI has the most significant impact when developed by cross-functional teams with a mix of skills and perspectives. Business and operational people working alongside analytics experts will ensure that initiatives address broad organizational priorities, not just isolated business issues (Fountaine, McCarthy and Saleh

2019). However, bringing these practices to companies whose structures, values, and governance rules are designed for cautious stability is difficult (Westerman, Soule and Eswaran 2019).

The above indicates that, in order to succeed with an AI transformation, manufacturing companies need to establish ways of working where they keep what works well while simultaneously being able to explore new ways. Such "*ambidextrous organizations*," as Charles A. O'Reilly and Michael L. Tushman call them, allow pioneering of radical or disruptive innovations while also pursuing incremental gains (O Reilly and Tushman 2004).

3.4.4. Scarcity of talent

Yet another challenge is the lack of skilled staff needed to develop and implement AI (Brock and Wangenheim 2019). Succeeding with an AI transformation requires access to the programmers able to build AI models (Fountaine, McCarthy and Saleh 2019). However, according to an article in the Economist from 2022, only 25 million people worldwide are fluent in standard programming languages. In 2025, there will be an estimated global shortage of about 4 million programmers (Economist 2022). In addition, it is often challenging to recruit and retain the non-technical experts required to develop and implement AI solutions, such as project managers and what Fountaine et al. refers to as "translators" (Fountaine, McCarthy and Saleh 2019). Naturally, the competition for both technical and non-technical talent is fierce, and it can be difficult for manufacturing companies to compete for this talent. Even if they manage to recruit them, they may struggle to retain them. Manufacturing companies often have no natural place to put these people other than in a traditional technology function such as IT or engineering, where they risk succumbing under the bureaucratic processes and best practices that characterize traditional functions (Teesce 2014).

However, the technical and organizational barriers mentioned above do not fully explain the challenges in transforming the manufacturing industry. Challenges on a societal level also have an impact.

3.4.5. Lack of mimetic models for both companies and researchers

The American-French anthropologist and cultural theorist René Girard posited that humans turn to others to understand how we should act because humans imitate others' behavior. Girard refers to those we imitate as "mimetic models" (Palaver 2013). Govindarajan and Immelt, Fountaine et al., Sung et al., and Davenport posit that there are not many "AI-driven" manufacturing companies to benchmark against (Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019, Sung, et al. 2022, Davenport and Foutty 2018). Accordingly, the manufacturing industry lacks successful mimetic models to imitate and learn from. Instead, every company needs to make its exploratory journey, figuring out the best way to introduce AI. This exploratory journey often forces them to draw their game plans on the fly (Govindarajan and Immelt 2019). In an environment used to making long-term plans and steering innovation, this often leads to tension and, consequently, a slow transformation pace.

Additionally, the absence of successful examples means that researchers struggle to find suitable manufacturing companies to study. Researchers such as Govindarajan and Immelt, and Fountaine et al. speak of what companies need to do, drawing from their experience of working with other industries or from what they have learned in working with digital transformation in manufacturing companies (Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019). However, many of their conclusions seem to be based on isolated experiences and anecdotal evidence.

3.4.6. Keeping up with methodological progress

While the AI transformation of the manufacturing industry appears to be moving at a slow pace, the methodological progress within AI is moving at a high rate (Bommasani 2021). This creates additional challenges for manufacturing companies willing to explore AI. Within the field of AI, new methods, e.g. encapsulated in new programming frameworks, continuously emerge and replace current ones (Davenport and Foutty 2018, Govindarajan and Immelt 2019, Weill 2019, Bassellier, Benbasat and Reich 2003, Persaud 2021, Bommasani 2021). In less than a decade, this development has made it possible to develop AI applications that previously required in-depth knowledge of computer science, statistics, and mathematics (Bommasani 2021). Fountaine et al. and Davenport argue that the rapid progress within AI, which to some extent is reflected in a significant increase in the research literature on AI in manufacturing, makes it increasingly important to be able to use the latest methods to innovate rapidly (Fountaine, McCarthy and Saleh 2019, Davenport and Foutty 2018, Zeba, et al. 2021). Govindarajan and Immelt and Ancona point out that, failing to do so, manufacturing companies risk wasting time using obsolete methods and ultimately risk being outcompeted by those who quickly adopt the latest methods (Govindarajan and Immelt 2019, Ancona 2019).

3.5. The leadership perspective

From the above, it becomes clear that implementing AI is a leadership challenge. This has been confirmed in previous research showing that implementation of AI requires that leaders of manufacturing companies understand how to best work with AI and what consequences it may bring (Fountaine, McCarthy and Saleh 2019, Sung, et al. 2022, Brock and Wangenheim 2019). However, it can be difficult for leaders to understand how to best support an AI transformation (Fountaine, McCarthy and Saleh 2019, Brock and Wangenheim 2019) since different, and sometimes conflicting views are spread across different research disciplines. Additionally, much of the research focuses on the implementation of AI in isolated areas of the business and not a widespread implementation (Arinez, et al. 2020).

Despite the lack of direction on the best way for leaders to support an AI transformation, researchers such as Govindarajan, Immelt, and Ancona point out the urgency to change. They state that if leaders fail, their companies will be outcompeted by those who can offer new data-driven services, risking extinction (Govindarajan and Immelt 2019, Ancona 2019).

Different researchers have tried to advise leaders on how to effectively support an AI transformation from different perspectives (Arinez, et al. 2020, Govindarajan and Immelt 2019, Ancona 2019, Kolbjørnsrud, Amico and Thomas 2017). However, since not many companies beyond the tech industry leverage AI on a wide scale, it is a little speculative to understand how to best support a widespread implementation in a manufacturing company (Davenport and Foutty 2018). There is also little research integrating the technical implications of implementing AI on a wide scale in manufacturing companies with the leadership challenges that follow. In light of the above, further research is needed to understand what it takes for a leader to effectively support an AI transformation of a manufacturing company (Govindarajan and Immelt 2019, Davenport and Foutty 2018).

3.6. Research questions and aim

Many gaps still exist that could be addressed on what is required to integrate AI into manufacturing companies on a wide scale. This thesis aims to bridge some of these gaps by answering the following overarching research problem; what are leadership implications and challenges for the effective widespread implementation of AI in manufacturing companies?

I address the research problem through the following research questions:
(i) What capabilities can benefit leaders who wish to implement AI in a manufacturing company, and how do these capabilities affect how leaders support an AI transformation?

In order to fully understand capabilities required, it is necessary to consider the nature of AI technology itself and its potential implications on AI implementation. Therefore, I conducted a second study, aiming to investigate:

(ii) What is the impact of the nature of AI technology on efficiency, democratization and ethical concerns and how can the increasing abstraction in deep learning software frameworks reshape the relationship between technology and society?

After investigating the leadership aspects and technology aspects of effective AI implementation, I saw it necessary to gain practical experience in AI implementation. Against this background, I took an action research approach to answer the research question:

(iii) What challenges and considerations might arise when implementing an AI solution in a manufacturing company's daily operations? How can these companies adequately develop in-house AI capabilities and what factors influence their decision to continue developing these capabilities or outsource them?

4. OVERALL RESEARCH DESIGN

In this chapter, I discuss the research design of this thesis, which draws inspiration from practice-based research and utilizes action research methodology. I reflect on my role as a reflective practitioner and the challenges this poses to objectivity. Additionally, I introduce the company that funded my research, GKN Aerospace, and provide an overview of the study's results, study cases, and contributions.

I describe the three studies that I conducted as part of my research. The first study aimed to establish a leadership competency framework for leaders of manufacturing companies that wish to implement AI on a wide scale in their business. The second study focused on quantifying the increase in the abstraction of deep learning and discussing its implications with respect to technological advancement, democratization, ethical concerns, and concentration of power. The third study aimed to provide guidance on developing in-house AI capabilities and examined factors that influence the decision to continue developing these capabilities or outsource them.

4.1. Practice-based and action research

This research concerns the leadership practice of implementing AI in the manufacturing industry. Research that takes the nature of practice as its central focus is called ‘practice-based’ research and is usually carried out by practitioners such as designers, writers, programmers etc. (Candy 2006). This approach has influenced this research. The concept of practice-based research originates from the idea that knowledge can be partly advanced through practice in certain disciplines. In practice-based research, the researcher can take the practice of their discipline as the research subject. The research program then consists of a continual reflection upon that practice and on the resulting informing of practice (Candy 2006). Even though the term practice-based research is widespread, it is not yet defined in a way that is agreed upon across the various fields of research where it is in use (Candy 2006). Practices are often complex, and one is not an isolated instance; instead, one is often intertwined with other practices (Nicolini 2012). As practices emerge, new requirements of practice can occur. The current research has been shaped over almost four years through continuous reflections considering what was learned in the previous parts of the research. These reflections have informed the decisions made on which further research to pursue.

In the context of practice-based research, it is helpful to distinguish between a “pure” practitioner and a practice-based researcher. According to Scrivener, the critical difference is that the practice-based researcher aims to generate novel apprehensions that are “not just novel to the creator or individual observers” (Scrivener 2002). While parts of the research are a result of the researcher’s particular goals of the time, the aim is also to add to the shared store of knowledge around this subject in a more general sense, providing insights that are not only new to the creator. The research set out in this thesis is a result of both quantitative and qualitative studies. The direction of the research has been informed by practitioner needs identified in current research. The research as such has been conducted according to the structured research process common to professional practice.

Another main difference is the form that the generated knowledge takes. While the practitioner is mainly interested in furthering its own individual goals of the time rather, the practice-based research outcome is shared with a broader community and arises from a structured process. A vital element of this process is that the research results are transferable to a broader research community. In the current case, I have generated research articles and this thesis to transfer my research conclusions to the research community.

In addition to the research approach being practice-based, parts of it draw inspiration from a qualitative research approach that follows the methodology of so-called action research, a term commonly used in information systems research. In action research, the researchers insert themselves into the context of their investigation (Järvinen 2001). This approach is unusual compared to other methods, where the researcher typically acts as an unbiased observer (Reason and Bradbury 2001).

In action research, action can inform practice, as theory can be created through practice. Action research emphasizes the collaboration between practitioners and researchers (Avison, et al. 1999, Baskerville and Wood-Harper 1996, Brydon-Miller, Greenwood and Maguire 2003, McKay and Marshall 2001). The research in this thesis results from collaborations between the researcher in his role as a practitioner, other industry practitioners such as programmers, and researchers from various fields.

The research approach has also been inspired by an action research model created by Mathiassen et al. (Mathiassen, Chiasson and Germonprez 2012).

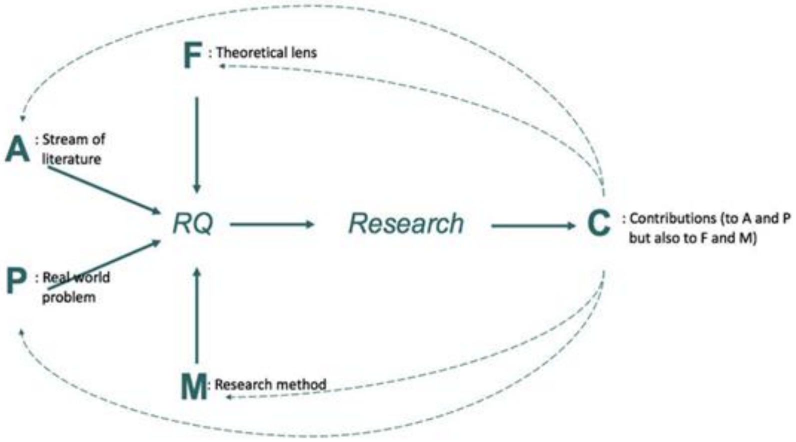

Fig 15: An illustration of the research model for action-based research based on the work by Mathiassen et al., where the research question is based both on relevant literature and a real world problem (Mathiassen, Chiasson and Germonprez 2012). The research contributes both to solving the real world problem and to relevant research within the area. Image taken from (Islind 2018).

According to this model, and in this research, the research questions, as well as the research, are continuously informed by relevant streams of research and real-world problems. Similarly, the research contributes to a solution to real-world problems and the shared store of knowledge around this subject in a more general sense. The stream of literature (A in the figure) consists of the work generated in this thesis and the results of the literature studies. The real-world problem (P in the figure) is the question of what is required from a manufacturing company leader to efficiently support an introduction of AI and the barriers to such an introduction. The theoretical framing (F in the figure) consists of the practice-based research lens as well as the technological framework as well as literature studies. The method (M in the figure) is a mixed-method. The research questions are in the middle of the figure below, and guides the researcher´s approach to its contributions, and the contribution (C in the figure) consists of three articles.

In this thesis, I am using a mixed methods research approach. In parts of the thesis, I have explored the complexities of social phenomena by achieving an empathic understanding of how the research subjects view the world. Collis and Hussey describe this as the interpretivism research philosophy. It sees reality as highly subjective because our perceptions shape it. In other parts, I have been inspired by a positivist approach, making use of statistical hypothesis testing where needed (Collis and Hussey 2014).

4.2. The role of the researcher - the reflective practitioner

Coming from a background in law, I have spent the last years leading digital transformation. Throughout this research, I have led a team responsible for digital innovation at GKN Aerospace. Holding two roles - both the role of researcher and practitioner has given me a position to conduct practice-based research, immediately trying out what I have learned in my research in the field. Similarly, the experiences from my position in the company have provided valuable insights for my research. Combining the outside-in perspective of the researcher with the inside-out perspective of the practitioner, I base the foundation of this thesis on industry and practitioner needs that I have identified in the research literature and have experienced first-hand. These needs and works have informed my writing at every level while ensuring academic stringency in my research studies and historical accounts. In addition, theories learned during the research have furthered the practical experiences learned in the manufacturing industry.

When I started this research, the first challenge I faced was where to begin. My role as a practitioner was to ensure that the company I worked for leveraged AI in the best possible way. In this work, I asked myself the question; what is required from a person in my position to advance the implementation of AI within a manufacturing company? I was looking for guidance in this regard. An AI transformation requires the support of the business leadership and must begin at a leadership level (Fountaine, McCarthy and Saleh 2019, Brock and Wangenheim 2019). This view seems to be shared by Ngwenyama and Nørbjerg, who suggest that large-scale software transformations depend on a firm commitment from top management (Ngwenyama and Nørbjerg 2010). However, as I mentioned above, even though researchers have identified a need for research on how to support a widespread implementation of AI, there is limited research on what it takes to be an AI-driven manufacturing company leader.

Practice-based research, where the researcher is partly subject to the research, can be criticized for the notion that it can be challenging for researchers to be objective about the research context while being heavily involved within the context (Baskerville and Wood-Harper 1996, Bryman 2015). In this research, I have not always been "a fly on the wall", but rather "a fly in the soup". By that, I mean that as a leader and researcher, I have aimed to be a part of the change process and intervene. It has been complex being a leader responsible for implementing AI and a researcher simultaneously, as this approach naturally eliminates the role of a neutral observer. Aiming to maintain objectivity as a researcher (Baskerville and Wood-Harper 1996, Bryman 2015), I have ensured to include other researchers and practitioners. In the first study I conducted, the aim was to understand what was required from a leader to advance the implementation of AI within a manufacturing company. Throughout this study, my principal supervisor (Assoc. Prof. Thomas Hamelryck), and Prof. Ulrika Lundh Snis, University West, were heavily involved. In the second study, I participated in the practical steps of programming and analysis of GitHub repositories with an industry programmer. I also led the development of the theory. The theory, however, was developed together with my principal supervisor, a researcher with a background in anthropology and philosophy (Cadell Last, PhD, University of Brussels, VUB), and Prof. Ulrika Lundh Snis In the third study, I provided direction in the development of the models and led the discussions with the internal and external stakeholders and the external service provider. All this while simultaneously studying the subject. However, the model was used in the study was developed with an industry expert and my principal supervisor with assistance from a researcher within computer science specialized in probabilistic models in close collaboration with the inspection operators of the business. The theory, however, was developed together with Prof. Ulrika Lundh Snis.

Much published research does not contribute to advances in academic knowledge while at the same time enlightening professional practice (Van De Ven 2007). Donald Schön describes academia as "institutions committed to a particular epistemology, a view of knowledge that fosters selective inattention to practical competence and professional artistry" (Schön 2016). Similarly, I have experienced a reluctance among practitioners to adopt the thoughtfulness and systematic methodology used in academia. Against this background, I want to promote a view of knowledge, as promoted by Schön and Jonna Bornemark, that bridges the gap between the knowledge honored in academia and the kinds of competence valued in professional practice, where both views contribute to one another (Schön 2016, Bornemark 2018).

In light hereof, the thesis investigates, from a highly multidisciplinary point of view, some of the challenges I have faced in my learning journey over the past three years when leading an introduction of AI in a manufacturing company. I combine quantitative methods and qualitative methods with real-life examples.

4.3. The aerospace research context

The aerospace component manufacturing corporation GKN Aerospace Engines – a branch of GKN Aerospace, funds this research. Initially founded as Dowlais Ironworks Co in 1759 in Dowlais, South Wales, GKN's aerospace branch, GKN Aerospace, today employs about 15,000 employees at 38 manufacturing locations and 4 R&D-centers in 12 countries around the world. It serves the world's leading aircraft and aero-engine manufacturers with its advanced technologies that improve the performance of more than 100,000 flights daily. GKN Aerospace has a unique reputation for quality and innovation and collaborates closely with universities, knowledge institutes, suppliers, and customers. It leads the industry in developing new technology to improve aircraft efficiency: lowering aircraft cost, weight, and emissions (GKN 2023).

GKN Aerospace would like to understand the potential that AI may bring to its business and it runs pilots in parts of its business. Like many other manufacturing companies, GKN Aerospace would benefit from increased machine uptime, increased production throughput and labor productivity, and more accurate forecasting. However, it still assesses the best ways AI could help with this. Against this background, GKN Aerospace Engines decided to fund this PhD research. The research in this thesis is concerned with implementation of AI in the manufacturing industry. However, I have studied this through the lens of a practitioner within an aerospace manufacturing company. The aerospace manufacturing industry has thus been a study object of the manufacturing industry.

5. RESULTS AND CONTRIBUTIONS

5.1. Paper 1: An AI leadership competency framework, Appendix 1 - *published at the 26th Biennial Nordic Academy of Management Conference, Örebro, Sweden, August 24-26, 2022*

There are principles guiding other parts of leadership within manufacturing companies than AI implementation. An example of such a leadership framework is the LEAN principles, which provide a framework for creating an efficient and effective organization, allowing managers to reduce waste and deliver better customer value (hence the name "Lean") (Dombrowski and Mielke 2013). However, there has yet to be a consensus on what capabilities leaders of manufacturing companies require to implement AI. Different views are spread across different research disciplines. Neither are there many "AI-driven" manufacturing companies to benchmark against (Govindarajan and Immelt 2019, Davenport and Foutty 2018, Teesce 2014). In light hereof, researchers within general management have pointed out the need for a framework describing what capabilities leaders of manufacturing companies need to develop and establish to implement AI (Felin, et al. 2012, Govindarajan and Immelt 2019, Davenport and Foutty 2018).

In this context, against these practitioner needs identified in the literature, I initiated my first study to establish a leadership capability framework for leaders of manufacturing companies that wish to implement AI on a wide scale in their business. I also wanted to understand how differences in leaders' ability to live up to the constituents of my framework affected how they implement AI and, consequently, the AI maturity of their companies. Against this background, I designed a study using a mixed methods approach, combining a literature study, a survey, and in-depth interviews. I conducted the study with my supervisor Thomas Hamelryck, Dpt. of Computer Science, KU and Ulrika Lundh Snis from the Department of Business and IT at University West.

Literature study

I first made a literature study on articles on AI and leadership where I identified and analyzed relevant streams of research on AI competence, leadership, and management by studying articles from high impact journals within general management (the study of the techniques, practices, or sciences related to managing a company), innovation management (the study of the process of managing innovative ideas) and information systems (the study of the interaction between hardware, software, users and business processes) respectively. Based on this literature study, I found that even if there is no consensus in the literature on what characterizes an "AI leader" and hence no playbook to adhere to (Davenport and Foutty 2018, Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019), there were some recurring themes of abilities and routines that characterize leaders that successfully implement AI. These themes came to form my proposed AI-leadership capability framework. In summary, I found that leaders must be "learners" and proficient strategists, provide and communicate a compelling vision and have strong social skills. Finally, they must be prepared to change both how they lead the companies and how their companies work.

Survey

In the second step of this study, we identified companies with different AI maturity to understand how their leaders' ability to live up to the capability framework's constituents affected their AI maturity. We created a web-based survey to assess the AI maturity of companies. The survey focused on deep learning to avoid confusion about the meaning of AI. We distributed the survey to senior executives within general management, IT, engineering, HR, and R&D at 11 major high-tech multinational manufacturing companies within the aerospace sector. We received 21 responses from leaders representing ten different companies. On the question of to what extent they had implemented deep learning solutions, four companies (referred to as "Laggards") scored significantly lower compared to the other six companies (the "Other Companies"). In this context, it should be mentioned that the survey did not measure the extent to which the companies had successfully implemented AI. The AI maturity of the Laggards and the Other Companies was self-assessed and not an absolute measure of their AI maturity. Nor was there any way we could fact-check to what extent the respondents had implemented AI.

Deep interviews

We deep interviewed twelve senior leaders, including the CEOs, from the executive teams of the Laggards based on the capability framework. In the in-depth interviews, we discussed the leaders' understanding of AI, how they strategically work to implement AI, the need to develop and implement an AI vision, and their and their colleagues' aptitude to change.

5.2. Paper 2: Abstraction, Mimesis and the Evolution of Deep Learning, Appendix 2 – *published in the journal AI & Society, May 2023*

To fully understand leadership capabilities required for an effective AI implementation, it is necessary to consider the nature of AI technology itself and its use. In a second study, we specifically investigate the role of abstraction in AI.

Abstraction is the process to derive general rules and concepts from individual facts or situations and from these build formal descriptions of the underlying facts or situations. "An abstraction" is the outcome of this process. Instead of accounting for every detail of each fact or situation, a higher abstraction level generalizes the relevant patterns that characterize a larger set. The opposite of abstraction is specification, which describes the process of breaking down general rules and concepts into concrete facts or situations (Ganascia 2015). Ganascia provides the following illustrative description of abstraction and the meaning of higher abstraction levels:

"*[…] figures such as triangles, spheres or pyramids are abstractions of shapes of objects, geometry is an abstraction of figures and algebraic geometry is an abstraction of geometry. In the same way, integers are abstractions of sets of objects, real numbers are abstraction of measures and associative rings – algebraic structures – are abstractions of integers and real numbers*" (Ganascia 2015)

In building deep learning algorithms, programmers typically use so-called deep learning software frameworks ("DLSF") – simply described as prepackaged programming tools. New DLSFs progressively encapsulate mathematical, statistical, and computational complexity and provide higher levels of abstraction (Bommasani 2021). Advancements have made it much easier for people to collaborate and build off each other's work allowing for a surge in technological innovation and progress (Bommasani 2021). In less than a decade, this development has made it possible to develop AI applications that previously required in-depth knowledge of computer science, statistics, and mathematics.

However, these new higher levels of abstraction pose new challenges for leaders of manufacturing companies. Manufacturing company leaders need to be aware of the implications of increased abstraction and be prepared to adapt to the changing landscape of AI technology to effectively support the widespread implementation of AI in their organization.

The second study involved collaboration with researchers in philosophy, computer science, IT, and business management, as well as an industry expert specializing in AI programming. The aim was to quantify the increase in deep learning abstraction, the rate of development, and its implications, using data from GitHub - a leading software repository. To achieve this, we searched for deep learning repositories on GitHub using keywords such as "*deeplearning*", "*deep learning*", and "*deep-learning*". We identified 605 403 repositories that fit these descriptions. The intention was to download and analyze a random sample amounting to half of these 605 403 repositories. To ensure a representative sample, we employed uniform sampling that selects a random subset of data from a larger dataset, with each data point having an equal probability of being chosen. After constructing the random sample and initiated the download process (that took over eight months to complete), some of the repositories of the sample had changed from public to private or been deleted. We therefore ended up with a random sample of 317 428 repositories. After removing all instances of forked repositories (i.e. repositories copied from other repositories) we ended up with a final dataset of 37 915 repositories.

From the downloaded repositories, we extracted information such as the creation date, last commit, programming language used, and number of lines of code. As most of the repositories used Python, we investigated the increase in the number of projects and the reduction in the number of lines of code in view of the initial release dates of important Python DLSFs.

We found that the number of deep learning projects increased significantly during the studied period, while the median number of lines of code used in the repositories decreased substantially. These findings were discussed from a technological advancements perspective, exploring the potential implications on AI models' competence, governance, security, bias, usability, and reliability.

We concluded that this process, which we call “abstraction explosion”, contributes to "ephemeralization," which is the ability to do more with less through technological advancement. Additionally, we discussed that the abstraction explosion has led to technological feedback loops that have enabled advancements within adjacent areas such as GPU development. This, in its turn, leads to the enablement of rapid paradigm shifts in technology.

The abstraction explosion also contributes to democratizing deep learning, making it easier for people to collaborate and build on each other's work. However, the abstraction explosion also makes it increasingly important to keep up with the development and to adopt timely levels of abstraction. Further, we found that the democratization of deep learning can lead to mimetic deadlocks and herd behavior which in its turn can lead to convergence towards the use of suboptimal solutions. The interplay between abstraction and mimesis also risks leading to a gradual reduction of nuance and complexity, contributing to a homogenized and streamlined perception of reality.

The democratization of deep learning increases the risk of fairness, privacy, and quality assurance problems as more people are able to use the higher abstraction levels. Non-experts often lack skills in quality control and reliability and organizations are not yet prepared on how to respond to AI failures. Incomprehensible error messages further exacerbate this issue, making it challenging to correct malfunctioning algorithms. As a consequence, external experts are often needed to solve these issues, but they are scarce and expensive to hire. Therefore, unfair algorithms can unintentionally (or intentionally) be unleashed due to the decreased explainability of deep learning and the lack of understanding of its workings or for which tasks they are suitable.

Finally, large corporations have outcompeted academia in the development of new abstraction levels, influencing and steering the democratization of deep learning. These companies control a significant amount of data shared on the internet and employ an increasing fraction of experts, influencing the progress of abstraction and providing access to experts in developing and using these abstractions.

This study highlights opportunities and challenges that leaders may face in adapting to the changing landscape of AI technology. It explores important concepts and implications related to the nature of AI itself and provides insights into the challenges and opportunities presented by widespread AI implementation, which can inform decision-making and strategy development for leaders of manufacturing companies.

5.3. Paper 3: AI Implementation and Capability Development in Manufacturing: An action research case, Appendix 3 – *accepted for publication at HICSS-57 (Hawaii International Conference on System Sciences), 2024*

My first two studies allowed me to gain at least some understanding of the leadership aspect associated with implementing AI in a manufacturing company and a more general understanding of the nature of AI itself and its potential societal implications. However, I also found it important to gain practical experience with the challenges a leader can face when implementing AI in a manufacturing company which has been called for in current research (Arinez, et al. 2020, Govindarajan and Immelt 2019, Davenport and Foutty 2018). The opportunity to gain practical experience showed itself when my team (the Team) was requested to assess a model from an external service provider specializing in detecting defects in images with limited labeled data (the Proprietary model). I viewed this as an opportunity to gain insights into the challenges and considerations that arise during an AI implementation project. Against this background, a case was set up in close collaboration between practitioners, researchers, and third-party experts and was designed to assess AI as a method to enhance the generation of insights and inform decision-making. The work resulted in an action research study where I gained experience in developing and implementing an advanced AI model. The aim with the study was to investigate the implications, considerations, and trade-offs of introducing AI into daily operations of a manufacturing company, leading up to the decision of whether to develop AI capabilities in-house or outsource them and the factors that influenced this decision. The case study focuses on the in-house development of an AI model for defect detection in X-rays of welds of aerospace components (the In-House Model).

In the current case, the GKN Aerospace (the Company) was increasing production of a critical component, the turbine exhaust case (TEC), and an increasing number of welds required inspection using X-ray. Three operators spent thousands of hours inspecting hundreds of thousands of images per year, and the process was expensive and prone to human error. In this case, the Company's senior leadership requested my digital innovation team (the Team) to create an in-house model that allowed for benchmarking against the

Proprietary model. The Team had limited experience of developing AI models in general and models for defect detection in particular and therefore needed to develop these capabilities. The Team collaborated with external experts in deep probabilistic programming (the Experts) and internal X-ray inspection operators (the Operators) who provided insights and recommendations based on their expertise. To further conceptualize the results, ensuring they were research-grounded and contributed to the academic discourse, I collaborated with Prof. Ulrika Lundh-Snis, which resulted in insights that enhanced the practical experiences gained in the manufacturing industry, making the results both practically and scientifically grounded.

Considering the Proprietary model's ability to operate with limited labeled data, and the fact that there was a fair amount of unlabeled data available, the Experts advised the Team to develop a model that could efficiently utilize unlabeled data without relying on extensive labeled data. With this in mind, they recommended the Team to construct a SS-VAE. To expedite the development process and meet the leadership's goal of making a decision soon, the Experts recommended the Team to adopt a publicly accessible model (the Baseline model) from https://pyro.ai/examples/ss-vae.html. This approach, they suggested, would provide an adequate benchmark for comparison purposes.

Through this research, I gained in-depth practical experience of the process of model selection, creation of training and validation datasets, adopting an AI model based on a publicly available model, as well as training of the same along with final model assessment. It also describes how decisions were made and addresses the organizational considerations and obstacles that needed to be overcome and, most importantly, how decisions and considerations were made. The Proprietary model was found to be superior to the In-house Model, both in terms of training time and accuracy. The Company therefore concluded that further work on the In-House Model therefore should be discontinued.

In the work of identifying and developing necessary capabilities of AI implementation, the hands-on work of developing a solution and dealing with the obstacles and considerations encountered along the way eventually proved to be more important than the solution itself and provided several learnings. These learnings include the need to acquire diverse skill sets and to strengthen the IT infrastructure and the IT department's capabilities for AI development.

The study also revealed broader concepts and insights that contribute to a more generalized understanding of AI capability development:

- The Company decided against developing an in-house solution due to challenges integrating the model with sub-images, long training time, and the superior performance of a proprietary model. This underscores the importance of evaluating costs and benefits in AI development, considering expertise, resources, and time. Leveraging existing AI solutions or outsourcing to external providers may be more efficient. However, overcoming such efforts can also enhance organizational capabilities and aid future decision-making on in-house development or outsourcing opportunities.
- Collaboration with Experts and Operators played an important role in overcoming challenges for the Team. This shows internal or external expertise is important for gaining knowledge, overcoming challenges, and advancing organizational capabilities in AI development. Collaboration is an essential component in building in-house AI capabilities, fostering open communication, knowledge sharing, and strengthening collaborative capabilities among employees, management, and external experts. This enhances creativity, innovation, and helps organizations overcome complex problems.
- The Team achieved progress by adopting an agile and experimental approach that emphasized flexibility and collaboration. Continuous feedback from stakeholders allowed the Team to refine the project design and enhance the In-house Model's performance iteratively.
- Subject matter expertise proved integral to the AI development process. The Operators' domain-specific knowledge was crucial in developing both the dataset and the model in a way that accurately detected defects. By leveraging subject matter experts' knowledge and expertise, organizations can develop AI models that are more accurate, effective, and useful, ultimately benefiting both workers and the organization as a whole.
- Involving internal subject matter experts in AI development can help address concerns about obsolescence and highlights the value of workers in the process. In this case, it became evident that neither the In-house nor Proprietary models could fully replace human capabilities. This experience indicates the importance of

> acknowledging the limitations of AI models and using them as tools to augment human labor rather than entirely replacing it. By integrating AI into the workforce, job satisfaction can be improved as workers can focus on higher-level tasks that require human expertise, rather than repetitive and mundane tasks. Emphasizing the collaborative relationship between AI models and humans allows organizations to foster a culture that values the contributions of both parties and maximizes the benefits of AI in the workforce.

Apart from providing me with practical experience from AI implementation, the study showed how capability development can be facilitated through hands-on experiences and collaboration among technical experts, business leaders, end-users, and researchers, as well as through the integration of AI models and human expertise. The implications we identified can support manufacturing companies in making informed decisions about sustained AI capability development. The study provides practical guidance on how to balance in-house development with external acquisition. While in-house development can provide control and potential competitive advantages, external acquisition can offer quick access to expertise. Companies must carefully consider factors such as cost, time, expertise, and long-term benefits. Even though in-house development can prove challenging, we argue that such efforts can strengthen organizational capabilities and enable informed decisions about future in-house development or outsourcing. The study also contributes to theory on AI implementation, confirming the need for a balanced evaluation of in-house versus outsourced solutions, considering costs, expertise, and performance. It emphasizes the importance of collaboration with internal and external stakeholders as well as researchers, agile and experimental methodologies, and the integration of human expertise in AI development. Furthermore, the study underscores the role of AI as a tool to augment rather than replace human labor, adding to the discourse on human-machine collaboration, organizational strategy, and AI capability development.

6. CONCLUSION

This thesis is about AI transformation in the manufacturing industry and the leadership challenges and implications that follow. The research highlights the key findings from three papers, addresses gaps in existing literature, and emphasizes the practical implications of implementing AI solutions. In the following chapter I present my research contributions, findings, discuss these in relation to current research and propose areas for further research.

6.1. Research contributions

There seems to be a consensus in the research literature across various disciplines that AI can benefit manufacturing companies. While there is research investigating the implementation of AI in manufacturing in isolated areas, the limited adoption suggests that implementing it company-wide may be challenging. Additionally, research within the field of general management indicates that AI transformations require support from business leaders and should begin at the leadership level (Fountaine, McCarthy and Saleh 2019, Govindarajan and Immelt 2019). This literature highlights the need for further research to help leaders understand how to best support an AI transformation.

Therefore, based on the personal need to develop knowledge in this area and the needs identified in the research literature, this research aims to contribute to the research on leadership implications and challenges for the effective widespread implementation of AI in manufacturing companies. The research uses a mixed methods approach, combining qualitative literature studies, surveys, interviews, and practical hands-on work with quantitative research.

The three research questions set out in this thesis have resulted in the generation of three papers. Due to the multidisciplinary nature of the research, they contribute to different bodies of knowledge.

The first article aims to answer the research question “What capabilities can benefit leaders who wish to implement AI in a manufacturing company, and how do these capabilities affect how leaders support an AI transformation?”. Its contribution is addressing the explicit need identified in the general management literature: understanding what capabilities are required from a manufacturing company leader to support the effective widespread implementation of AI.

The second article aims to answer what impact does the nature of AI technology the research question “What is the impact of the nature of AI technology efficiency, democratization and ethical concerns and how can the increasing abstraction in deep learning software frameworks reshape the relationship between technology and society?”. Here, the contribution concerns the rapid development of AI, which has made it possible to develop increasingly powerful AI applications but has also increased the risk of fairness, privacy, and quality assurance problems with AI models. Current research has identified that there is a constant and rapid development of new technologies and techniques and that this development affects leaders (Govindarajan and Immelt 2019, Davenport and Foutty 2018, Persaud 2021). This research aims to address these concerns and provide guidance for leaders on how to navigate them.

The third article aims to answer the research question “What are some key challenges and considerations in implementing an AI solution in the daily operations of a manufacturing company, how companies can adequately develop in-house AI capabilities and what factors influence their decision to continue developing these capabilities or outsource them?”. The contribution of this article concerns the understanding of the practical implications of an AI transformation. This research study aims to contribute to bringing clarity to leadership challenges and considerations, and the practical implications of implementing AI.

Together, and individually, the three papers all aim to contribute to the research on general management, information systems, work-integrated learning, the ethical implications of AI, manufacturing science and engineering, and innovation management.

6.2. Findings and propositions for further research

The capability framework presented in the first article provides a preliminary description of the attributes that may benefit manufacturing company leaders in their implementation of AI. However, it is not exhaustive and requires further refinement through additional research. Further investigation is needed to fully understand the leadership capabilities required for a successful AI transformation.

As more leaders engage in the widespread adoption of AI, it is important to encourage knowledge sharing and collaboration with the research community. However, some leaders may be hesitant to share their experiences due to fear of losing power or facing criticism for their failures, as noted by Ancona, Fountaine et al. It is crucial for researchers to establish a trust-building environment that fosters open communication and enables leaders to feel comfortable sharing their experiences.

In addition, researchers can play a critical role in mitigating these fears by emphasizing the importance of collaboration between academia and practitioners (Bommasani 2021). Given the rapid developments in AI, close collaboration between these two groups is essential for advancing the field.

In the second article, we discussed how the development within the field of AI has reshaped the relationship between technology and society in terms of efficiency, democratization and ethical concerns. The implications raised in the study are important for manufacturing company leaders who need to understand the consequences of increased abstraction and spread of deep learning to effectively support the widespread implementation of AI in their organization.

In this article, the GitHub results indicated that in the last few years, there has been a significant reduction in the number of lines of code used to build deep learning applications. The reduction in the number of lines of code is an indication of the ongoing advancements in AI technology and the increasing accessibility of AI development. The results imply that AI development is becoming more accessible and easier to use, requiring less specialized knowledge and expertise. Solutions that once required in-depth knowledge of computer science, statistics, and mathematics can now be developed by a wider range of individuals. We discussed how the ongoing advancements in AI technology, often referred to as ephemeralization, allow us to do "more and more with less and less until eventually, you can do everything with nothing" (Fuller 1938). These advancements are driving the democratization of AI, making it increasingly important for manufacturing companies to stay up-to-date with the latest methods and trends. Otherwise, they risk wasting time using obsolete methods and falling behind in the competitive landscape.

Recent developments within the field of AI confirm this trend. Examples include AI systems that generate art based on text prompts (Midjourney 2023) and predict the three-dimensional shape of proteins (Jumper, et al. 2021) or the introduction of large language models such as ChatGPT (OpenAI 2023).

However, in the second article we conclude that, while the adoption of AI technology may seem urgent, it is important for leaders to carefully consider the implications of implementing the latest AI methods. In many cases, we do not have a full understanding of how AI models make their decisions (Castelvecchi 2016), and are only informed of the final outcome (Holweg 2022). Additionally, there is a growing homogenization in the way

AI models are selected and used (Bommasani 2021), which could lead to the convergence of suboptimal solutions. Instead of finding the best way to solve a problem, developers may end up using the same algorithms and methods as their peers, without fully considering their potential drawbacks.

Additionally, the widespread use of AI risks increasing the number of cases of AI failures that many organizations are not yet strategically prepared to handle (Holweg 2022). The more widespread the implementation of AI becomes, the more instances we could have of such failures. It is possible to counter many drawbacks by hiring skilled staff responsible for developing and implementing AI. Optimally, when this staff has skills in standard processes for debugging and testing for quality control and reliability of AI models, they can correct these errors. However, recruiting and retaining these competencies could be a challenging endeavor.
Suppose the above is not considered and dealt with appropriately. In that case, leaders could initiate AI transformations without sufficient understanding of its implications, which could have fatal consequences. An illustrating example from the aerospace industry, when refitting the 737 Max with larger engines, instead of redesigning the airframe (body of the aircraft), Boeing relied on an AI system called AI Maneuvering Characteristics Augmentation System (MCAS). Unfortunately, the MCAS later showed to contribute to two plane crashes that killed a total of 346 people (Nicas, et al. 2019).

Such events risk adding to the reluctance and general fear of AI among leaders and workers that several researchers have identified (Babic, et al. 2020, Fountaine, McCarthy and Saleh 2019, Schepman and Rodway 2020). This fear can add to the resistance to change, pointed out as one of the main barriers to widespread adoption of AI (Govindarajan and Immelt 2019, Fountaine, McCarthy and Saleh 2019). Resisting change can be a good thing and can contribute to leaders not bringing too much change at the same time or changing things too quickly. Companies that push the transformation too hard and fail could risk destroying what works well, losing their credibility, and scaring off their best employees (Westerman, Soule and Eswaran 2019, Fitzgerald, et al. 2013). However, when the resistance to change is based on fear rather than informed decision-making, it can risk preventing a sustainable AI transformation. Ultimately, this could make companies take one step back from becoming more efficient rather than moving forward. Against this background, further research is needed to understand the implications of technological advancements within AI and the leadership implications that this could bring. Research within explainable AI, causality, informative error messages and ethical implications are important in this regard (Brown 1983, Ko 2014, Holweg 2022).

Through the practical work conducted in the third study, I was able to apply the knowledge gained from previous parts of my research and gain important insights into specific challenges and considerations that can only be gained through hands-on experience. The challenges discussed throughout the thesis, such as uninformative error messages and a scarcity of experts to correct errors, were encountered and overcome in this work. The experience also highlighted the difficulty in predicting required capabilities beforehand and emphasized the importance of co-creation and collaboration with internal and external stakeholders, recognizing the value of human expertise, and leveraging AI models as productivity tools that augment human labor. While it is possible to talk about the challenges and considerations of an AI transformation on a general level, this study exemplified what it could mean from a more practical standpoint. Through the study we identified a need for further research investigating the relationships between in-house and outsourced AI development, exploring how different industries, organizational sizes, or technological complexities influence the decision-making process. Additionally, studies examining human-AI collaboration across various sectors could provide insights into optimizing the blend of human expertise and AI, potentially leading to new models for organizational efficiency, innovation, workforce satisfaction, and capability development.

To some extent, the experiences from the practical work are anecdotal. However, they are experiences that can only be made when practically working with implementing AI. Discussions with my peers in other manufacturing companies have confirmed that they often face similar problems. I therefore also see a need of further research where manufacturing company leaders share their practical experiences implementing AI in manufacturing companies. Perhaps, when we have access to enough of these anecdotal stories, we can start to generalize across cases.

6.3. Transferability of the results to other industries

While parts of this research have been conducted based on the needs and requirements of the aerospace manufacturing industry, the results are transferable to other manufacturing industries, as well as other industries, such as transportation, mining, healthcare, energy, and agriculture. Leaders of these industries can use the insights from this research to inform their decisions on implementing an AI transformation, taking into account the potential consequences of moving too quickly or implementing AI without a full understanding of its implications.

By emphasizing the significance of the research in the broader context of AI transformation in the manufacturing industry, our study contributes valuable knowledge to the ongoing discussions and has the potential to inform future research and practical applications.

6.4. Concluding remarks

In this thesis, I have explored leadership challenges and implications of AI transformation in the manufacturing industry. The thesis aims to make significant contributions to our understanding of AI transformation in the manufacturing industry, particularly concerning leadership challenges and implications. By addressing key research questions, identifying gaps in the literature, sharing practical experiences, and emphasizing the importance of a multidisciplinary approach, the research intends to inform both future research and real-world applications. As the research has shown, implementing AI in the manufacturing industry presents numerous challenges that leaders must navigate. It requires that technical, organizational and societal aspects are taken into consideration.

The long-term implications of an AI transformation are difficult to predict, as American scientist and futurist Roy Amara noted, "*we tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run"* (Searls 2012). Furthermore, the historical underestimation of challenges related to AI indicates the complexity and unpredictable nature of integrating AI into our systems.

While incremental improvements in manufacturing processes through AI integration are undeniably valuable, there is potential for transformative change by reimagining the role of AI in this sector. To unlock this potential, I would argue that it is imperative to adopt a multidisciplinary approach that brings together researchers, industry practitioners, and experts from diverse fields such as philosophy, art, or other disciplines that can provide fresh insights and perspectives. By challenging the boundaries of conventional thinking and promoting an open dialogue among stakeholders from diverse areas of expertise, it becomes possible to identify novel ways to leverage AI that can transform the manufacturing landscape and create unprecedented value. It is our hope that this research will inspire further exploration into the transformative potential of AI within the manufacturing industry and beyond, leading to innovative solutions that can transform the way we work, create, and live.

7. References

Acerbi, Federica, Dai Andrew Forterre, and Marco Taisch. 2021. "Role of artificial intelligence in circular manufacturing: A systematic literature review." *IFAC-PapersOnLine* 54 (1): 367-372.

Adiga, S. 1993. "Object-oriented software: Relevance to manufacturing." In *Intelligent Manufacturing*, 25-37. Springer, Dordrecht.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction machines: the simple economics of artificial intelligence.* Harvard Business Press.

Ahmed, Mohiuddin, Raihan Seraj, and Syed Mohammed Shamsul Islam. 2020. "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation." *Electronics* 9 (8): 1295.

Ancona, Deborah. 2019. "Five Rules for Leading in a Digital World." *MIT Sloan Management Review* 61 (1): 1-4.

Angluin, Dana, and Carl H. Smith. 1983. "Inductive inference: Theory and methods." *ACM computing surveys (CSUR)* 15 (3): 237-269.

Arinez, Jorge F., Qing Chang, Robert X. Gao, Chengying Xu, and Jianjing Zhang. 2020. "Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook." *Journal of Manufacturing Science and Engineering* 142 (11).

Avison, David, Francis Lau, Michael Myers, and Peter Axel Nielsen. 1999. "Action Research." *Communications of the ACM.* 94-97.

Babic, Boris, Daniel L. Chen, Theodoros Evgeniou, and Anne-Laure Fayard. 2020. "A better way to onboard ai understand it as a tool to assist rather than replace people." *Harvard Business Review* 2-11.

Baskerville, Richard L., and A. Trevor Wood-Harper. 1996. "A critical perspective on action research as a method for information systems research." *Journal of Information Technology* 11 (3): 235-246.

Bassellier, Geneviève, Izak Benbasat, and Blaize Horner Reich. 2003. "The Influence of Business Managers’ IT Competence on Championing IT." *Information Systems Research* 14 (4): 317-336.

Bisong, Ekaba. 2023. Accessed May 05, 2023. https://ekababisong.org/assets/seminar_IEEE/activation-function.png.

Bommasani, R., Liang, P., Hudson, D., Adeli, E., Altman, R., Arora, S. Von Arx, S., Bernstein, M., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. Doumbouya, M. Durmus, E., Ermon, S., Etchemendy,. 2021. "On the Opportunities and Risks of Foundational Models." *Arxiv.* https://arxiv.org/abs/2108.07258.

Bornemark, Jonna. 2018. *Det omätbaras renässans : En uppgörelse med pedanternas världsherravälde.* Volante.

Bottou, Léon. 2010. "Large-Scale Machine Learning with Stochastic Gradient Descent." In *Proceedings of COMPSTAT'2010*, 177-186. Physica-Verlag HD.

Brock, Jürgen Kai-Uwe, and Florian von Wangenheim. 2019. "Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence." *California Management Review* 61 (4): 110-134.

Brown, P. 1983. "Error messages: the neglected area of the man/machine interface." *Communications of the ACM* 26 (4): 246-249.

Brydon-Miller, Mary, Davydd Greenwood, and Patricia Maguire. 2003. "Why Action Research?" *Action Research* 9-28.

Bryman, Alan. 2015. *Social Research Methods.* Oxford University Press.

Candy, Linda. 2006. *Practice Based Research: A Guide.* CCS-Report, Sydney: Creativity & Cognition Studios, University of Technology.

Castelvecchi, D. 2016. "Can we open the black box of AI?" *Nature News* 538 (7623): 20.

Chang, Daniel T. 2021. "Probabilistic deep learning with probabilistic neural networks and deep probabilistic models." *arXiv preprint arXiv:2106.00120.*

Chen, Tin Chih Toly, and Yi Chi Wang. 2022. "Artificial Intelligence in Manufacturing." *SpringerBriefs in Applied Sciences and Technology* 13-35.

Chien, Chen-Fu, Stéphane Dauzère-Pérès, Woonghee Tim Huh, and Young Jae. 2020. "Artificial intelligence in manufacturing andlogistics systems: algorithms, applications, and case studies." *International Journal of Production Research* 58 (9).

Chollet, François. 2017. *Twitter.* Accessed Sep 15, 2022. https://twitter.com/fchollet/status/942733414788190209.

Cioffi et al., Raffaele. 2020. "Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions." *Sustainability* 12 (2): 492.

Collis, Jill, and Roger Hussey. 2014. *Business research: A practical guide for undergraduate and postgraduate students.* Palgrave Macmillan.

Cowan, Robin. 2001. "Expert Systems: Aspects of and Limitations to the Codability of Knowledge." *MERIT- Infonomics Research Memorandum series* .

Crevier, Daniel. 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence.* BasicBooks.

Dada, Emmanuel Gbenga, Joseph Stephen Bassi, Haruna Chiroma, Shafi'i Muhammad Abdulhamid, Adebayo Olusola Adetunmbi, and Opeyemi Emmanuel Ajibuwa. 2019. "Machine learning for email spam filtering: review, approaches and open research problems." *Heliyon* 5 (6): e01802.

Dahlmeier, Daniel. 2017. "On the Challenges of Translating NLP Research into Commercial Products." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).*

Davenport, Thomas J., and Janet Foutty. 2018. *AI-Driven Leadership.* August 10. Accessed February 12, 2020. https://sloanreview.mit.edu/article/ai-driven-leadership/.

Deveraux, Doug. 2019. "Smaller Manufacturers Get Lean with Artificial Intelligence." *National Institute of Standards and Technology, Manufacturing Innovation Blog.*

Dombrowski, Uwe, and Tim Mielke. 2013. "Lean Leadership fundamental principles and their application." *Procedia CIRP*.

Economist. 2022. *What if all workers wrote software, not just the geek elite?* January 29. Accessed February 02, 2022. https://www.economist.com/business/2022/01/29/what-if-all-workers-wrote-software-not-just-the-geek-elite?giftId=e827aa1a-7c4e-4ca0-85c3-f9d2953ab719.

Fahle, Simon, Christopher Prinz, and Bernd Kuhlenkötter. 2020. "Systematic review on machine learning (ML) methods for manufacturing processes–Identifying artificial intelligence (AI) methods for field application." *Procedia CIRP 93* 413-418.

Felin, Teppo, Nicolai J. Foss, Koen H. Heimeriks, and Tammy L. Madsen. 2012. "Microfoundations of Routines and Capabilities: Individuals, Processes, and Structures." *Journal of Management Studies* 49 (8): 1351-1374.

Fitzgerald, Michael, Nina Kruschwitz, Didier Bonnet, and Michael Welch. 2013. "Embracing Digital Technology - a new strategic imperative." *MIT Sloan Management Review.*

Fortuna, Paweł, and Oleg Gorbaniuk. 2022. "What Is Behind the Buzzword for Experts and Laymen: Representation of "Artificial Intelligence" in the IT-Professionals' and Non-Professionals' Minds." *Europe's Journal of Psychology* 18 (2): 207-218.

Fountaine, Tim, Brian McCarthy, and Tamim Saleh. 2019. "Building the AI-Powered Organization." *Harvard Business Review* 97 (4): 62-73.

Fuller, B. 1938. *Nine Chains to the Moon, Anchor Books, 1938, 1973.* Anchor Books.

Ganascia, J-G. 2015. "Abstraction of levels of abstraction." *Journal of Experimental & Theoretical Artificial Intelligence* 23-35.

GE Aviation. 2022. *TrueChoice Flight Hour.* Accessed Sep 26, 2022. https://www.geaviation.com/services/true-choice/flight-hour.

2022. *GKN.* www.gkn.com.

Glorot, Xavier, Antoine Bordes, and Yoshua Bengio. 2011. "Deep sparse rectifier neural networks." *Proceedings of the fourteenth international conference on artificial intelligence and statistics.* JMLR Workshop and Conference Proceedings.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* The MIT Press.

Govindarajan, Vijay, and Jeffrey R Immelt. 2019. "The Only Way Manufacturers Can Survive." *MIT Sloan management review* 60: 24-33.

Grosan, C., and A. Abraham. 2011. "Rule-based expert systems." *Intelligent systems: A modern approach* 149-185.

Hady, Mohamed Farouk Abdel, and Friedhelm Schwenker. 2013. "Semi-supervised learning." In *Handbook on Neural Information Processing* , 215-239.

Hansen, Emil Blixt, and Simon Bøgh. 2021. "Artificial intelligence and internet of things in small and medium-sized enterprises: A survey." *Journal of Manufacturing Systems* 58: 362-372.

Hinton, G.E., and R.R. Salakhutdinov. 2006. "Reducing the Dimensionality of Data with Neural Networks." *Science* 313: 504-507.

Holmström, Jonny. 2022. "From AI to digital transformation: The AI readiness framework." *Business Horizons* 65 (3): 329-339.

Holweg, M., Younger, R., Wen, Y. 2022. "The Reputational Risks of AI." *California Management Review* 64 (2).

Humboldt-Universität. 2022. *Ueber den Schachspieler des Herrn von Kempelen und dessen Nachbildung.* Accessed Sep 09, 2022. This thesis investigates, from an interdisciplinary point of view, some of the challenges associated with introducing AI in aerospace manufacturing. .

Humlum, Anders, and Bjørn Bjørnsson Meyer. 2020. "Artificial Intelligence and College Majors." July 29. Accessed May 09, 2023. https://rockwoolfonden.s3.eu-central-1.amazonaws.com/wp-content/uploads/2022/08/RFF-Study-Paper-172-Artificial-Intelligence-and-College-Majors.pdf?download=true.

Hunter, Richard. 2015. *Apollonius of Rhodes: Argonautica Book IV.* Cambridge University Press.

Islind, Anna Sigríður. 2018. *PhD Thesis: Platformization, Co-Designing Digital Platforms in Practice.* University West.

Jamwal et al., Anbesh. 2022. "Machine learning applications for sustainable manufacturing: A bibliometric-based review for future research." *Journal of Enterprise Information Management* 35 (2): 566-596.

JavaPoint. 2023. *Gradient Descent in Machine Learning.* Accessed May 05, 2023. https://www.javatpoint.com/gradient-descent-in-machine-learning.

Johnson, Deborah G., and Mario Verdicchio. 2017. "AI Anxiety." *Journal of the Association for Information Science & Technology.*

Johnson-Laird, Phil. 2010. "Deductive reasoning." *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (1): 8-17.

Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 2021. "Highly accurate protein structure prediction with AlphaFold." *Nature 596* 583-589.

Järvinen, Pertti. 2001. *On Research Methods.* Opinpajan kirja.

Kaplan, Andreas. 2022. *Artificial Intelligence, Business and Civilization - Our Fate Made in Machines.* Routledge.

Kettner, Matthias. 1991. "Peirce’s notion of abduction and psychoanalytic interpretation." In *Semiotic Perspectives on Clinical Theory and Practice*, 163-179.

Kim et. al, Sung Wook,. 2022. "Recent advances of artificial intelligence in manufacturing industrial sectors: A review." *International Journal of Precision Engineering and Manufacturing* 1-19.

Kingma, Diederik, and Max Welling. 2014. "Auto-Encoding Variational Bayes." *arXiv preprint arXiv:1312.6114* .

Ko, A. J.,. 2014. *Programming languages are the least usable, but most powerful human-computer interfaces ever invented.* March 26. Accessed February 14, 2022. https://medium.com/bits-and-behavior/programming-languages-are-the-least-usable-but-most-powerful-human-computer-interfaces-ever-invented-7509348dedc.

Kolbjørnsrud, Vegard, Richard Amico, and and Robert J. Thomas. 2017. "Partnering with AI: how organizations can win over skeptical managers." *Strategy & Leadership* 45 (1): 37-43.

Krizhevsky, Alex, llya Sutskever, and Geoffrey Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems 25* 1097-1105.

LeCun, Yann. 2022. "A Path Towards Autonomous Machine Intelligence." *Open Review* 62.

LeCun, Yann, Leon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 2278-2324.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep learning." *Nature* (521): 436-444.

Lee, J., T. Suh, D. Roy, and M. Baucus. 2019. "Emerging technology and business model innovation: the case of artificial intelligence." *Journal of Open Innovation: Technology, Market, and Complexity* 5 (3): 44.

Lee, Jay,, Behrad Bagheri, and Hung-An Kao. 2015. "A cyber-physical systems architecture for industry 4.0-based manufacturing systems." *Manufacturing letters* 3: 18-23.

Legg, Shane, and Marcus Hutter. 2007. "A Collection of Definitions of Intelligence." *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms : Proceedings of the AGI Workshop 2006.* IOS Press. 17-24.

Lim, Seung-Hwan, Steven R. Young, and Robert M. Patton. 2016. "An analysis of image storage systems for scalable training of deep neural networks." *system* 5 (7): 11.

Lwakatare, Lucy Ellen, Aiswarya Raj, Ivica Crnkovic, and Jan Bosch. 2020. "Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions." *Information and Software Technology* 127.

Maheshwari, Anega, Priyanka Mitra, and Bhavna Sharma. 2022. "Autoencoder: Issues, Challenges and Future Prospect." *Recent Innovations in Mechanical Engineering.* Springer. 257-266.

Manyika, James. 2022. "Getting AI right: Introductory notes on AI & society." In *Daedalus*, 5-27.

Marcus, Gary. 2018. "Deep Learning: A Critical Appraisal." Accessed February 14, 2022. https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf.

MarketsandMarkets. 2022. "Artificial Intelligence in Manufacturing Market by Offering (Hardware, Software, and Services), Industry, Application, Technology, (Machine Learning, Natural Language Processing, Context-aware Computing, Computer Vision), & Region (2022-2027)." April. Accessed May 09, 2023. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-manufacturing-market-72679105.html?gclid=CjwKCAjw3ueiBhBmEiwA4BhspIFSNmHQlY9AxCyPol-TRuLt7STK_IEnhZQEDZqcqOTifLFQQVwVzxoC1tcQAvD_BwE.

Mathiassen, Lars, Mike Chiasson, and Matt Germonprez. 2012. "Style Composition in Action Research Publication." *MIS Quarterly* 36 (2): 347-363.

McCarthy, John. 2007. *What Is Artificial Intelligence?* Technical report, Stanford University. http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.

McCorduck, Pamela. 2004. *Machines Who Think.* A K Peters/CRC Press.

McKay, Judy, and Peter Marshall. 2001. "The dual imperatives of action research." *Information Technology & People* 46-59.

McKinsey, & Company. 2022. *Capturing the true value of Industry 4.0.* April 13. Accessed May 09, 2023. https://www.mckinsey.com/capabilities/operations/our-insights/capturing-the-true-value-of-industry-four-point-zero.

—. 2022. "Value creation in the metaverse." June. Accessed May 09, 2023. https://www.mckinsey.com/~/media/mckinsey/business%20functions/marketing%20and%20sales/our%20insights/value%20creation%20in%20the%20metaverse/Value-creation-in-the-metaverse.pdf.

Midjourney. 2023. *Midjourney.* Accessed May 09, 2023. https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F.

Monett, Dagmar, Collin Lewis, and Kristinn R. Thórisson. 2020. "Introduction to the JAGI Special Issue “On Defining Artificial Intelligence” —Commentaries and Author’s Response." *Journal of Artificial General Intelligence* 1-100.

Newquist, HP. 2020. *The Brain Makers: Genius, Ego, And Greed in the Quest For Machines That Think.* The Relayer Group.

Ngwenyama, Ojelanki, and Jacob Nørbjerg. 2010. "Software process improvement with weak management support: an analysis of the dynamics of intra-organizational alliances in IS change initiatives." *European Journal of Information Systems* 19: 303-319.

Nicas, Jack, Natalie Kitroeff, David Gelles, and James Glanz. 2019. *Boeing Built Deadly Assumptions Into 737 Max, Blind to a Late Design Change.* June 1. Accessed December 05, 2022. https://www.nytimes.com/2019/06/01/business/boeing-737-max-crash.html.

Nicolini, Davide. 2012. *Practice Theory, Work, and Organization. An Introduction* . Oxford University Press.

Nwankpa, Chigozie Enyinna, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. 2018. "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning." *arXiv preprint arXiv:1811.03378* .

O Reilly, Charles A., and Michael L. Tushman. 2004. "The ambidextrous organization." *Harvard business review* 82 (4): 74-83.

O’Shea, Keiron, and Ryan Nash. 2015. "An Introduction to Convolutional Neural Networks." *arXiv preprint arXiv:1511.08458.*

Okoli, Chitu, and Kira Schabram. 2010. "A guide to conducting a systematic literature review of information systems research." *Social Science Research Network* 10 (26).

OpenAI. 2023. *Introducing ChatGPT.* openAI.com. Accessed March 15, 2023. https://openai.com/blog/chatgpt.

Ouali, Yassine, Céline Hudelot, and Myriam Tami. 2020. "An overview of deep semi-supervised learning." *arXiv preprint arXiv:2006.05278.*

Palaver, Wolfgang. 2013. *René Girard's Mimetic Theory (Studies in Violence, Mimesis, and Culture (SVMC)).* Michigan State University Press.

Palensky, Peter, Dietmar Bruckner, Anna Tmej, and Tobias Deutsch. 2008. "Paradox in AI – AI 2.0: The Way to Machine Consciousness." *International Conference on IT Revolutions.* 194-215.

Pan, Yunhe. 2016. "Heading toward Artificial Intelligence 2.0." *Engineering* 2 (4): 409-413.

Pandl, Konstantin D., Scott Thiebes, Manuel Schmidt-Kraepelin, and Ali Sunyaev. 2020. "On the Convergence of Artificial Intelligence and Distributed Ledger Technology: A Scoping Review and Future Research Agenda." *IEEE Access* 8: 57075-57095.

Persaud, Ajax. 2021. "Key competencies for big data analytics professions: a multimethod study." *Information Technology & People* 34 (1): 178-203.

Pin, Wang, Fan En, and Wang Peng. 2021. "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning." *Pattern Recognition Letters* 141: 61-67.

Ramachandran, Prajit, Barret Zoph, and Quoc V. Le. 2017. "Searching for activation functions." *arXiv preprint arXiv:1710.05941.*

Rattanaprateep, Chuleerat, and Suphamit Chittayasothorn. 2006. "A frame-based objectrelational database expert system architecture and implementation." *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases.*

Reason, Peter, and Hilary Bradbury. 2001. *The SAGE Handbook of Action Research, Participative Inquiry and Practice.* Sage.

Russell, Stuart, and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach.* Pearson Education.

Santa Barbara Museum of Art. 2022. *Artificial Intelligence and Art: The Case of Harold Cohen.* Accessed Sep 06, 2022. https://www.sbma.net/events/underhillai.

Saran, Aynur. 2018. "A Paradigmatic Frame in the Definition of Leadership: “Transformational Leadership”." In *Chaos, Complexity and Leadership 2016 4*, 175-184. Springer International Publishing.

Schaffer, Simon. 1999. *Enlightened Automata in The Sciences in Enlightened Europe.* University of Chicago Press.

Schepman, Astrid, and Paul Rodway. 2020. "Initial validation of the general attitudes towards Artificial Intelligence Scale." *Computers in Human Behavior Reports* 1: 100014.

Schön, Donald A. 2016. *The Reflective Practitioner, How Professionals Think in Action.* Routledge.

Scrivener, Stephen. 2002. "The art object does not embody a form of knowledge."

Sculley, D., Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Franc¸ois Crespo, and Dan Dennison. 2015. "Hidden Technical Debt in Machine Learning Systems." *Advances in Neural Information Processing Systems 28.*

Searls, Doc. 2012. *The intention economy: when customers take charge.* Harvard Business Press.

Sivakumar, S., Soumya Ranjan Nayak, S. Vidyanandini, J. Ashok Kumar, and G. Palai. 2018. "An empirical study of supervised learning methods for breast cancer diseases." *Optik* 175: 105-114.

Soldatos, John, and Dimosthenis Kyriazis. 2020. *Trusted Artificial Intelligence in Manufacturing; Trusted Artificial Intelligence in Manufacturing: A Review of the Emerging Wave of Ethical and Human Centric AI Technologies for Smart Production; A Review of the Emerging Wave of Ethical and Human Centric.*

Sung, Kim Wook, Jun Ho Kong, Sang Won Lee, and Seungchul Lee. 2022. "Recent Advances of Artificial Intelligence in Manufacturing Industrial Sectors: A Review." *International Journal of Precision Engineering and Manufacturing* 23: 111-129.

Teesce, David. 2014. "The Foundations of Enterprise Performance: Dynamic and Ordinary Capabilities in an (Economic) Theory of Firms." *The Academy of Management Perspectives* 28 (4): 328–352.

Tran, Dustin. 2020. *Probabilistic Programming for Deep Learning.* Columbia University.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* (236): 433-460.

Van De Ven, Andrew H. 2007. *Engaged Scholarship.* Oxford University Press on Demand.

Wang et al., Baicun. 2021. "Smart manufacturing and intelligent manufacturing: A comparative review." *Engineering* 738-757.

Wang, Y., H. S. Ma, J. H. Yang, and K. S. Wang. 2017. "Industry 4.0: a way from mass customization to mass personalization production." *Advances in manufacturing* 5: 311-320.

Webster, Jane, and Richard T. Watson. 2002. "Analyzing the past to prepare for the future: Writing a literature review." *MIS quarterly* xiii-xxiii.

Weill, P., Apel, T., Woerner, S. L., Banner, J. S. 2019. "It Pays to Have a Digitally Savvy Board." *MIT Sloan Management Review* 60 (3): 41-45.

Westerman, George, Deborah Soule, and Anand Eswaran. 2019. "Building Digital-Ready Culture in Traditional Organizations." *MIT Sloan Management Review* 59-68.

Xiaojin, Zhu, and Goldberg Andrew. 2009. "Introduction to Semi-Supervised Learning." *Synthesis Lectures on Artificial Intelligence and Machine Learning* 3: 1-130.

Yang, Xin, Kong Lingshuang, Shi Liu, and et al. 2018. "Machine Learning and Deep Learning Methods for Cybersecurity." *IEEE Access.*

Yann, LeCun, Cortes Corinna, and Christopher Bruges. 2022. *The MNIST Database of Handwritten Digits.* Accessed Sep 16, 2022. http://yann.lecun.com/exdb/mnist/.

Zeba, Gordana, Marina Dabic, Mirjana Čičak, and Tugrul Daim. 2021. "Technology mining: Artificial intelligence in manufacturing." *Technological Forecasting and Social Change* 171.

APPENDIX 1

The 26th Biennial Nordic Academy of Management Conference
Örebro, Sweden, August 24-26, 2022

An AI leadership competency framework for manufacturing companies: An interdisciplinary and mixed methods approach

Eklöf, Jon[a]; Hamelryck, Thomas[b] and Lundh Snis, Ulrika[c]

[a] University of Copenhagen – Department of Computer Science, GKN Aerospace, [b] University of Copenhagen – Department of Computer Science, [c] University West – Department of Business Economics & IT

Summary
In this study, we create and assess a framework for leaders of manufacturing companies that wish to introduce AI in their businesses. In addition, we investigate how differences in leaders' ability to apply the framework affect their capacity to deploy AI.

Keywords or phrases: AI leadership; AI & Manufacturing; AI management; How to implement AI; AI leadership competences

Track: 6.2 Digitalization: Digital Tools and Organizational Transformation

Introduction

While some researchers question AI's current direction (Marcus 2018), others argue that AI has the potential to disrupt the manufacturing industry as we know it and that manufacturing companies that do not commit fully to an AI transformation will be outcompeted by those who can offer new data-driven services (Govindarajan 2019). An AI transformation requires support of the business leadership and must begin at leadership level (Brock 2019) (Fountaine 2019). However, it can be difficult for leaders to understand how to best support an AI transformation. Different views are spread across different research disciplines. Neither are there many "AI-driven" manufacturing companies to benchmark against (Govindarajan 2019) (Davenport 2018) (Teesce 2014). Therefore, researchers have pointed out the need of a framework describing what abilities and routines leaders of manufacturing companies need to develop and establish to deploy AI (Felin 2012), (Govindarajan 2019) (Davenport 2018).

Objective

The aim of this study is to establish a leadership competency framework to aid leaders of manufacturing companies that wish to introduce AI in their business. In addition, we investigate how differences in leaders' ability to apply the framework affect their capacity to deploy AI.

AI, Machine Learning and Deep Learning

Our study indicated that many leaders do not understand the difference between artificial intelligence (AI), machine learning and deep learning. Below, we briefly introduce these concepts.

John McCarthy, one of the founders of the term AI, defines it as the science and engineering of intelligent machines that mimic cognitive functions associated with the human mind, i.e. the ability to sense, reason, act, learn and adapt (McCarthy 2007).

Machine learning is a sub-field of AI that describes algorithms that learn to perform certain tasks based on training data (Mitchell 1997). Algorithms able to perform certain tasks without requiring training data, for example rule based decision trees, fall under the AI definition, but are not considered machine learning - because there is no learning involved. Within machine learning there is what is called traditional machine learning and deep learning. Traditional machine learning relies on established statistical methods such as linear regression or support vector machines to perform its tasks. Deep learning uses neural networks (algorithms loosely inspired by the structure of biological neural networks such as the human brain) with many layers to make its predictions (Marcus 2018). Consequently, all deep learning methods are machine learning methods, but not vice versa. While a simple linear regression and the most advanced deep learning algorithm both fall within the AI and machine learning definitions, it goes without saying that the deep learning model has a higher level of "intelligence" compared to the linear regression. The "intelligence" of AI is thereby not binary, i.e. intelligent or not, but rather falls somewhere on a sliding scale.

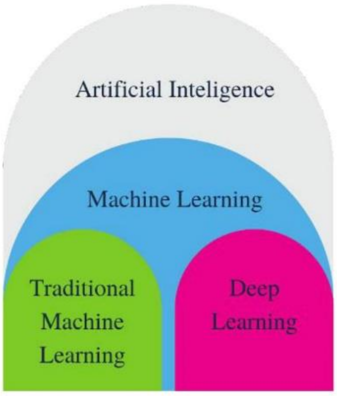

Figure 1. The relationship between AI, machine learning, traditional machine learning and deep learning. The figure shows that machine learning is a subfield of AI while traditional machine learning and deep learning, both are subfields of machine learning.

Methods

In this study, we have used a mixed methods approach, combining a literature review, a survey and deep interviews.

Literature review

In the literature review, we identified and analyzed relevant research on AI competence, leadership and management. The literature review covered 20 journals with the highest impact scores within general management (the study of the techniques, practices, or sciences related to managing a company), innovation management (the study of the process of managing innovative ideas) and information systems (the study of the interaction between hardware, software, users and business processes) respectively. We reviewed articles that covered abilities and routines required from leaders that wish to introduce AI in their business. We searched for articles using the following keywords: "*Artificial Intelligence, leadership*", "*Artificial Intelligence, management*", "*Artificial Intelligence leaders*", "*How to implement Artificial Intelligence*", "*AI, leadership*", "*AI, management*", "*AI leaders*", "*How to implement AI*" "*Leaders, digital transformation*", "*Management, digital transformation*". We also included a search on Google scholar with the same keywords for articles covering these topics but not included in the journals with the highest impact scores. In total, we reviewed 57 articles - 3 articles from innovation management journals, 12 articles from information systems journals and 42 articles from general management journals. The literature review refers to 33 of these 57 articles from 17 different journals where the majority of the articles are from general management journals. We did not include references to the other 34 articles that we reviewed since they cover other topics than what we investigate in this article. We have attached a detailed list of the articles ultimately included in the literature review, Appendix 1.

Survey

Much of the research on AI leadership focuses on what characterizes leaders that successfully implement AI. Rather than focusing on the success stories, we wanted to understand if there is a correlation between companies with low AI maturity and their leaders' lack of the constituents of the framework. To identify companies that struggle with the deployment of AI, based on two AI maturity models proposed by Alsheiabni et al. (Alsheiabni 2019) and Saari et al. (Saari 2019), respectively, we created a web-based survey to assess the AI maturity of companies. To avoid the general confusion about the meaning of AI, the survey focused on deep learning. In the interviews that followed (as well as in parts of this article where the arguments have a more general application) the discussions also included the term "AI" – without specifying a certain technology.

In the survey, we asked the respondents to rate, on a scale from 1-5 (where 1 was the lowest score and 5 the highest), if they had deployed deep learning solutions, if they have a strategy for using deep learning, if they have a data strategy, if they have the roles required to develop and deploy deep learning solutions, if they have an ecosystem of external service providers, if they have access to real time data and if their leadership is committed to the use of deep learning and AI. We distributed the survey to senior executives within general management, IT, engineering, HR and R&D at 11 major high tech manufacturing companies. We received 21 responses from leaders representing 10 different companies. On the question to what extent they had deployed deep learning solutions, four companies (hereinafter referred to as "Laggards") scored significantly lower compared to the other six companies (the "Other Companies").

Deep interviews

We deep interviewed twelve senior leaders, including the CEOs, from the executive teams of the Laggards based on the competency framework. Instead of following linear steps in the interview process, the interviews were iterative. This enabled us to incorporate what we learned at one point in the research into the remainder of the research and to confront the leaders with statements made by other leaders in their teams. In the deep interviews, we discussed the leaders' understanding of AI, how they strategically work to deploy AI, the need to develop and deploy an AI vision as well as their and their colleagues' aptitude to change.

Table 1: Methods summary describing how many articles from how many journals we reviewed in the literature review, how many responses we received in the survey from how many companies and how many leaders we interviewed from how many companies.

Literature Review:	33 articles	17 journals
Survey:	21 responses	10 companies
Interview:	12 interviewees	4 companies

Results

Literature Review

During the literature review we found that, even if there is no consensus in the literature on what characterizes an "AI-leader" and hence no playbook to adhere to, there are a number of recurring themes of abilities and routines that characterize leaders that successfully deploy AI. Based on these themes, we propose an AI-leadership competency framework. In summary, leaders need to be "learners", proficient strategists, provide and communicate a compelling vision and have strong social skills. Finally, it is necessary that they be prepared to change both how they lead the companies and how their companies work.

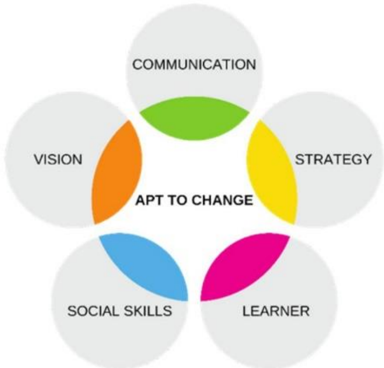

Figure 2: Model of AI-leadership competency framework setting out the different abilities and routines required from leaders that wish to implement AI in their companies.

Learner

Fountaine et al. claim that a failed implementation of AI often is due to the lack of fundamental understanding of AI among senior executives (Fountaine 2019). AI is not one single technology, but many, each with its own application areas, requirements, strengths, and limitations (Davenport 2018). Leaders need a basic understanding of the technologies to successfully engage in its deployment. They need it, both when planning internally and when dealing with external experts (Pyle 2015). Further, within the field of AI, there is a constant and rapid development of new technologies and techniques. Leaders need to commit to learning as the field grows (Davenport 2018) (Govindarajan 2019) (Weill 2019) (Geneviève 2003) (Persaud 2020).

Leaders need to understand the existence of digital data, how to collect and convert it in a robust and reliable way, its management, and its analysis. In addition, they need at least basic knowledge about AI model building as well as basic knowledge and understanding of modern programming languages and modern IT architecture (Brinch 2020). When the amount of data grows, data security skills become vital (Brock 2019) (Davenport 2018).

Understanding AI and its requirements enables leaders to create realistic understandings of what AI can do for the business and to assess what AI technologies are most critical to their organizations' success (Brock 2019) (Davenport 2018) (Ng 2016). It also enables them to plan for full-scale implementation at the beginning of each project (Davenport 2018). When a leader starts learning about AI, it often facilitates collective learning in the organization and provides the leader with an opportunity to demonstrate its commitment to explore AI (Fountaine 2019).

There is research indicating that business leaders' learning of AI provides a financial benefit. For example, according to Weill et al., companies with three or more board members with knowledge and understanding of digital technologies – including AI – can expect 17% higher profit margins than those with two or fewer, 38% higher revenue growth, 34% higher return on assets, and 34% higher market cap growth (Weill 2019).

Strategist

Learning about AI technologies and their requirements is not enough. Strategic abilities are also important. Leaders with abilities within digital strategy and digital business development derive more value from AI compared to leaders with weaker skills (Brock 2019) (Weill 2019). Leaders must understand how to organize for AI deployment, how to create digital strategies such as data strategies and AI-strategies and how to run AI projects using agile methods rooted in a culture of experimentation and creativity (Sousa 2019) (Mukherjee 2020) (Steiber 2020) (Davenport 2018).

Organizing an AI-function

Many organizations have overlooked digital transformation for so long that their leaders do not know where to begin (Kane 2020). One of the first moves is to decide how to organize the team responsible for developing and deploying AI solutions. Govindarajan and Immelt and Teece note that it can be tempting to outsource the creation of AI capabilities or place them under traditional technology functions such as IT. However, outsourcing AI capabilities means that you lose the competitive advantage that AI solutions can provide, only getting access to technology and competence available on the open market (Govindarajan 2019) (Teesce 2014). Therefore, according to Teece, outsourcing only works in weak competitive environments (Teesce 2014). Further, organizing an AI function under a traditional technology function such as IT puts the AI function at risk of succumbing under the bureaucratic processes and best practices that often characterizes traditional functions (Denning 2020). Additionally, traditional functions are often not used to the cross-functional ways of working required for successfully deploying AI on a wider scale. Only when the AI function is located outside the traditional functions will it find ways, or be allowed, to disrupt the business (Govindarajan 2019).

Data strategy

All AI technologies require data to function. Therefore, a data strategy (a strategy on how to collect, store, process and share the data) is crucial since a large part of the work in implementing any AI solution is to prepare the data for the models (Pyle 2015). A data strategy also encourages sharing of data. This is important since departments often hoard information and politicize access to it (Pyle 2015). Data must be collected, stored and processed to a form and format that suits the relevant AI model. Only when these steps are completed is it possible to apply an AI model. Rogati has illustrated this in a model that she calls the AI hierarchy of needs pyramid (Rogati 2017). Leaders need to be able to create a data strategy that lives up to these requirements and encourages sharing of data (Rogati 2017).

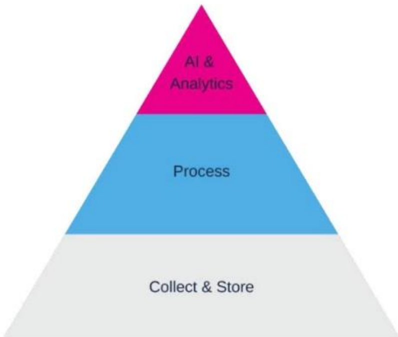

Figure 3: AI hierarchy of needs. The model by Rogati illustrates how data must be collected, stored and processed to a form and format that suits the relevant AI model before it is possible to deploy an AI model.

AI project management

Efficient AI project management is critical to quickly scale up AI projects to full production status (Davenport 2018). When managing AI projects, leaders need to abandon the mindset that an idea needs to be fully developed before it is deployed (Fountaine 2019) (Govindarajan 2019). Instead, they should promote ways of working that are agile, experimental, and adaptable (Westerman 2019) (Steiber 2020) (Davenport 2018).

There are different proposals on how to do this. Thomke suggests an order where companies run thousands of small experiments per year. Even if only 10% of the experiments pays off, this still translates into “a significant number of successes, which, in turn, diminish the financial and emotional costs of the failures” (Thomke 2020). Ancona proposes that leaders build flexible teams able to collaborate internally and link to knowledge, resources, and external innovation partners (Ancona 2019). Ancona’s proposal is in line with the conclusions of Kane et al. and Westerman et al. who argue that leaders need to give people high levels of discretion about what to do rather than relying on formally structured coordination and policies (Kane 2020) (Westerman 2019). Or, as Mukherjee puts it, “you can’t lead for creativity by seeking uniformity” (Mukherjee 2020).

However, the statements by Thomke, Ancona, Kane et al., Westerman et al. and Mukherjee stand in contrast to those by Obwegeser et al. and Fountaine et al., who promote a top down approach. Obwegeser proposes creating an efficient and transparent pipeline for evaluation and prioritization of AI initiatives and establishing metrics for each AI initiative (Obwegeser 2020). Fountaine et al. mean that leaders should control all AI projects and stop projects that are difficult to implement or require more than a year to launch since this could harm both current and future AI projects (Fountaine 2019). Sousa and Rocha suggest that leaders should be careful giving every person in the organization free reign to do whatever they want. Instead, they should balance feasibility, time investment, and value of AI initiatives (Sousa 2019).

Social skills

Determining which AI projects that benefit the business the most, identifying and harvesting associated data, is not something business leaders can or even should do on their own. More than traditional technologies, deploying AI requires teamwork. Leaders need the ability to build collaborative internal and external networks and strategic alliances, and to work cross functionally across all levels (Sousa 2019) (Reck 2019) (Mukherjee 2020). They need to orchestrate collaborative, self-organizing teams that establish priorities, determine the implications for technology architectures and human skills, and assess the implications for key functions (Davenport 2018) (Kane 2020). Introducing AI risks exposing or heightening internal rivalry between those who promote traditional ways of working and those who explore new models. Leaders need to be able to mediate in such conflicts between the old and the new digital business (Govindarajan 2019) (Kane 2020).

Envisioning & Communication Skills

Envisioning

Leaders need to provide the necessary "transformational energy" for firms to succeed with AI (Brock 2019). Leaders can provide this energy through a compelling story about why and how the business will benefit from AI – an AI vision (Caridi-Zahavi 2016) (Kane 2020).

Fitzgerald et al. have found that the ability of leaders to form and share a vision has a large impact on the acceptance of a digital transformation in a company. In companies where the CEO had shared its vision on digital transformation, 93% of employees agreed that a digital transformation was the right thing for their companies, and 73% strongly agreed. However, only 36% of the CEOs in that study had shared their vision (Fitzgerald 2013).

An AI vision helps employees make sense of the transformation and creates a sense of connectivity (Ancona 2019) (Caridi-Zahavi 2016). In their vision, leaders have the opportunity to provide information on how they wish to use AI or how competitors or similar companies use AI to improve their performance. The vision can help create the necessary "pull effect" from the organization (Obwegeser 2020) and allows the leader to address obstacles to the transformation such as workers' fear of becoming obsolete and other factors within its company culture that contribute to resistance. It gives the leader the opportunity to frame these obstacles as opportunities rather than threats, perhaps by stating that they intend to use AI to increase people's productivity rather than to replace them (Babic 2019) (Fountaine 2019) (Schepman 2020). When they communicate a compelling vision, leaders also get the opportunity to create a climate of psychological safety and mutual trust and to show their commitment to the digital transformation (YukI 2012) (Fountaine 2019).

Communication

When leaders communicate their AI vision, they need to be able to do so in a convincing way. Further, mediating in the conflicts that can arise during an AI transformation requires a motivating leadership that clearly communicates why the change is necessary. Different leaders have different communication styles. However, Yukl suggests that leaders, irrespective of communication style, use colorful, emotional language, vivid imagery, metaphors, stories, symbols, and slogans when communicating their vision (YukI 2012).

Apt to change

Govindarajan and Immelt argue that since a widespread introduction of AI "changes everything", AI-leadership requires courage. Leading a digital transformation is about being willing to challenge incumbency, ignorance, and the status quo and requires leaders to rethink both how they and their companies work (Govindarajan 2019). Thomke argues that leaders need to embrace a "new model of leadership" which encourages curiosity, drives data-informed decisions, experiments ethically, sets major challenges and creates systematic training and support for rapid experimentation (Thomke 2020). For some leaders these fundamental shifts do not come easy. Not only do their organizations need to change, but also the leaders themselves need to be prepared to change (Sousa 2019) (Kane 2020) (Steiber 2020).

While leaders can tell that they committed to change, their actions may tell otherwise. Kegan and Lahey point out that this is common, but not always a result of opposition or laziness. Instead, "competing commitments – subconscious, hidden goals prevent them from realizing their commitment" (Kegan 2001). Kegan and Lahey argue that it is possible to identify and remedy the competing commitment by identifying what "worrisome outcome" a person is committed to avoid by displaying a certain behavior (Kegan 2001).

Complacency

Fitzgerald et al. and Govindarajan and Immelt highlight complacency among leaders, often rooted in ignorance about AI, as one of the main obstacles to a successful transformation (Fitzgerald 2013) (Govindarajan 2019). Leaders that do not understand AI and its benefits will not be able to drive a change that is fast enough (Fitzgerald 2013). Interestingly, complacency seems to exist primarily at the leadership level. While 53% of the CEOs in one study thought the pace of the digital transformation in their companies was right, fast, or very fast, 78% of their employees thought that the pace was too slow (Fitzgerald 2013).

Perceived need to maintain business "as is"

Teece points out the leaders' perceived need to maintain best practices and high productivity (Teesce 2014). An AI transformation is a balancing act between keeping what functions well in the company against the need to change (Govindarajan 2019). Short term, it is often a safe bet to stick to existing ways of working - for the simple reasons that they work. AI investments often do not provide a return until after the leaders have left the company (Govindarajan 2019). Additionally, if leaders push the transformation too hard and fail, not only do they risk losing their own credibility, they also risk scaring off their best employees (Westerman 2019) (Fitzgerald 2013). This approach of sticking with what works is reinforced by bureaucratic systems of HR, budget, planning, strategy, finance and risk management - that generally are designed to support the continuation of the business "as is". These systems reward those who do what the top wants and offer disincentives to any who attempt different approaches (Denning 2020). Leaders often know more about these bureaucracies than about the organizational forms that could take their place (Ancona 2019).

The fear of losing power

Ancona and Fountaine et al. argue that many leaders are afraid to lose power and worry that chaos will ensue if they loosen the reins (Ancona 2019) (Fountaine 2019). In a study of leaders' attitudes towards AI, although 84% of the participants expected AI to make their work more effective and interesting, 36% feared that it would threaten their jobs (Kolbjørnsrud 2017). A successful AI transformation could render its leader obsolete. Naturally, the incentives to drive any form of change become even lower when your own career is at risk (Yukl 2012).

Assessing the Leadership Competency Framework

In our survey, on the question to what extent they had deployed deep learning solutions, the Laggards scored between 1 and 2, with an average of 1.6/5 – compared to an average of 3.8/5 (with most scores ranging between 4 and 5 and only one score of 3) for the Other Companies. Also on the other questions, the Laggards scored significantly lower.

An illustration of the differences in the score that we received to the questions in the survey is set out in the table below:

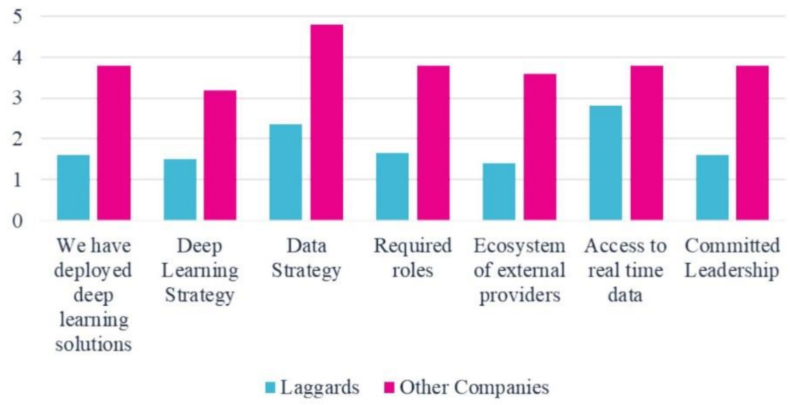

Figure 4: Average survey scores between Laggards and Other Companies on a scale of 1-5 to the questions if they had deployed deep learning solutions, if they have a strategy for using deep learning, if they have a data strategy, if they have the roles required to develop and deploy deep learning solutions, if they have an ecosystem of external service providers, if they have access to real time data and if their leadership is committed to the use of deep learning and AI.

In the interviews that followed, we found that the leaders of the Laggards had not developed the abilities or established the routines of the competency framework. A majority believed that AI and deep learning would have a significant impact on their business in the coming years and most were curious and excited about the change it would bring. In the interviews, they acknowledged the need to create a culture that cultivates curiosity, emphasizes data-informed decisions and supports agile and rapid experimentation. Paradoxically, their actions did not reflect their perceived commitment.

Learning

Almost all the leaders that we interviewed completely lacked an understanding of, or were confused about, the terms AI, machine learning and deep learning. Often, they believed the terms to be synonymous or expected something more from these technologies than what they actually deliver. One of the leaders said:

"Deep learning means that you have deep knowledge about something."

Some had heard of deep learning but could not describe it, or any other AI technology, accurately. Most of the leaders acknowledged they needed to improve their knowledge and understanding of AI. However, hardly anyone had made any efforts in educating themselves on the subject or expressed any plans on doing so. None of the interviewees provided any form of explanation to why they had not gained this understanding, other than the fact that it is a matter of priorities. Others said that the field was so new to them that they did not know where to start. One CEO pointed out his and his colleagues' lack of understanding as the main obstacle to deployment of AI in his company. As the interviews progressed, we noticed a shift in attitude.

As one CEO put it:

> *"If I don't bother learning about these things, then why should someone else in my leadership team?"*

Strategy
None of the leaders that we interviewed had allocated the responsibilities to develop and deploy AI solutions. Since the leaders had not secured the competence necessary to develop and deploy AI solutions – there was simply no one to allocate the responsibility to. Neither had they built an ecosystem of external service providers that could help develop and deploy AI. However, some leaders acknowledged the need to appoint someone to drive the AI deployment. One CEO gave this insightful answer when we asked him to point out the person in his company responsible for the introduction of AI:

> *"Since I have not delegated this responsibility to anyone in my team, the responsibility lies with me. However, realize that I have not fulfilled this responsibility since I have not even made the effort to learn about these things."*

Due to the lack of knowledge, understanding and responsible person for driving AI initiatives, none of the leaders had established a data strategy.

Vision and Communication
None of the Laggards had communicated any vision for how they wished to use AI to support the overall business strategy. When we asked the leaders why, the lack of learning once again came up as one of the main reasons.

> *"How can I create a vision if I don't know what the vision should be about?"* – said one CEO.

However, while other leaders within this CEO's team acknowledged that their CEO lacked foundational knowledge of AI, they also believed that the lack of AI vision was due to a general resistance to change.

Social skills & Aptitude to change
Companies that succeed with deploying AI, often display a greater level of leadership support for the introduction of AI (Brock 2019). In the survey, we asked the leaders to rate the commitment of the rest of the company leadership for the introduction of deep learning and AI. The average score of 1.6/5 amongst the Laggards (compared to 3.8/5 in the other companies that responded to the survey) indicated that the perceived level of commitment to use AI was low.

As mentioned above, Kegan and Lahey state that competing commitments could make leaders resist change. Kegan and Lahey call this that leaders are "Immune to Change" and that this can undermine the success of a company. Kegan and Lahey argue further that change immune people usually do not see that they have a responsibility that things have become the way they are. Instead, they tend to blame the current situation on others (Kegan 2001). Apart from the CEOs, our interviewees rarely described themselves as part of the reasons that their companies fall behind in AI deployment. Instead, the problem was usually due to someone else. Some described their colleagues as conservative, lacking an understanding of how to use data as a resource. Many of the leaders also expected or assumed that someone else had taken the lead in deploying AI and deep learning.

One interviewee made the following illustrating statement:

> *"Even if I have not personally seen an AI strategy, I am sure someone must have created one."*

Kegan and Lahey argue that the competing commitment can be identified and remedied by identifying what "worrisome outcome" a person is committed to avoid by displaying a certain behavior (Kegan 2001). Since many of the interviewees did not even admit that they were part of the problem, it would have been surprising if they suddenly decided to disclose their innermost fears. However, many of the interviewees were quick to guess the competing commitment of other leaders in the companies. Here they mentioned the fear of appearing ignorant or foolish, fear of rivalry, fear of becoming obsolete or losing status when the size of their organizations shrink.

Conclusions

In this study, we create and assess a framework for leaders of manufacturing companies that wish to introduce AI in their businesses. Learning about AI – what it is and its requirements, potential and limitations – is one of the main constituents of the framework. Without an understanding of what it is, it is e.g. impossible to create relevant strategies or create or communicate a vision. We acknowledge that it can be difficult for anyone that wants to take the initial steps to gain a fundamental understanding of AI and its elements. Many news reporters and bloggers seem equally confused as they constantly suggest that the terms are synonymous. A Google search of the definition of AI results in over 32 definitions. Articles that intend to aid leaders in deploying AI, only broadly speak of AI without specifying any technology. Even researchers provide no coherent definition of AI. The terms AI, machine learning and deep learning have so many meanings that it can be difficult for leaders to distinguish what is what.

The fact that there is a constant and rapid development of new technologies and techniques makes the learning even more difficult. In the 16 years from 2002 to 2018, AI patent applications increased by more than 100%. Over the same period, according to the United States Patent and Trademark office, the share of all patent applications that contain AI grew from 9% to nearly 16%. (U.S. Patent and Trademark Office 2020) However, this does not mean that AI is frequently used. According to a study referred to in the Danish newspaper Børsen, only three percent of Danish companies use AI in their operations (Børsen 2022).

Another critical aspect of the framework is the willingness to change. Evolutionarily, resisting change can be a good thing. It ensures that we keep in our known physical and social realms and stops us from doing things that risk our social bonds or status. Ancient regions of our brains control our unwillingness to change and "we often have little choice but to follow the affective dictates that they provide" (Panksepp 2012). However, if we are to trust Govindarajan and Immelt and Ancona, leaders of manufacturing companies either embrace the change or risk extinction (Govindarajan 2019) (Ancona 2019).

References

Alsheiabni, S., Cheung, Y., Messom, C. 2019. "Towards an Artificial Intelligence Maturity Model: From Science Fiction to Business Facts". *Twenty-Third Pacific Asia Conference on Information Systems*.

Ancona, D. 2019. "Five Rules for Leading in a Digital World". *MIT Sloan Management Review*, Volume 61, Issue 1, pages 1-4.

Babic, B., Chen, D. L., Evgeniou, T., Fayard A-L,. 2019. "A better way to onboard AI". *Harvard Business Review*, July-August Issue.

Berman, S., Heller Baird, C., Eagan, K., Marshall, A. 2020. "What makes a Chief Digital Officer successful". *Strategy & Leadership*, Volume 48 (2), pages 32-38.

Brinch, M., Gunasekaran, A., Wamba, S. F. 2020. "Firm-level capabilities towards big data value creation". *Journal of Business research*, Volume 131, pages 539-548.

Brock, J. K-U. Wangenheim, F. von. 2019. "Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence". *California Management Review*. Volume 61 (4), pages 110-134.

Børsen. 2022. "Humanister skal oversætte teknologi – så alle forstår det: Men virksomhederne ansætter dem ikke". May 06. Accessed June 09, 2022. *https://borsen.dk/nyheder/okonomi/ai-virksomheder-gaar-uden-om-humanister.*

Caridi-Zahavi, O., Carmeli, A., Arazy, O. 2016. "The Influence of CEOs' Visionary Innovation Leadership on the Performance of High-Technology Ventures: The Mediating Roles of Connectivity and Knowledge Integration". *Journal of Product Innovation Management*, Volume 33, Issue 3, pages 356-376.

Davenport, T.H., Foutty, J. 2018. "AI-Driven Leadership". August 10. Accessed February 12, 2020. *https://sloanreview.mit.edu/article/ai-driven-leadership/.*

Denning, S. 2020. "Why a culture of experimentation requires management transformation". *Strategy & Leadership*, Volume 48 (4), pages 11-16.

Felin, T., Foss, N. J., Heimeriks, K. H., Madsen, T. L. 2012. "Microfoundations of Routines and Capabilities: Individuals, Processes, and Structures". *Journal of Management Studies*, Volume 49, Issue 8, pages 1351-1374.

Fitzgerald, M., Kruschwitz, N., Bonnet, D., Welch, M. 2013. "Embracing Digital Technology - a new strategic imperative". *MIT Sloan Management Review*, pages 3-12.

Fountaine, T., McCarthy, B., Saleh, T. 2019. "Building the AI-Powered Organization". *Harvard Business Review*, July-August Issue, pages 62-73.

Geneviève, B. Benbasat, I., Horner Reich, B. 2003. "The Influence of Business Manager's IT Compentence on Championing IT". *Information Systems Research*, Volume 14, No. 4, pages 317-336.

Govindarajan, V., Immelt, J.R. 2019. "The Only Way Manufacturers Can Survive". *MIT Sloan management review*, Volume 60 (3), pages 24-33.

Kane, G. C., Phillips, A. N., Copulsky, J., Andrus, G. 2020. "How Digital Leadership Is(n't) different". *MIT Sloan Management Review*, Volume 60 (3), pages 34-39.

Kegan, R., Lahey, L. L. 2001. "The Real Reason People Won't Change". *Harvard Business Review*, November Issue, pages 77-85.

Kolbjørnsrud, V., Amico, R., Thomas, R. J. 2017. "Partnering with AI: how organizations can win over skeptical managers". *Strategy & Leadership*, Volume 45 (1), pages 37-43.

Marcus, G. 2018. "Deep Learning: A Critical Appraisal". Accessed February 14, 2022. *https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf.*

McCarthy, J. 2007. "What Is Artificial Intelligence?" Technical report, Stanford University. *http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.*

Mitchell, T. 1997. *Machine Learning*. McGraw Hill.

Mukherjee, A. S. 2020. "5 Musts for Next-Gen Leaders". *MIT Sloan Management Review*, Volume 61, Issue 3, pages 1-4.

Ng, A. 2016. "What Artificial Intelligence Can and Can't Do Right Now". *Harvard Business Review*. Accessed March 03, 2022. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now.

Obwegeser, N., Yokoi, T., Wade, M., Voskes, Tom. 2020. "7 Key Principles to Govern Digital Initiatives". *MIT Sloan Management Review*, Volume 61, Issue 3, pages 1-9.

Panksepp, J., Biven, L. 2012. *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. Norton Series on Interpersonal Neurobiology.

Persaud, A. 2020. "Key competencies for big data analytics professions: a multimethod study". *Information Technology & People*, Volume ahead-of-print No. ahead-of-print. .

Pyle, D., San José, C. 2015. "An executive's guide to machine learning". *McKinsey Quarterly*.

Reck, F., Fliaster, A. 2019. "Four Profiles of Successful Digital Executives". *MIT Sloan Management Review*, Volume 60, Issue 3, pages 1-7.

Rogati, M. 2017. "The AI Hierarchy of Needs". June 12. Accessed March 03, 2022. *https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007*.

Saari, L., Kuusisto, O., Pirttikangas, S. 2019. "AI Maturity Web Tool Helps Organisations Proceed with AI". VTT Technical Research Centre of Finland,. doi:https://doi.org/10.32040/Whitepaper.2019.AIMaturity.

Schepman, A., Rodway, P. 2020. "Initial validation of the general attitudes towards Artificial Intelligence Scale". *Computers in Human Behavior Reports*, Volume 1, January-July Issue.

Sousa, M. J., Rocha, A. 2019. "Skills for disruptive digital business". *Journal of Business Research*, Volume 94, pages 257-263.

Steiber, A., Alänge, S. Swapan Ghosh, Dulce Goncalves,. 2020. "Digital transformation of industrial firms: an innovation diffusion perspective". *European Journal of Innovation Management*, Volume 24, No. 3, pages 799-819.

Teesce, D. J. 2014. "The Foundations of Enterprise Performance: Dynamic and Ordinary Capabilities in an (Economic) Theory of Firms". *The Academy of Management Perspectives*, Volume 28 (4), pages 328–352.

Thomke, S. 2020. "Building a Culture of Experimentation". *Harvard Business Review*, March-April Issue, pages 40-48.

U.S. Patent and Trademark Office. 2020. Inventing AI, Tracing the diffusion of artificial intelligence with U.S. patents. October. Accessed June 09, 2022. *https://www.uspto.gov/sites/default/files/documents/OCE-DH-AI.pdf*.

Weill, P., Apel, T., Woerner, S. L., Banner, J. S. 2019. "It Pays to Have a Digitally Savvy Board". *MIT Sloan Management Review*, Volume 60, Issue 3, pages 41-45.

Westerman, G., Soule, D. L., Eswaran, A. 2019. "Building Digital-Ready Culture in Traditional Organizations". *MIT Sloan Management Review*, Volume 60, Issue 4, pages 59-68.

YukI, G. 2012. "Effective Leadership Behavior: What We Know and What Questions Need More Attention". *Academy of Management Perspectives*, Volume 26, No. 4, pages 66-85.

Appendix 1 – Articles included in Literature Review

Research Discipline	Title	Author	Publication	Year
Information Systems	The Influence of Business Managers' IT Competence on Championing IT	Geneviève Bassellierz Izak Benbasat, Blaize Horner Reich	Information Systems Research	2003
Information Systems	Information Technology Leadership Challenges – Past, Present, and Future	Ray Hoving	Information Systems Management	2007
Information Systems	Key competencies for big data analytics professions: a multimethod study	Ajax Persaud	Information Technology & People, Vol. ahead-of-print No. ahead-of-print.	2020
Information Systems	Initial validation of the general attitudes towards Artificial Intelligence Scale	Astrid Schepman, Paul Rodway	Computers in Human Behavior Reports	2020
Information Systems	Beyond design and use: How scholars should study intelligent technologies	Diane E. Bailey, Stephen R. Barley	Information and Organization	2020
General Management	The Foundations of Enterprise Performance: Dynamic and Ordinary Capabilities in an (Economic) Theory of Firms	David J. Teesce	The Academy of Management Perspectives	2001
General Management	Effective Leadership Behavior: What We Know and What Questions Need More Attention	Gary YukI	The Academy of Management Perspectives	2003
General Management	Microfoundations of Routines and Capabilities: Individuals, Processes, and Structures	Teppo Felin, Nicolai J. Foss, Koen H. Heimeriks and Tammy L. Madsen	Journal of Management Studies	2012
General Management	A Model of Strategic Entrepreneurship: The Construct and its Dimensions	R. Duane Ireland, Michael A. Hitt, David G. Sirmon	Journal of Management	2012
General Management	Partnering with AI: how organizations can win over skeptical managers	Vegard Kolbjørnsrud, Richard Amico and Robert J. Thomas	Strategy & Leadership	2013
General Management	Demystifying AI: What Digital Transformation Leaders Can Teach You about Realistic Artificial Intelligence	Jürgen Kai-Uwe Brock and Florian von Wangenheim	California Management Review	2014
General Management	AI-Driven Leadership	Davenport, Thomas H; Foutty, Janet	Cambridge: Massachusetts Institute of Technology, Cambridge, MA (blog post)	2017
General Management	Building a Culture of Experimentation	Stefan Thomke	Harvard Business Review	2018
General Management	Building the AI-Powered Organization	Tim Fountaine, Brian McCarthy, Tamim Saleh	Harvard Business Review	2018
General Management	The Real Reason People Won't Change	Robert Kegan, Lisa Laskow Lahey,	Harvard Business Review	2019
General Management	What makes a Chief Digital Officer successful	Saul Berman, Carolyn Heller Baird, Kevin Eagan and Anthony Marshall	Strategy & Leadership	2019

Appendix 2

AI & SOCIETY
https://doi.org/10.1007/s00146-023-01688-z

OPEN FORUM

Abstraction, mimesis and the evolution of deep learning

Jon Eklöf[1,3] · Thomas Hamelryck[1] · Cadell Last[2] · Alexander Grima[3] · Ulrika Lundh Snis[4]

Received: 23 June 2022 / Accepted: 4 May 2023

Abstract
Deep learning developers typically rely on deep learning software frameworks (DLSFs)—simply described as pre-packaged libraries of programming components that provide high-level access to deep learning functionality. New DLSFs progressively encapsulate mathematical, statistical and computational complexity. Such higher levels of abstraction subsequently make it easier for deep learning methodology to spread through mimesis (i.e., imitation of models perceived as successful). In this study, we quantify this increase in abstraction and discuss its implications. Analyzing publicly available code from Github, we found that the introduction of DLSFs correlates both with significant increases in the number of deep learning projects and substantial reductions in the number of lines of code used. We subsequently discuss and argue the importance of abstraction in deep learning with respect to ephemeralization, technological advancement, democratization, adopting timely levels of abstraction, the emergence of mimetic deadlocks, issues related to the use of black box methods including privacy and fairness, and the concentration of technological power. Finally, we also discuss abstraction as a symptom of an ongoing technological metatransition.

Keywords Deep learning · Evolution of deep learning · Abstraction · Mimesis

1 Introduction

Deep learning is a branch of machine learning that plays a vital role in science, medicine, technology and governance. In the past decade, deep learning software frameworks (DLSFs) have emerged that make it possible to develop applications that previously required in-depth knowledge of computer science, statistics and mathematics. Typically, DLSFs are open source, i.e., they are publicly available and free to use, which promotes the spread of new deep learning methodologies through mimesis. Abstraction (from the Latin *abs*, meaning "away from" and *trahere*, meaning to "draw") is the process of removing characteristics in order to focus on essential characteristics (Ganascia 2015). DLSFs can be interpreted as attempts to provide increasingly higher levels of abstraction to developers of deep learning. In this article, we first quantify the progress of abstraction in deep learning by examining software projects on Github, a large software repository (GitHub 2022a, b), and subsequently discuss some of its implications.

1.1 Abstraction in deep learning

Abstraction is the process to derive general rules and concepts from individual facts or situations and from these build formal descriptions of the underlying facts or situations. "An abstraction" is the outcome of this process. Instead of accounting for every detail of each fact or situation, a higher abstraction level generalizes the relevant patterns that characterize a larger set (Langer 1953). Abstractions help us understand and explain reality in a simpler way, and can be useful and powerful when used correctly. However, they can also limit our understanding. Since abstractions simplify reality to help us focus on certain parts, they risk leaving out other important aspects. Consequently, certain nuances may be lost as particular features are distilled through abstraction.

✉ Jon Eklöf
jon.g.eklof@gmail.com

Cadell Last
https://philosophyportal.online

1 Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

2 Center Leo Apostel (CLEA), Vrije Universiteit, Brussels, Belgium

3 GKN Aerospace Engines, Trollhättan, Sweden

4 School of Business, Economics and IT, University West, Trollhättan, Sweden

The opposite of abstraction is specification, which describes the process of breaking down general rules and concepts into concrete facts or situations. Ganascia provides the following illustrative description of abstraction and the meaning of higher abstraction levels:

> "[…] figures such as triangles, spheres or pyramids are abstractions of shapes of objects, geometry is an abstraction of figures and algebraic geometry is an abstraction of geometry. In the same way, integers are abstractions of sets of objects, real numbers are abstraction of measures and associative rings—algebraic structures—are abstractions of integers and real numbers" (Ganascia 2015).

Abstraction is critical for the development of science, as a scientific theory is a kind of abstraction. Mechanics, thermodynamics, electromagnetism, relativity and quantum mechanics all correspond to various levels of abstraction in physics. Abstract concepts such as gene expression, transcription and regulation are important in biology. Also in the humanities, abstractions are widely used. Historians use concepts that abstract past reality, such as Max Weber's conceptual apparatus, with the notions of rationalization and secularization (Ganascia 2015).

DLSFs are software libraries containing programming building blocks that provide generic deep learning functionality. They shield their users from many mathematical, statistical and computational complexities and make it possible to utilize advanced hardware such as graphics processing units (GPUs) (Demirović et al. 2018; Shi et al. 2016). DLSFs make it possible to take advantage of other programmers' work, with only a summary understanding of the underlying principles or technologies. If needed, a DLSF's functionality can be selectively changed by additional user-written code to provide application-specific functionality. However, this is often a complex process, as it requires the user to go below the current abstraction level.

1.2 Mimetic theory

The American-French anthropologist and cultural theorist René Girard (1923–2015) posited that human desire is mimetic: humans tend to desire objects that are desired by those around them (Girard et al. 2002; Girard 1965). Mimetic theory indicates that mimesis is a complex and generative phenomenon at the heart of human culture, religion, economy, politics and technology. Desire has a triangular nature, involving a subject, a model and an object of desire. The subject looks to a model to learn about what is desirable. In addition, desire is mutually reinforcing; one person's desire can render an object more attractive for another person, which in turn increases the interest of the first person. This aspect of desire is called double mediation and lies at the heart of many hypes, fashions, bubbles, trends and frenzies. Advertisement is easily understood from the point of view of mimetic theory: it aims to provide credible and attractive models that induce subjects to desire certain objects through triangular mimesis. Humans are embedded in vast networks of desire in which they act as models and subjects of desire to each other, and compete over objects. Today, such networks of desire increasingly exist in the digital realm of the internet (Girard 2009). In the context of this article, we consider the interplay of mimesis and abstraction. We argue that abstraction, as for example in the case of DLSFs, makes it easier to adopt the concepts, methods and algorithms of mimetic models that are successful, or that are perceived as successful.

2 Objective

In this article, we quantify the increase of abstraction in deep learning by examining publicly available code on Github. Our hypothesis is that, if abstraction levels within deep learning increase, the number of lines of code in deep learning projects should decrease. We view the reduction in the number of lines of code as an indication of the ongoing advancements in AI and the increasing accessibility of AI technology. The quantification of abstraction using the number of lines of code used for deep learning applications serves as a starting point for discussing its philosophical and societal implications. The main contribution of the paper is to provide an analysis of the opportunities and challenges resulting from increased abstraction. We discuss and argue the importance of abstraction in deep learning with respect to ephemeralization, technological advancement, democratization, the need to use timely abstractions, the emergence of mimetic deadlocks, issues related to the use of black box methods including privacy and fairness, and the concentration of technological power. Finally, we also discuss abstraction as a symptom of an ongoing technological metatransition.

3 Methods

Github.com is a provider of Internet hosting for software development and version control where developers store their software projects. In January 2021, when we started to collect the data for this study, Github had over 57 million users and contained over 160 million projects stored in so-called repositories (GitHub 2022a, b). To identify deep learning repositories, we searched through all these repositories using the following keywords: '*deeplearning*', '*deep learning*' and '*deep-learning*'. We identified 605 403 repositories that fit these descriptions. The intention was to download and analyze a random sample amounting to half of these 605,403 repositories. To ensure a representative sample, we employed uniform sampling using the np.random.choice command in Python. This command selects a random subset of data from a larger dataset, with each data point having an equal probability of being chosen. However, after we constructed the random sample and initiated the download process (that took over eight months to complete), some of the repositories of the sample had changed from public to private or been deleted. We therefore ended up with a random sample of 317,428 repositories. After removing all instances of forked repositories (i.e., repositories copied from other repositories) we ended up with a final dataset of 37,915 repositories. From this dataset, we extracted the following information: "*date*"—when the repository was created, "*last commit*"—when the repository was last changed, "*language*"—what programming language that was used for the project and "*code_count*"—the number of lines of code used in the repository. First, we investigated the number of deep learning projects based on when they were created. Second, we investigated the evolution of the median of the number of lines of code. In Fig. 2, we plot the median of the number of lines of code (y-axis) of all projects whose last activity on GitHub occurred in the given year (x-axis). Finally, we investigated the increase in the number of projects and the reduction in number of lines of code in view of the initial release dates of the major DLSFs: Keras (version 0.2.0), TensorFlow (version 0.5.0), PyTorch (alpha-1), Jax (version 0.1.47) and TensorFlow 2.0.

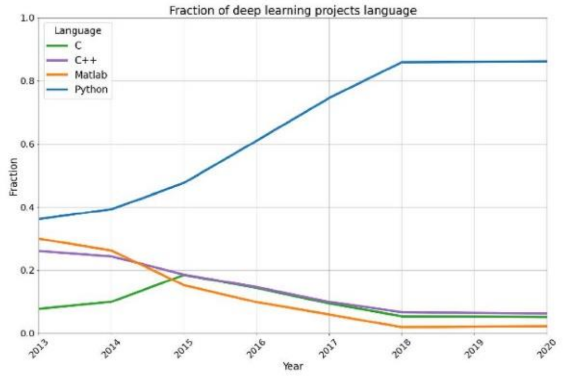

Fig. 1 Yearly development of the proportionate use of the programming languages C (green line), C++ (purple line), Matlab (orange line) and Python (blue line) for deep learning projects on GitHub. The *y*-axis shows the fraction of projects in each language; the *x*-axis shows the year of the first commit. A Github commit is a record of changes made to a code repository. It is a snapshot of the changes made to the code, typically with a brief description of the changes made

4 Results

4.1 Python is the predominant programming language

Python has been the dominant programming language for deep learning projects on GitHub for as long as our data extends. As we can see in Fig. 1, between 2014 and 2018, its popularity increased steadily. Since then, Python has maintained its position as the programming language of choice for deep learning projects. Today, over 85% of the deep learning projects use Python and, as can be seen in Fig. 1, this number shows no signs of diminishing. Since Python accounts for such a large share of the deep learning projects on GitHub, we have focused our further investigations on this programming language only.

4.2 Significant increase of deep learning projects

Between 2013 and the end of 2015, the number of deep learning projects in the sample only increased slightly. Beginning in 2016, until 2018, the number of deep learning projects increased drastically. In Fig. 2, we can see changes in the rate at which the increase in the number of deep learning projects has occurred (orange dotted line). Some of these changes roughly coincide with the release of DLSFs in Python. The first increase in the beginning of 2015 coincides with the release of Keras (Keras 2016). The second coincides with the release of TensorFlow (Google AI 2015) and the third change that occurred at the end of 2016 coincides with the release of PyTorch, which was released in September 2016 (PyTorch 2018; GitHub and PyTorch 2022a, b).

4.3 Reduced number of lines of code

In this study, we aim to quantify the increase in abstraction using a reasonable proxy: we investigate the number of lines of code used in deep learning projects. Our hypothesis is that, as abstraction levels within deep learning increase, the number of lines of code in deep learning projects should decrease. The GitHub data showed exactly this. We could see that as the DLSFs were launched, the median number of lines of code of the deep learning projects changed. From 2014 until 2016, the median number of lines of code of active deep learning repositories on GitHub increased. However, since the end of 2016, the median number of lines of code of active deep learning repositories has decreased significantly. In 2016, the median number of lines of code used was about 3700. At the end of 2016, there was a sharp drop in the number of lines of code. By the end of 2017, the median had reduced to about 1200 lines of code. This sharp drop coincides with the release of TensorFlow and PyTorch

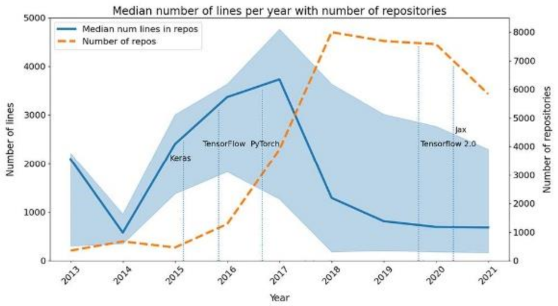

Fig. 2 Development of the median of the number of lines of Python code of deep learning projects (blue solid line) together with the number of deep learning repositories (orange dotted line) on Github. The blue shaded area shows the spread of the median number of lines using the 25th and 75th quartile of the sample. The left y-axis shows the median number of the lines of code of projects whose last commit on GitHub was in the year specified on the x-axis; the right y-axis shows the number of repositories initiated in the year specified on the x-axis. The blue dotted vertical lines display the initial release dates of the stable versions of important DLSFs (Keras 0.2.0, TensorFlow 0.5.0, PyTorch alpha-1, TensorFlow 2.0, and Jax 0.1.47)

(Google AI 2015; PyTorch 2018; GitHub and PyTorch 2022a, b). The next drop occurred at the end of 2017. After 2017, there have been no major reductions in the number of lines of code used similar to those in the end of 2016 and 2017. The decrease continues, but at a slower rate. By the end of 2020, the median was down to about 600 lines of code. This means that over the past years, the median number of lines of code has reduced by more than 80%.

5 Implications

The launch of open source DLSFs such as TensorFlow and PyTorch correlates both with significant increases in the number of deep learning projects as well as with major reductions in the number of lines of code used to build deep learning models. The reduction in the number of lines of code is an indication of the ongoing advancements in AI technology. An increasing amount of people use deep learning solutions based on abstractions that allow them to do more with less. We call this ongoing process "abstraction explosion". We discuss the possible implications of the progress of abstraction in deep learning below.

5.1 Abstraction contributes to ephemeralization

The architect and systems theorist Buckminster Fuller (1895–1983) coined the term ephemeralization—the ability through technological advancement to do "*more and more with less and less until eventually you can do everything with nothing*" (Fuller 1938). DLSFs make it possible to build deep learning models that can perform increasingly difficult tasks (Wang et al. 2020; Oneto et al. 2019) while requiring less and less actual code. This can be interpreted as the ephemeralization of deep learning. Even if our results indicated that the reduction of the number of lines of code used for deep learning models seems to have halted over the past years, deep learning models based on the existing DLSF can solve increasingly difficult tasks (Wang et al. 2020; Oneto et al. 2019). One explanation for the success of abstraction is the increase in the amount of available data. From 2010 to 2020, the amount of data created, captured, copied, and consumed in the world increased by almost 5000% growth, from 1.2 trillion gigabytes to 59 trillion gigabytes (Press 2020). Deep learning models rely upon large amounts of data and their predictions become more accurate the more data they are fed (Marcus 2018).

5.2 Technological advancement through feedback loops

Today, deep learning models typically run on graphics processing units (GPUs) since this shortens the training and inference time of the models compared to running the models on central processing units (CPUs) (Demirović et al. 2018; Shi et al. 2016). Using GPUs for deep learning models used to be a cumbersome process that required a lot of manual work to distribute the calculations on the GPU. Now, GPUs are easy to use for deep learning models because their complexity is effectively abstracted away (Demirović et al. 2018). As a result, more and more people use GPUs to run deep learning models. This has created a feedback loop that has promoted the further development of GPUs. This can be interpreted as an example of "the law of accelerating returns" proposed by Kurzweil; as a process becomes more effective, greater resources are available to progress that process further (Kurzweil 2001).

5.3 Fast succession of paradigm shifts

Abstraction can stimulate the emergence of new paradigmatic technologies such as deep learning, deep probabilistic programming (Bingham et al. 2019), quantum computing (Havlíček et al. 2019) and causal inference (Pearl 2000) by making them comparatively easy to use. New solutions involving so-called "low coding" or even "no coding" (low-code, no-code) are already replacing conventional software development (Kaye 2022; Economist 2022). For example, through its ChatGPT model, OpenAI has launched a system that is able to translate natural language into multiple programming languages (OpenAI 2023). This takes coding to a completely different level since it makes it possible to build models without any actual writing of code. To some extent, it even renders our method in this study, trying to quantify the abstraction explosion by counting lines of code, obsolete. Tying such high levels of abstraction with emerging technologies can promote the emergence of rapid and decisive paradigm shifts.

5.4 Democratization of deep learning

Many DLSFs, including TensorFlow and PyTorch, are open source DLSFs, which means that they are free for anyone to use, modify and redistribute. The release of open source software, along with a broader socio-cultural shift towards participation in media and cultural production, increases the opportunities for democratization of production, governance and knowledge exchange (Powell 2012). The DLSFs have made it much easier for people to collaborate and build on each other's work, allowing for a surge in technological innovation and progress (Bommasani et al. 2021).

In addition, DLSFs provide access to technologies that previously required in-depth knowledge of mathematics, statistics and programming, which lowers the demands with respect to required education. This, in turn, facilitates faster adoption of technology by a wider range of users of DLSFs, as specialized or in-depth knowledge is not required. We note that, although the DLSFs have democratized deep learning, access to the most advanced deep learning models are moving in the opposite direction. Models such as GPT-3 are only accessible through an API, which is shared with a limited pool of people (Bommasani et al. 2021).

5.5 Adopting timely levels of abstraction

The rapid progress of abstraction in DLSFs makes it increasingly important to use the latest programming frameworks that allow adopting timely, appropriate levels of abstraction. Failing to do so, companies or whole societies could waste time and resources using obsolete abstractions and ultimately risk being outcompeted by those who use a more timely level of abstraction. Because of this rapid development, keeping up with the abstraction explosion is a challenge in itself and requires a commitment to learning about new, emerging levels of abstraction. This commitment requires a proactive mindset and the aptitude to acknowledge one's limitations in order to identify the need for further knowledge. In addition, the commitment entails fostering critical thinking skills and the ability to go beyond superficial understandings of technological advancements to evaluate their potential implications and limitations. Such commitment requires adequate resources in terms of money, time and cognitive capacity. Keeping up with the progress of abstraction will remain a challenging task in governance, industry, medicine, the military and academia.

5.6 Mimetic deadlocks

Increased abstraction has simplified the adoption and use of new methods and the democratization has made them publicly available. This risks leading to dead ends if solutions are solely sought within the familiar and convenient limits of the abstractions that everybody is using. We call this kind of unproductive herd behavior a "mimetic deadlock". The reason we risk ending up in these deadlocks is that we let others guide our behavior. Girard describes our tendency to rely on others to determine how we should behave:

> "Man is the creature who does not know what to desire, and he turns to others in order to make up his mind. We desire what others desire because we imitate their desires." (Girard 1988).

To some extent, the DLSFs that we discuss in this article are a result of beneficial mimesis, where groups of individuals commonly engage in collective practices to generate new knowledge and understanding. The application interfaces of TensorFlow, PyTorch and Jax all are based on the older software framework NumPy. In 1935, Ludwik Fleck (1896–1961), a Polish physician, microbiologist, and philosopher of science, referred to such a mimetic process in science as a "*thought collective*" (Sady 2001). The democratization of deep learning and the rise of collaboration and community-driven development point to the formation of such a thought collective. As more people are able to access and contribute to the development of deep learning, new thought collectives might emerge that may shape the future of the field.

As Fleck already understood, mimesis also risks narrowing interest and constraining innovation. Bommasani et al. point out that we see a homogenization in the selection of models based on these DLSFs and how they are used (Bommasani et al. 2021). Rather than finding the best way to solve a problem, developers risk converging to the use of algorithms and methods used by successful models around them. This can lead to a mimetic deadlock: a convergence to an agreed convenient solution that is not necessarily optimal or even good. Abstraction exacerbates this phenomenon as it makes it easy to copy without understanding the underlying issues.

Further, the interplay between abstraction and mimesis risks creating a feedback loop where the simplified representations of reality are further imitated. This can lead to a gradual reduction of nuance and complexity, contributing to a homogenized and streamlined perception of reality. As people are exposed to these simplified and imitative versions of reality, they may begin to adopt a more uniform way of thinking and perceiving the world. This can hinder their ability to appreciate the full complexity and diversity of reality, ultimately limiting their understanding and perspectives.

5.7 The black box problem

Often, we cannot account fully for how deep learning algorithms make their decisions (Castelvecchi 2016) but are only informed of the final decisions made by the algorithm (Holweg et al. 2022). Higher abstraction levels hide part of the computational complexity of the lower levels. This makes it increasingly difficult both to confirm that the models do what they are supposed to do and to troubleshoot them. Bommasani et al. argues that any flaws in the lower levels of abstraction "*are blindly inherited*" by the higher levels of abstraction (Bommasani et al. 2021). For example, rounding errors can change the topological properties of activation functions in deep networks, paradoxically resulting in improved performance (Naizat et al. 2020). In other words, the improved performance results from an error deeply hidden in layers of abstraction. Such phenomena obviously make the correct interpretation of the performance of machine learning models challenging.

Not knowing if the abstraction level works as intended could have serious implications, particularly where deep learning applications make decisions that are directly affecting human well-being or are used in areas where malfunctions could have fatal consequences, such as the aeronautics or nuclear industry (Holweg et al. 2022; Ganascia 2015). It can also amplify already existing problems with deep learning in terms of fairness, privacy and quality assurance problems.

5.8 Fairness, privacy and quality assurance problems

Decreased explainability in combination with the democratization of deep learning increases the risk of fairness, privacy and quality assurance problems. Even without the democratization of deep learning, the algorithms can amplify societal stereotypes (Wang et al. 2019; Bommasani et al. 2021). For example, a recruiting tool for the fields of science, technology, engineering and math jobs believes men are more qualified and shows bias against women (Kiritchenko and Mohammad 2018), and facial recognition software has proven to perform poorly for females with darker skin (Buolamwini 2018). However, there might be multiple conditions of fairness that (with mathematical certainty) cannot be satisfied at the same time (Kleinberg et al. 2017). Increasing fairness with respect to one condition can thus very well result in decreasing fairness with respect to other conditions (Kleinberg et al. 2017).

When deep learning becomes increasingly accessible and easier to use, and more organizations use it to increase efficiency and effectiveness, the number of cases where its application violates social norms and values rises (Holweg et al. 2022). Through the abstraction explosion, non-experts have access to tools to create deep learning solutions. However, non-experts often lack fundamental skills in standard processes for debugging as well as testing for quality control and reliability. They also lack adequate strategies to deal with the effects of malfunctioning algorithms. Most organizations are not strategically prepared on how to respond to AI failures. If people are using the increased abstraction levels without being able to understand if they work as intended, or what they may or may not be used for, they can unintentionally unleash "unfair" algorithms.

Attempts have been made trying to combat this increased risk of "unfair" algorithms. Singapore has released a software toolkit aimed at helping financial institutions ensure they are using AI responsibly (Yu 2022). Although this to some extent may work to increase the transparency of algorithms, it does not correct the root cause of the problem where the algorithms in some way are malfunctioning. Correcting errors of deep learning models requires an understanding of the underlying abstraction levels, which in its turn often requires an in-depth understanding of both math and programming.

5.9 Abstraction introspection and informative error messages

The difficulty in correcting malfunctioning algorithms is amplified further by the fact that the error messages generated by malfunctioning code are often vague, imprecise, confusing and at times seemingly incorrect—particularly for novices (Brown 1983). The professor in human-computing interaction Amy J. Ko from Washington University, has even referred to programming languages as the least usable, but most powerful human–computer interfaces ever invented (Ko 2014). Incomprehensible error messages break down the ability to see through layers of abstraction and are thereby often unhelpful in correcting the code. This is particularly problematic as these messages act as the primary source of information to help novices rectify their mistakes (Becker et al. 2019). For the average person to be able to use error messages to correct malfunctioning code requires another level of abstraction that allows the algorithms themselves to "reflect" about their processes and—if not rectifying itself—at least provide an understandable explanation of the problem that enables the user to rectify the problem. However, with today's abstraction levels, understanding error messages is comparable to the difficulty of reading source code (Becker, et al. 2019). Consequently, external experts that may be hard to find are often needed to solve these issues.

5.10 Experts are needed but scarce

When abstractions fail or turn out to be inadequate, the challenge is to "*make the abstraction more concrete*",[1] moving from abstraction to the underlying specific abstraction levels (Ganascia 2015). The average user has limited time, resources, competence and cognitive capacity to deal with the complexities of the DLSF other than on a superfluous level. As we have seen in this study, increased abstraction leads to more and more people using the higher levels of abstraction. Further, since non-experts often lack fundamental skills in standard processes for debugging as well as testing for quality control and reliability an increasing amount of errors will go unnoticed. In addition, when something goes wrong, comparatively few users will know how to correct these errors. Heylighen has pointed out that increasing system complexity and information overload can negate the advantages of ephemeralization since it puts an increasing amount of pressure on the reducing number of experts able to control the ephemeralized systems (Heylighen, Accelerating Socio-Technological Evolution: from ephemeralization and stigmergy to the global brain, 2007). With just 25 million people around the world fluent in standard programming languages, the proportionate number of people able to correct errors in higher abstraction levels decreases as abstraction increases (Economist 2020). The scarcity of expertise is worsened by the fact that the industry provides higher salary than academia and often a more satisfying work environment (Wolff et al. 2020; Woolston 2021). In addition, academia rarely has access to the most advanced deep learning models (Bommasani et al. 2021).

Experts are, therefore, more likely to choose an industry career rather than pursuing academic careers in statistics, computer science or mathematics. However, deep learning builds upon these traditional sciences and it is to these fundamental levels we need to revert when something goes wrong. Ultimately, we may end up in a situation where the experts able to correct malfunctioning abstraction levels are exceedingly hard to find. If there are any experts left, they become increasingly expensive to hire.

5.11 Abstraction is power

Influencing the progress of abstraction and having access to experts when developing and using these abstractions becomes a considerable source of power. The philosopher Bertrand Russell pointed to the connection between abstraction and practical power:

> "A financier, whose dealings with the world are more abstract than those of any other 'practical' person, is also more powerful than any other practical person. Financiers can deal in wheat or cotton without needing ever to have seen either: all they need to know is whether the price will go up or down. This is abstract mathematical knowledge, at least as compared to the knowledge of the agriculturist. Similarly the physicist, who knows nothing of matter except certain laws of its movements, nevertheless knows enough to be able to manipulate it." (Russell 2009).

Initially, academia often developed new levels of abstraction within deep learning, such as when the Montreal Institute for Learning Algorithms at University of Montreal developed the DLSF Theano (Brownlee 2016).[2] To some extent, this guaranteed a certain level of transparency on the considerations made when producing the abstraction levels. Today, large corporations have outcompeted academia in the development of new abstraction levels (Bommasani et al. 2021; Economist 2020). One such example is the discontinuation of development of DLSF Theano. According to the development team behind Theano, it was outcompeted because of "*strong industrial players*" (Lamblin 2017).

These large corporations influence and steer the democratization of deep learning by developing and releasing DLSFs to the public. As such, they also act as mimetic models, both to other corporate agents and academia, pointing to specific abstractions and the types of problems they can solve. In addition, the same companies control a large portion of the data that is shared on the internet (Norwegian Consumer Council 2018) and they employ an increasing fraction of experts (Bommasani et al. 2021). Abstraction, data and the experts to make use of them are thus sources of considerable mimetic and practical power.

6 Conclusions and outlook

In this article, we quantify the increase of abstraction in deep learning by examining publicly available code on Github and discuss its philosophical and societal implications.

A metasystem integrates a number of initially independent components and creates a qualitative different system by steering or controlling their interactions (Turchin 1977; Heylighen 2003). Our analysis highlights the importance of considering the broader societal and philosophical implications of the abstraction explosion in deep learning, as it may signify a larger technological transition that moves from

[1] Note that this is akin to the abstract, the negative and the concrete of Hegelian dialectics.

[2] See for example the groundbreaking research by Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. (Krizhevsky et al. 2012).

the physical into the digital metasystem with consequences for how we understand and interact with technology (Last 2017). Turchin and Heylighen have referred to such major changes as metasystem transitions (Turchin 1977; Heylighen 2003). Several transitions have followed the same process through key evolution in information storage and replication (Szathmáry et al. 1995; Maynard Smith and Szathmary 1997; Szathmáry 2015; Turchin 1977). One of the drivers of the current metasystem transition is an increase in abstraction, but also what Kurzweil refers to as the law of accelerating returns (Kurzweil 2001). As abstraction increases, feedback loops will lead to more and more people using these new tools, driving the progress further. Reaping the benefits of abstraction, while avoiding associated potential problems such as mimetic deadlock, black box issues concerning privacy and fairness, and excess power concentration, will remain an ongoing challenge in the foreseeable future for academia, industry and governance.

Acknowledgements We acknowledge funding from the University of Copenhagen, Department of Computer Science, Denmark; GKN Aerospace Engines, Sweden and the Center Leo Apostel (CLEA), Vrije Universiteit Brussel, Belgium.

Funding Open access funding provided by Royal Danish Library.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

References

Becker B, Denny P, Pettit R, Bouchard D, Bouvier D, Harrington B, Prather J (2019) Compiler error messages considered unhelpful: the landscape of text-based programming error message research. IN: Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education, pp 177–210

Bingham E, Chen J, Jankowiak M, Obermeyer F, Pradhan N, Karaletsos T, Goodman N (2019) Pyro: deep universal probabilistic programming. J Mach Learn Res 20:973–978

Bommasani R, Liang P, Hudson D, Adeli E, Altman R, Arora SV, Etchemendy (2021) On the opportunities and risks of foundational models. arXiv preprint, http://arXiv.org/2108.07258

Brown P (1983) Error messages: the neglected area of the man/machine interface. Commun ACM 26(4):246–249

Brownlee J (2016) Deep learning with python: develop deep learning models on Theano and TensorFlow using Keras. In: Machine learning mastery

Buolamwini JG (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In: Proceedings of machine learning research, 77–91

Castelvecchi D (2016) Can we open the black box of AI? Nat News 538(7623):20–23. https://doi.org/10.1038/538020a

Demirović D, Skejić E, Šerifović-Trbalić A (2018) Performance of some image processing algorithms in Tensorflow. In: 25th International Conference on Systems, Signals and Image Processing (IWSSIP), 1–4. https://doi.org/10.1109/IWSSIP.2018.8439714

Economist (2020) Businesses are finding AI hard to adopt. Retrieved March 15, 2023, from https://www.economist.com/technology-quarterly/2020/06/11/businesses-are-finding-ai-hard-to-adopt

Economist (2022) What if all workers wrote software, not just the geek elite? *Economist*. Retrieved March 15, 2023, from https://www.economist.com/business/2022/01/29/what-if-all-workers-wrote-software-not-just-the-geek-elite?giftId=e827aa1a-7c4e-4ca0-85c3-f9d2953ab719

Fuller B (1938) Nine chains to the moon, Anchor Books, 1938, 1973. Anchor Books

Ganascia J-G (2015) Abstraction of levels of abstraction. J Exp Theor Artif Intell 27(1):23–35

Girard R (1965) Deceit desire and the novel. Johns Hopkins Press

Girard R (1988) Generative Scapegoating. Violent origins: Walter Burkert, René Girard, and Jonathan Z. Smith on Ritual Killing and cultural formation. Stanford University Press, pp 73–105

Girard R (2009) Battling to the end: conversations with Benoît Chantre. MSU Press

Girard R, Oughourlian J-M, Lefort G (2002) Things hidden since the foundation of the world. Stanford University Press

GitHub (2022a) *About GitHub*. Retrieved Feb 14, 2022a, from https://github.com/about

GitHub (2022b) *PyTorch*. Retrieved Feb 02, 2022b, from GitHub: https://github.com/pytorch/pytorch/releases/tag/v0.1.1

Google AI (2015) Google AI Blog. Retrieved 02 Apr 2022, from https://ai.googleblog.com/2015/11/tensorflow-googles-latest-machine.html

Havlíček V, Córcoles A, Temme K, Harrow A, Kandala A, Chow J, Gambetta J (2019) Supervised learning with quantum-enhanced feature spaces. Nature 567(7747):209–212. https://doi.org/10.1038/s41586-019-0980-2

Heylighen F (2003) Introduction to principia cybernetica. Retrieved February 09, 2022, from Principia Cybernitica Web: http://pespmc1.vub.ac.be/INTRO.html

Heylighen F (2007) Accelerating socio-technological evolution: from ephemeralization and stigmergy to the global brain. Globalization as evolutionary process. Routledge, pp 304–329

Holweg M, Younger R, Wen Y (2022).The reputational risks of AI. Calif Manag Rev Insights, 64(2)

Kaye K (2022) Should just anyone be given keys to the AI machine? Why low-code AI tools pose new risks. (protocol.com) Retrieved Mar 15, 2023, from https://www.protocol.com/enterprise/no-low-code-ai-risks

Keras (2016) Keras Blog. Retrieved Feb 14, 2022, from https://blog.keras.io/introducing-keras-10.html

Kiritchenko S, Mohammad S (2018) Examining gender and race bias in two hundred sentiment analysis systems. In: Proceedings of the 7th Joint Conference on Lexical and Computational Semantics (*SEM), New Orleans)

Kleinberg J, Mullainathan S, Raghavan M (2017) Inherent trade-offs in the fair determination of risk scores. In: 8th Innovations in Theoretical Computer Science Conference, pp. 43:1–43:23. https://doi.org/10.4230/LIPIcs.ITCS.2017.43

Ko AJ (2014) Programming languages are the least usable, but most powerful human-computer interfaces ever invented. Retrieved Feb 14, 2022, from https://medium.com/bits-and-behavior/programming-languages-are-the-least-usable-but-most-powerful-human-computer-interfaces-ever-invented-7509348dedc

Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105

Kurzweil R (2001) The law of accelerating returns. Retrieved Mar 15, 2023, from https://www.kurzweilai.net/the-law-of-accelerating-returns

Lamblin P (2017) Theano Users Google group. Retrieved Feb 14, 2022, from https://groups.google.com/g/theano-users/c/7Poq8BZutbY

Langer S (1953) Feeling and form: a theory of art developed from philosophy in a new key. Charles Scribner's Sons, New York

Last C (2017) Global commons in the global brain. Technol Forecast Soc Chang 114:48–64. https://doi.org/10.1016/j.techfore.2016.06.013

Marcus G (2018) Deep learning: a critical appraisal. arXiv preprint, https://arxiv.org/1801.00631. Retrieved Mar 15, 2023, from https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf

Maynard Smith J, Szathmary E (1997) The major transitions in evolution. Oxford University Press

Naizat G, Zhitnikov A, Lim L-H (2020) Topology of deep neural networks. J Mach Learn Res 21(1):7503–7532

Norwegian Consumer Council (2018) Deceived by design, how tech companies use dark patterns to discourage us from exercising our rights to privacy. Retrieved Mar 15, 2023, from https://fil.forbrukerradet.no/wp-content/uploads/2018/06/2018-06-27-deceived-by-design-final.pdf

Oneto N, Anguita L, Navarin A, Sperduti D (2019) Recent advances in big data and deep learning. In: Proceedings of the INNS Big Data and Deep Learning Conference INNSBDDL2019, held at Sestri Levante, Genova, Italy 16–18 April 2019

OpenAI (2023) Introducing ChatGPT. (openAI.com) Retrieved Mar 15, 2023, from https://openai.com/blog/chatgpt

Pearl J (2000) Causality: models, reasoning and inference, vol 19.2. Cambridge University Press, Cambridge

Powell A (2012) Democratizing production through open source knowledge: from open software to open hardware. Media Cult Soc 34(6):691–708

Press G (2020) 54 predictions about the state of data In: 2021. Retrieved Mar 15, 2023, from forbes.com: https://www.forbes.com/sites/gilpress/2021/12/30/54-predictions-about-the-state-of-data-in-2021/?sh=6af0ecb0397d

PyTorch (2018) PyTorch blog. Retrieved Feb 14, 2022, from https://pytorch.org/blog/a-year-in/

Russell B (2009) The ABC of relativity. Routledge

Sady, W. (2001). Ludwik Fleck–Thought Collectives and Thought Styles. In *Polish philosophers of science and nature in the 20th century* (pp. 197–205). Brill.

Shi S, Wang Q, Xu P, Chu X (2016) Benchmarking state-of-the-art deep learning software tools. In: 7th International Conference on Cloud Computing and Big Data (CCBD), 99–104. https://doi.org/10.1109/CCBD.2016.029

Szathmáry E (2015) Toward major evolutionary transitions theory 2.0. Proc Natl Acad Sci 112(33):10104–10111. https://doi.org/10.1073/pnas.1421398112

Szathmáry E, Maynard Smith J (1995) The major evolutionary transitions. Nature 374(6519):227–232. https://doi.org/10.1038/374227a0

Turchin V (1977) The phenomenon of science: a cybernetic approach to human evolution. Columbia University Press, New York

Wang T, Zhao J, Yatskar M, Chang K, Ordonez V (2019) Balanced datasets are not enough: estimating and mitigating gender bias in deep image representations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision

Wang X, Zhao Y, Pourpanah F (2020) Recent advances in deep learning. Int J Mach Learn Cybern 11(4):747–750

Wolff G, Anderson J, Viry P (2020) Europe has an artificial-intelligence skills shortage. Retrieved Mar 15, 2023, from https://www.bruegel.org/2020/08/europe-has-an-artificial-intelligence-skills-shortage/

Woolston C (2021) Stagnating salaries present hurdles to career satisfaction. Nature 599:519–521. https://doi.org/10.1038/d41586-021-03041-0

Yu E (2022) Singapore releases software toolkit to guide financial sector on AI ethics. Retrieved Mar 15, 2023, from https://www.zdnet.com/article/singapore-releases-software-toolkit-to-guide-financial-sector-on-ai-ethics/

Appendix 3

AI Implementation and Capability Development in Manufacturing: An Action Research Case

Jon Eklöf
University of Copenhagen, GKN Aerospace
jon.eklof@gmail.com

Ulrika Lundh Snis
University West
ulrika.snis@hv.se

Thomas Hamelryck
University of Copenhagen
thamelry@bio.ku.dk

Alexander Grima
GKN Aerospace
alexander.grima@gknaerospace.com

Ola Rønning
University of Copenhagen
ola@di.dk.ku

Abstract

This action research article presents a case study of a global manufacturing company deploying artificial intelligence (AI) to develop capabilities and enhance decision-making. This study explores considerations and trade-offs involved in introducing AI into daily operations, leading up to the decision to develop AI capabilities in-house or outsource them.

The case study offers in-depth technical descriptions of model selection, dataset creation, model adoption, model training and evaluation while addressing organizational obstacles and decision-making processes. The study's findings highlight the importance of collaboration between technical experts, business leaders, and end-users, as well as the interaction and collaboration between AI systems and human employees in the workplace.

The article contributes a practical perspective on AI implementation in manufacturing, emphasizing the need to balance in-house capability development with external acquisition. Although the case study company managed to create an in-house model, factors such as implementation, debugging, data requirements, training time, and performance led to outsourcing the capabilities. However, making this informed decision required capabilities and insights that were acquired through practical work. Consequently, although in-house development can be challenging, it can also enhance organizational capabilities and provide the necessary knowledge to make informed decisions about future development or outsourcing.

Keywords: AI capability development, AI implementation, AI in manufacturing

1. Introduction

The adoption of AI in manufacturing has received significant attention in both academia and industry in recent years (Zeba, Dabic, Čičak, & Daim, 2021). Most of the research on the implementation of AI in manufacturing has been conducted independently from daily operations (Arinez, Chang, Gao, Xu, & Zhang, 2020). Further investigations are needed to better understand the implications of integrating AI into daily operations and manage AI project and deployment risk. It also raises questions on how organizations should manage, staff and coordinate AI development. While contemporary AI research typically describes a computational solution to a specific problem, this article investigates the practical process of developing such a solution, the obstacles encountered, and the considerations made during the process. This approach is supported by (Amabile, 1996) and (Govindarajan & Trimble, 2012), who argue that the solution-finding process is important in enhancing creativity and innovation and overcoming complex problems.

When considering integrating AI into their operations, companies must decide whether to develop capabilities in-house or outsource them to external service providers (Ransbotham, Gerbert, Kiron, & Reeves, 2017). Outsourcing AI capabilities allows companies to quickly acquire necessary expertise but may result in a loss of competitive advantage provided by AI solutions, only gaining access to technology and competence available in the open market (Govindarajan & Immelt, 2019; Teesce, 2014). In-house development provides organizations with greater control over the development process, but this comes with a cost. Companies must therefore weigh the resources and time needed against the benefits of acquiring the same capability externally. There is no clear guidance on the decision-making process or criteria for making such decisions (Ransbotham, Gerbert, Kiron, & Reeves, 2017; Govindarajan & Immelt, 2019; Teesce, 2014).

This paper aims to investigate the implications, considerations, and trade-offs of introducing AI into daily operations of a manufacturing company, leading up to the decision of whether to develop AI capabilities in-house or outsource them and the factors that influence this decision. The case study focuses on the in-house development of an AI model for defect detection in X-rays of welds of aerospace components (the In-House Model). It provides IS professionals' accounts of managing challenges faced by contemporary corporations and organizations. Although the current study investigates capability development in one company only, it aims to generalize within the specific setting, focusing on theoretical abstraction and insights that are well-grounded in the selected case.

2. Identifying the problem domain

2.1. The case

In the current case, an aerospace component manufacturing company (the Company) was increasing production of a critical component, the turbine exhaust case (TEC), the final stage of a commercial jet engine. The TEC is vital for heat dissipation and performance. It is a complex load-bearing structure designed to withstand high temperatures and loads while also maintaining aerodynamic performance, low weight, and cost-effectiveness. The TEC has a diameter of approximately 1 meter and weighs roughly 75 kgs. It is built by welding several segments together, and each TEC comprises more than 100 welds that undergo inspection using X-rays, a method commonly used to detect defects such as gas pores and cracks in welds (Tyystjärvi et al., 2022).

Figure 1: Image of the turbine exhaust case - the component subject to weld inspection.

As the production rate was ramping up, an increasing amount of welds required inspection. The X-ray images of the TEC were digitally scanned, and each scan contained between two and four X-ray images of welds. Three operators spent thousands of hours inspecting hundreds of thousands of images per year. As pointed out by Bertović, such a manual inspection process is time-consuming, fail-prone, and operator-dependent (Bertović, 2016). In the current case, the process of training and certifying these inspectors was expensive and cumbersome, taking up to two years of training and practical experience for operators to become certified for inspection. Although minor defects were present, critical defects were rare and did not often render components defective. As a result, the monotony of inspecting welds and the rarity of critical defects increased the risk of human-related errors.

The Company had previously assessed automated inspection methods utilizing machine learning, but faced obstacles such as low or inconsistent contrast and brightness, and defect-like anomalies or geometries producing excessive false positives. These issues have been identified as recurring problems with automated inspection (Nacereddine, Zelmat, Belaïfa, & Tridi, 2005; Ronneberger, Fischer, & Brox, 2015). Despite having a database of hundreds of thousands of X-ray images, the Company lacked labeled data suitable for training AI models.

2.2. Related research

The study builds on previous research suggesting that machine learning techniques, including deep learning, can be used to detect defects in X-ray images (Bertovic & Virkkunen, 2021). To explore this further, a literature search was conducted on Google Scholar using the inclusion criteria "*defect detection,*" "*X-ray,*" and "*welds,*" in conjunction with the keywords "*machine learning*" and "*deep learning*" respectively. The search identified 18 articles investigating various AI techniques for detecting defects in X-ray images. These techniques include traditional machine learning methods such as support vector machines (Wang Y. e., 2008) as well as deep learning techniques such as segmentation networks (Tyystjärvi et al., 2022), generative adversarial networks (GANs) (Akcay, Atapour-Abarghouei, & Breckon, 2019; Guo et al., 2021), autoencoders (AEs) (Presenti et al., 2022), variational autoencoders (VAEs) (Banko et al., 2021; Lindgren & Zach, 2021), and, to the greatest extent, convolutional neural networks (CNNs) (Jiang et al. L. , 2021; Yaping & Weixin, 2019; Wang, Shi, & Tong, 2019; Yang et al. D. , 2021; Deng et al., 2021; Naddaf-Sh et al., 2021; Ajmi et al., 2020; Wen-ming, 2019; Jiang et al. H. , 2021; He et al., 2017) (Yang et al. L. , 2021). However, to fully leverage the benefits that deep learning techniques provide, it is typically necessary to have a large data base of labeled examples (Presenti et al., 2022). Creating labeled training data for image analysis in non-destructive evaluation can be challenging due to the complex and often subtle nature of the features that need to be identified. Because manual labeling is labor-intensive and susceptible to errors or inconsistency between labels, it becomes increasingly problematic when labeling large data volumes. As a result, there are limited public data resources available for training machine learning models to perform this task (Tyystjärvi et al., 2022; Mery et al., 2015).

More recently, semi-supervised VAEs (SS-VAEs) have emerged as a tool for image classification in instances where, as in the current case, there are low amounts of labeled data available, but a fair amount of unlabeled data (Kong & Ni, 2020). However, our literature search revealed no studies investigating use of SS-VAEs for defect detection in X-ray images.

2.3. About the SS-VAE

The SS-VAE builds on the VAE, which, in turn, builds on the AE architecture. The AE contains an encoding function that maps an input to a compressed latent space representation and a decoding function that maps from the latent space back into the original space (Maheshwari, Mitra, & Sharma, 2022).

The VAE maps an input to a probability distribution over the latent space, allowing it to model complex and multimodal data distributions and generate new data by sampling the distribution (Kingma & Welling, 2014).

VAEs have been used in various domains, including image recognition and anomaly detection (Maheshwari, Mitra, & Sharma, 2022). In instances where labeled data is limited, it is possible to estimate the label using a SS-VAE. The SS-VAE decoder (i.e. the probabilistic model) represents missing labels as latent variables sampled from a prior distribution while the encoder (i.e. the approximating distribution or guide) infers missing labels from the image data. This approach has been used to improve classification in various domains with a limited amount of labeled data (Kingma & Welling, 2014; Wu et al., 2021).

3. Methodological framework

3.1. Action research

The study uses Mathiassen et al.'s action research model, which involves collaboration between practitioners and researchers to address real-world issues (Mathiassen, Chiasson, & Germonprez, 2012). Put simply, action research is a "learning by doing" process where a group of individuals identifies a problem, plans how to tackle the problem, takes action to resolve it, evaluates the success of their efforts, and iterates as necessary (Susman, 1983; O'Brien, 1998). The distinguishing feature of action research compared to professional practices is its scientific approach, wherein the problem is systematically studied, and the intervention is grounded in theoretical considerations (O'Brien, 1998). Through collaboration between practitioners and researchers, this study seeks to contribute to solving the Company's problem of establishing necessary capabilities while addressing the needs identified in current research, how to identify necessary capabilities and optimal ways to develop them.

Furthermore, this study adopts a practice-based research approach that leverages the main author's expertise in leading digital transformation, enabling the development of necessary capabilities and constant reflection on and refinement of the practice (Candy, 2006). By adopting this approach, the study aims to not only address the Company's problem but also to contribute to the academic literature on capability development in practice-based contexts.

3.2. Agile approach with experiment-first-mindset

Rather than adhering to a strict plan, setting out each step and the requirements of the end product, the study utilized an agile approach, emphasizing experimentation and creativity in the project. The project was executed as a series of iterative experiments, refining the project design based on continuous feedback, collaboration, and flexibility, which are important components both in action research and in addressing the challenges of AI projects (Sousa, 2019; Mukherjee, 2020; Steiber, Alänge, Ghosh, & Goncalves, 2020; Curcio et al., 2018).

3.3. Action Planning

In this case, the Company's senior leadership requested their in-house digital innovation team (the Team), led by the main author, to develop a model in-house (In-house Model) that could be used as a benchmark against a proprietary model (Proprietary Model) from an external service provider specializing in detecting defects in images with limited labeled data. The Team had no insight into the Proprietary Model other than that it was partly trained in an unsupervised manner. The Company's goal was to make a decision within a few months, so the Team was reminded of the importance of time and tasked with creating a model that allowed them to benchmark the Proprietary Model. The Team had limited experience of developing advanced AI models in general and models for defect detection in particular and therefore needed to develop these capabilities.

The Team had a proved ability to implement digital solutions into various areas of the Company. However, given the Team's limited experience working with AI solutions, the study adopted an action research approach that involved forming a collaborative group of both external and internal stakeholders to provide a broader range of expertise and perspectives (Reason & Bradbury, 2001). The Team collaborated with external experts in deep probabilistic programming (the Experts) and internal X-ray inspection operators (the Operators) who provided insights and recommendations based on their expertise. The Experts and Operators also served as a benchmark for evaluating the effectiveness of the interventions implemented during the research process.

As the Team leader, the study's main author was responsible for AI capability development within the

Company. In this study, he worked alongside an industry programmer to perform the practical work, enabling practice-based research. Simultaneously, the main author researched leadership aspects of AI implementation in the manufacturing industry allowing for the immediate application of research findings in the field.

To further conceptualize the results, ensuring they were research-grounded and contributed to the academic discourse, the main author collaborated with the second author and engaged with the research community. This approach resulted in insights that enhanced the practical experiences gained in the manufacturing industry, making the results both practically and scientifically grounded.

4. Taking action

This section describes the chronological order of the different stages of the practical work being undertaken, namely model selection, dataset creation, model development, model training, and model evaluation.

4.1. Selecting the In-house Model

The Company's senior leadership had requested the Team to develop an In-house Model that could serve as a benchmark against the Proprietary Model. However, the leadership did not provide explicit criteria for the In-House Model, apart from its role as a benchmark. Faced with a lack of clear guidelines, the Team collaborated with the Experts to determine the properties the model should possess to function as a benchmark. Considering the Proprietary Model's ability to operate with limited labeled data, and the fact that there was a fair amount of unlabeled data available, the Experts advised the Team to develop a model that could efficiently utilize unlabeled data without relying on extensive labeled data. With this in mind, they recommended the Team to construct a SS-VAE. For a comprehensive description of the SS-VAE, please see Kingma and Welling, 2014).

While creating a model from scratch may have resulted in better performance, the Team faced time constraints. To expedite the development process and meet the leadership's goal of making a decision soon, the Experts recommended that the team adopt a publicly accessible model (the Baseline model) implemented in the deep probabilistic programming language Pyro (https://pyro.ai/examples/ss-vae.html). This approach, they suggested, would provide an adequate benchmark for comparison purposes.

4.2. Creating the dataset

The inspection criteria for the TEC state that several minor defects, such as gas pores located within a specific area, could render the component defective. In consultation with the Operators, the Team identified that the In-House Model needed to be able to process 256x256 pixel images to capture all known types of defects, based on the inspection criteria.

According to the external service provider, around 20 images per class were required to effectively showcase the capabilities of the Proprietary Model. The Experts advised that for an SS-VAE, 20 labeled images per class was low. The Baseline model was trained using a varied range of labeled MNIST images (28x28 pixel black and white representations of the handwritten digits 0-9), spanning from 100 to 3000 instances (Pyro, 2017). This means that there were approximately 10 to 300 training images available per individual handwritten digit. It is worth noting that the MNIST images were considerably smaller compared to the 256x256 pixel images intended for use in the In-house Model.

Although the lower number of labeled training images in the Baseline model potentially could have an adverse effect on the accuracy of the In-house Model, the Team deemed that an important aspect of benchmarking the Proprietary Model would be to use the same amount of labeled training data. The Team therefore decided to use 20 around labeled images per class for training.

Together with the Operators, the Team generated a labeled dataset from 10 full-size scans (Scans), containing between four to five X-ray images of welds (Weld-images) of the TEC. For each Weld-image, the regions that could contain defects and the defects in each region were marked. From these regions, the Weld-images were divided into 256x256 pixel images (Sub-images).

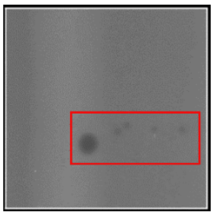

Figure 2: An example of a 256x256 pixel Sub-image with marked defects.

From these Sub-images the Team created a labeled training dataset containing 21 defect and 21 non-defect Sub-images. This was the dataset that was provided to the external consultant under a confidentiality agreement to train the Proprietary Model. Due to the low amount of labeled training data, the Team augmented the Sub-images to increase the number of data points in the training dataset. By flipping the Sub-images, a commonly known method to increase the number of data points (Shorten & Khoshgoftaar., 2019), the Team generated an additional three images per labeled image. This resulted in a total of 168 Sub-images for the training dataset - 84 Sub-images per class. The Team decided together with the Experts to leverage the SS-VAE's ability to use unlabeled data for training.

The Baseline model used a ratio of 1:500 to 1:16 between labeled and unlabeled images. The Experts initially recommended similar ratios for the project. However, due to the differences in size between 28x28 pixel MNIST images used in the Baseline model and the 256x256 pixel Sub-images in the current project, the Team recognized that using the same ratios as the Baseline model would significantly increase the training time of the model, to the point of being infeasible (for example, a ratio of 1:500 would have resulted in a training time of over eight years).

The Team thus faced a trade-off between model optimization and speed of execution. Using more labeled images would likely increase the accuracy of the In-House Model but also increase the training time. In consultation with the Experts, the Team ultimately opted for a ratio of 1:10 between labeled and unlabeled Sub-images. This decision was motivated by the fact that the labeled dataset already contained a sufficient number of training examples per class, and that a ratio of 1:10 was deemed to provide enough information to be able to evaluate the In-House Model.

The Team thus generated an unlabeled dataset of 1680 Sub-images using the same process as the labeled training data, excluding augmentation.

4.3. Reducing dimensionality

The Sub-images initially had a 16-bit grayscale depth. The Experts advised mapping the pixel values to a categorical distribution. Instead of representing each pixel's intensity as a numerical value, it would be transformed into a probability distribution across categories, where each category represented a specific intensity level. Unfortunately, this approach caused a memory leak on the GPU due to a bug in Pyro - the software framework used to build the In-house Model (Pyro, 2022). The Team spent weeks trying to solve this issue. Eventually, the Team sought help from the Pyro development team who suggested using a previous version of Pyro, which inexplicably solved the memory leak.

However, after this issue was solved, the Team realized that, despite using powerful GPUs, training the In-House Model on the 16-bit grayscale Sub-images would take over six months, which was unfeasible given the timeline provided by the Company leadership. Consequently, the Team explored various methods of reducing the dimensionality of the images.

Hou et al. have emphasized the importance of pre-processing high-dimensional data to reduce the complexity of classification tasks and computational costs (Hou et al., 2020). To determine the necessary information level for Operators to detect potential defects in the Sub-images, the Team consulted with them and concluded that black and white images would suffice for capturing all defects. To achieve this, the Team used OpenCV, a Python framework, and the CV2 module, along with an adaptive binary Gaussian model, to reduce the dimensionality of the images and convert them to black and white (OpenCV, 2022).

The adaptive binary Gaussian algorithm sets pixel thresholds using surrounding regions, accommodating images with varying illumination. However, given the variability in optimal thresholds for different images, the Team adjusted them according to the weld inspection criteria.

Noise and artifacts in images complicated the conversion process, potentially impacting the model's accuracy. Thus, the Tam and Operators jointly reviewed all training images to ensure no defects were missed during conversion.

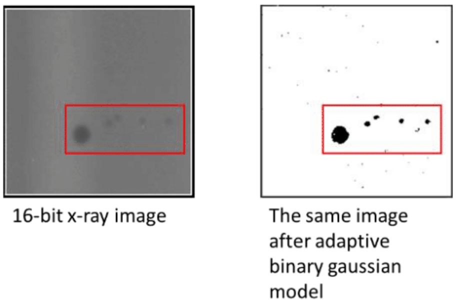

Figure 3: A comparison of a Sub-image with 16-bit grayscale depth and its conversion to black and white using an adaptive binary Gaussian model.

4.4. Adopting and training the In-house Model

The Team adopted the Baseline model, a semi-supervised variational autoencoder using the Python programming framework Pyro. Python code for the Baseline model can be found at https://github.com/pyro-ppl/pyro/blob/dev/examples/vae/ss_vae_M2.py. The In-house Model consisted of an encoder, a latent space, a classifier and a decoder.

The encoder utilized four tiers of convolutional and max pooling layers, each with 4, 8, 16, and 16 channels, respectively. These layers were defined by a kernel size of 3, a stride of 1, and a pooling size of 2. Following these layers, three fully connected layers were employed, having 3136 and 1000 nodes, and a pair of parallel layers each with 75 nodes. Softplus was used as the activation function for both the convolutional and fully connected layers. The encoder configured a 75-dimensional latent space, where each dimension was represented by a Gaussian distribution.

The classifier, sharing the same architecture as the encoder, differed by using 500 hidden units in the second fully connected layer. The output, rather than being mapped to a 75-dimensional space, was linked to two separate classes. The final layer of the classifier implemented a Softmax activation.

As for the decoder, it maintained a structure inverse to that of the encoder. It was composed of three sets of fully connected layers, followed by four tiers of upscale and transpose convolutional layers. The Sigmoid function was used for the final activation of each value, with the probability distribution of a single pixel denoted by a Bernoulli distribution.

In the adoption process, the latent space was a key consideration. Contrary to the approach suggested by e.g. Li et al., which advocates for meticulous optimization when configuring AI models (Li, Swersky, & Zemel., 2015), the Team, in collaboration with the Experts, opted for a less laborious approach. Instead of conducting an exhaustive search, they made an informed decision to expand the latent space from 50 as used in the Baseline model to 75 dimensions. This decision was influenced by considerations such as the size and complexity of the images.

To enhance the model's training time and ensure high defect detection accuracy in the Sub-images, convolutional neural networks (CNNs) were incorporated. Recognized for their proficiency in image analysis and pattern recognition tasks, CNNs can assign significance to spatial relationships within the data, capturing image-specific features while minimizing the parameter count (O'Shea and Nash 2015). The Team experimented with various CNN configurations, eventually integrating four sets of convolutional and max pooling layers, followed by three fully connected layers into the encoder. This decision was informed by the need to balance model complexity, computational efficiency, and image resolution.

4.5. Model training

Further delays were incurred by the global semiconductor shortage, which resulted in a lack of powerful hardware for model training, as well as specific IT department hardware and software requirements, before the Team was able to start training the In-house Model. The encoder was trained using mini-batches of 300 Sub-images. The encoder transformed the Sub-image's 65,536 pixels into the 75 dimension latent space. The classifier then categorized the Sub-image as defective or not, leveraging both labeled and unlabeled Sub-images for training and label estimation, respectively. The decoder, using the latent representation of a Sub-image and its estimated label, reconstructed the Sub-image, outputting 65,536 pixel values of 0 or 1 due to the Sub-images' conversion to black and white.

Prior to the In-house Model training, the Team conducted initial experiments using a fully supervised VAE with the labeled training dataset of Sub-images. Given that the VAE converged after approximately 85,000 epochs, the Team decided to train the In-house Model for the same number of epochs.

4.6. Evaluating performance and training time

The common evaluation measurements for these kinds of models are accuracy $\frac{true\ positives+true\ negatives}{total\ number\ of\ predctions}$, precision $\frac{true\ positives}{true\ positives+false\ positives}$ and, recall $\frac{true\ positives}{true\ positives+false\ negatives}$ (Tyystjärvi et al., 2022). Hence, these were used as evaluation measurements of the In-house Model after 85 000 epochs. A summary of the results can be found in Table 1. The results on accuracy, precision and recall of the Proprietary Model were received from the external consultant within a week after providing the training data.

Table 1: A comparison of training time, accuracy, precision, and recall between the In-house Model and the Proprietary Model

Model	Accuracy	Precision	Recall
In-house Model	75%	96%	46%
Proprietary Model	87%	100%	100%

As can be seen in Table 1, the In-house Model achieved an accuracy of 75% and a precision of over 90%, but had a recall of only 46%, indicating that it often misclassified images as not containing defects. The acceptance criteria for the TEC requires the detection of all defects. A recall as low as 46% would render it unsuitable for use in production. Even with additional data and training, it was deemed unlikely that the recall of the In-house Model would reach the required 100%. The Proprietary Model achieved an accuracy of 87% with 100% precision and recall.

The Baseline model used the MNIST dataset consisting of 60 000 28x28 pixel images. However, when the Team applied the In-house Model to its Sub-image dataset of 256x256 pixel images, even though it consisted of only 1848 images, the training time increased significantly. Despite training the In-house Model for nine weeks and 85,000 epochs, the reconstruction loss had not yet converged. The external consultant provided the results of the Proprietary Model within a week after receiving the training data. Even though the Proprietary Model was not perfect and misclassified some images as containing defects, the Team concluded that both the performance and training time of the Proprietary Model was superior to the In-house Model and that further work on the In-House Model therefore should be discontinued.

5. Lessons learned

In the work of identifying and developing necessary capabilities of AI implementation, the hands-on work of developing a solution and dealing with the obstacles and considerations encountered along the way eventually proved to be more important than the solution itself. As Govindarajan and Trimble and have emphasized, the process of finding solutions is crucial for enhancing creativity, innovation, and solving complex problems (Govindarajan & Trimble, 2012).

The development of the In-house Model revealed that a range of capabilities were required that were not anticipated when starting the project. This illustrates the difficulty of predicting the necessary capabilities beforehand, the importance of hands-on work for capability development and the need of investing in diverse skill development to cultivate a workforce capable of addressing the challenges and complexities of AI development. In the current case, these unforeseen capabilities were developed through a continuous learning process in close collaboration with the internal and external stakeholders as well as through formal and informal course work.

Below, we discuss key lessons learned in terms of capability development.

5.1. Lessons learned for the Team

To begin with, the Team needed to create an underlying data structure that could efficiently store and label the Sub-images. Once the defects in the Sub-images were marked, the Team had to develop a script that converted the marked regions of the Sub-images that contained defects into a dataset that could be merged with the remaining metadata from the Sub-images.

The Sub-images had a wide range of brightness levels, and their pixel values were represented in 16-bit grayscale. To convert them into black and white images, the Team employed thresholding, which involves setting a threshold value. However, determining the optimal threshold value for each image was challenging as it can vary depending on the image's characteristics and using a global thresholding approach could lead to information loss.

To tackle these challenges, the Team had to acquire new image processing skills to be able to handle the images' high dynamic range and the presence of noise and artifacts while identifying defects accurately. This required collaboration with Operators to gain a deep understanding of the inspection domain and from this understanding develop appropriate thresholding strategies to capture the necessary information accurately.

Further, to be able build the In-house Model, the Team had to acquire a basic understanding of the theoretical foundations of deep probabilistic models, including Bayesian statistics, variational inference, and deep generative models. Additionally, they needed to develop practical capabilities in hyperparameter tuning and distribution selection, crucial components of configuring the SS-VAE architecture.

The In-house Model relied on unlabeled data to learn the underlying distribution of the input data. The Team had to make informed decisions and trade-offs regarding the size and ratios of the training dataset to ensure that the SS-VAE architecture was robust and generalizable to new data.

5.2. IT infrastructure and capabilities

The project not only enhanced the capabilities of the Team but also provided valuable insights for the IT department within the Company. IT infrastructure and capabilities proved essential for the effective development of the In-house Model. The development required a robust IT infrastructure, comprising appropriate hardware and software tools.

As the Team developed the In-house Model, the IT department gained knowledge on the current hardware and software requirements. They had to gain expertise in selecting appropriate hardware for the project, requiring an understanding of the latest advancements in hardware technologies, such as GPU architectures.

In addition to high-performance hardware, the Team required specific programming frameworks - PyTorch, Pyro and OpenCV. The IT department thus had to acquire knowledge about these programming frameworks, ensuring compatibility with hardware and assessing support and documentation availability. The project provided an opportunity for the IT department to develop the necessary capabilities to support future AI projects.

6. Implications and generalizability

One of the limitations of practice-based research is its generalizability to wider populations. However, generalization across cases is not always necessary (Geertz, 1973). Instead, generalization can be achieved by generalizing within individual cases, identifying theoretical abstractions, and using these insights to generalize to theory (Lee & Baskerville, 2003). The current study investigates capability development in one company only. The intent of this in-depth engagement with a particular case, and with a particular technology, extends beyond the mere application of statistical generalizability across cases.

This research intends to provide theoretical insights that are applicable to the particular case under study and that contribute to a more generalized understanding of AI capability development. Below we discuss the implications and contributions based on the findings of the study and in light of current research.

6.1. Balancing cost and benefit of in-house solutions

The Company found that an in-house solution was not viable due to challenges with Sub-images, long training times, and the superior performance of the Proprietary Model. As highlighted by Ransbotham and others, this emphasizes the need to weigh the pros and cons of in-house AI development, considering expertise, resources, and time for model creation and validation (Ransbotham, Gerbert, Kiron, & Reeves, 2017; Govindarajan & Immelt, 2019; Teesce, 2014). In some cases, it may be more effective and efficient to leverage existing AI solutions or to outsource AI development to external service providers who can offer specialized expertise and resources. Although developing the In-House Model was challenging, such efforts can strengthen the capabilities of organizations, enabling them to better weigh the pros and cons of future in-house development or outsourcing opportunities.

6.2. Collaboration with internal and external stakeholders

The Team overcame many of its challenges through collaboration with the Experts and Operators. Firstly, this points to the importance of having a proactive mindset and the ability to recognize one's limitations to identify the need for further knowledge. Secondly, it shows the significance of seeking internal or external expertise to gain additional knowledge, overcome challenges, and advance the organization's capabilities. Collaboration is an essential component of building in-house AI capabilities (Govindarajan & Immelt, 2019; Mukherjee, 2020; Ancona, 2019). Encouraging open communication and knowledge sharing across departments, as well as with external partners, strengthens collaborative capabilities of employees, management, and external experts, which in turn can help enhance creativity and innovation, as well as allowing organizations to overcome complex problems (Govindarajan & Trimble, 2012).

6.3. Agile and experimental approach

By adopting an agile and experimental approach that prioritized flexibility and collaboration, the Team made progress and quickly adapted to changing requirements and unforeseen obstacles. This approach aligns with the works of several authors who have emphasized the importance of agility and experimentation in AI development (Sousa, 2019; Mukherjee, 2020; Steiber, Alänge, Ghosh, & Goncalves, 2020). Continuous feedback from all stakeholders enabled the Team to refine the project design and improve model performance iteratively, ensuring that the final product met the needs of the Company leadership. Moreover, by emphasizing creativity and experimentation, the Team was able to use innovative solutions to overcome complex problems, such as using OpenCV for dimensionality reduction.

6.4. Recognizing the value of human expertise in AI development

Subject matter expertise proved integral to the AI development process. The Operators' domain-specific knowledge about X-ray inspection was crucial in developing both the dataset and the model in a way that accurately detected defects. As suggested by Fountaine and Govindarajan, by leveraging subject matter experts' knowledge and expertise, organizations can develop AI models that are more accurate, effective, and useful, ultimately benefiting both workers and the organization as a whole (Fountaine, McCarthy, & Saleh, 2019; Govindarajan & Immelt, 2019).

6.5. AI Models as tools to augment human labor

Further, by involving such internal subject matter experts, organizations can alleviate concerns about obsolescence and demonstrate the value of workers in the AI development process (Babic, Chen, Evgeniou, & Fayard, 2020; Fountaine, McCarthy, & Saleh, 2019; Schepman & Rodway, 2020). In the present case, it soon became clear that neither modelthe In-house Model or the Proprietary Model fully could replace human capabilities in detecting defects in X-ray images. This experience illustrates the importance of recognizing the limitations of AI models and leveraging them as tools to augment human labor rather than replace it entirely. Incorporating AI into the workforce can improve job satisfaction by enabling workers to focus on higher-level tasks that require human expertise, rather than performing tedious and repetitive tasks (Babic, Chen, Evgeniou, & Fayard, 2020; Fountaine, McCarthy, & Saleh, 2019; Schepman & Rodway, 2020). By emphasizing the collaborative relationship between AI models and humans, organizations can foster a culture that values the contributions of both and maximize the benefits of AI in the workforce.

7. Conclusions and future research

This action research article explores the process of implementing AI in the manufacturing industry and developing AI capabilities. It offers practical insights into the technical aspects and trade-offs between training time and optimal performance. Given the time-sensitive nature of many projects, efforts to reduce training time of AI models can lead to suboptimal solutions. However, in a business context, such trade-offs are sometimes necessary to meet operational targets. Further, while external acquisition can offer quick access to expertise, in-house development can provide control and potential competitive advantages, Even though in-house development can prove challenging, we argue that such efforts can strengthen organizational capabilities and enable informed decisions about future in-house development or outsourcing.

The study also contributes to theory on AI implementation, confirming the need for a balanced evaluation of in-house versus outsourced solutions, considering costs, expertise, and performance. It emphasizes the importance of collaboration with internal and external stakeholders as well as researchers, agile and experimental methodologies, and the integration of human expertise in AI development. Furthermore, the study highlights the role of AI as a tool to augment human labor, adding to the discourse on human-machine collaboration, organizational strategy, and AI capability development.

Future research considering these theoretical implications in other organizations would allow for generalizations both within and across cases. In particular, we suggest further research investigating the relationships between in-house and outsourced AI development, exploring how different industries, organizational sizes, or technological complexities influence the decision-making process. Additionally, studies examining human-AI collaboration across various sectors could provide insights into optimizing the blend of human expertise and AI, potentially leading to new models for organizational efficiency, innovation, workforce satisfaction, and capability development.

8. References

Ajmi et al., C. (2020). Deep learning technology for weld defects classification based on transfer learning and activation features. *Advances in Materials Science and Engineering* .

Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2019). Ganomaly: Semi-supervised anomaly detection via adversarial training. *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14.* Perth: Springer International Publishing.

Amabile, T. M. (1996). *Creativity and innovation in organizations.* Boston: Harvard Business School.

Ancona, D. (2019). Five Rules for Leading in a Digital World. *MIT Sloan Management Review, 61*(1), 1-4.

Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., & Zhang, J. (2020). Artificial Intelligence in Advanced Manufacturing: Current Status and Future Outlook. *Journal of Manufacturing Science and Engineering, 142*.

Babic, B., Chen, D. L., Evgeniou, T., & Fayard, A.-L. (2020). A better way to onboard AI. *Harvard Business Review, 98*(4), 56-65.

Banko et al., L. (2021). Deep learning for visualization and novelty detection in large X-ray diffraction datasets. *Npj Computational Materials, 7*(1), 104.

Bertović, M. (2016). *Human factors in non-destructive testing (NDT): risks and challenges of mechanised NDT.*

Bertovic, M., & Virkkunen, I. (2021). NDE 4.0: New Paradigm for the NDE Inspection Personnel. In *Handbook of Nondestructive Evaluation 4.0* (pp. 1-31). Springer.

Candy, L. (2006). *Practice Based Research: A Guide.* Sydney: Creativity & Cognition Studios, University of Technology.

Curcio et al., K. (2018). Requirements engineering: A systematic mapping study in agile software development. *Journal of Systems and Software, 139*, 32-50.

Deng et al., H. (2021). Industrial laser welding defect detection and image defect recognition based on deep learning model developed. *Symmetry, 13*(9), 1731.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. *Harvard Business Review, 97*(4), 62-73.

Geertz, C. (1973). *The interpretation of cultures* (Vol. 5019). Basic books.

Govindarajan, V., & Immelt, J. R. (2019). The Only Way Manufacturers Can Survive. *MIT Sloan management review, 60*, 24-33.

Govindarajan, V., & Trimble, C. (2012). *Reverse innovation: Create far from home.* Boston: Win Everywhere.

Guo et al., R. (2021). Weld defect detection from imbalanced radiographic images based on contrast enhancement conditional generative adversarial network and transfer learning. *IEEE Sensors Journal, 21*(9), 10844-10853.

He et al., K. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision.*

Hou et al., W. (2020). Review on computer aided weld defect detection from radiography images. *Applied Sciences, 10*(5), 1878.

Jiang et al., H. (2021). Convolution neural network model with improved pooling strategy and feature selection for weld defect recognition. *Welding in the World, 65*, 731-744.

Jiang et al., L. (2021). Casting defect detection in X-ray images using convolutional neural networks and attention-guided data augmentation. *Measurement 170*, 108736.

Kingma et al., D. (2014). Semi-supervised learning with deep generative models. *Advances in neural information processing systems, 27*.

Kingma, D., & Welling, M. (2014). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114* .

Kong, Y., & Ni, D. (2020). A semi-supervised and incremental modeling framework for wafer map classification. *IEEE Transactions on Semiconductor Manufacturing, 33*(1), 62-71.

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information systems research, 14*(3), 221-243.

Lee, J., Suh, T., Roy, D., & Baucus, M. (2019). Emerging technology and business model innovation: the case of artificial intelligence. *Journal of Open Innovation: Technology, Market, and Complexity, 5*(3), 44.

Li, Y., Swersky, K., & Zemel., R. (2015). Generative moment matching networks. *International conference on machine learning*.

Lindgren, E., & Zach, C. (2021). Autoencoder-based anomaly detection in industrial x-ray images. *Quantitative Nondestructive Evaluation, 85529*.

Maheshwari, A., Mitra, P., & Sharma, B. (2022). Autoencoder: Issues, Challenges and Future Prospect. *Recent Innovations in Mechanical Engineering* (pp. 257-266). Springer.

Mathiassen, L., Chiasson, M., & Germonprez, M. (2012). Style Composition in Action Research Publication. *MIS Quarterly, 36*(2), 347-363.

Mery et al., D. (2015). GDXray: The database of X-ray images for nondestructive testing. *Journal of Nondestructive Evaluation, 34*, 1-12.

Mukherjee, A. S. (2020). 5 Musts for Next-Gen Leaders. *MIT Sloan Management Review, 61*(3), 1-4.

Nacereddine, N., Zelmat, M., Belaïfa, S. S., & Tridi, M. (2005). Weld defect detection in industrial radiography based digital image processing. *3rd International Conference: Sciences Of Electronic, Technologies Of Information And Telecommunications.*

Naddaf-Sh et al., M.-M. (2021). Defect detection and classification in welding using deep learning and digital radiography. *Fault diagnosis and prognosis techniques for complex engineering systems*, 327-352.

O'Brien, R. (1998). An overview of the methodological approach of action research.

OpenCV. (2022). *OpenCV Image Thresholding*. Retrieved Sep 26, 2022, from https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html)

Presenti et al., A. (2022). Automatic anomaly detection from X-ray images based on autoencoders. *Nondestructive Testing and Evaluation, 37*(5), 552-565.

Pyro. (2017, November 3). *The Semi-Supervised VAE.* Retrieved May 04, 2023, from https://pyro.ai/examples/ss-vae.html

Pyro. (2022). *Pyro Deep Universal Probabilistic Programming*. Retrieved Sep 16, 2022, from https://pyro.ai/

Ransbotham, S., Gerbert, P., Kiron, D., & Reeves, M. (2017). Reshaping Business With Artificial Intelligence. *MIT Sloan Management Review, 59*(1).

Reason, P., & Bradbury, H. (2001). *The SAGE Handbook of Action Research, Participative Inquiry and Practice.* Sage.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer.

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, 100014.

Shorten, C., & Khoshgoftaar., T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data, 6*(1), 1-48.

Sousa, M. J. (2019). Skills for disruptive digital business. *Journal of Business Research*, 257-263.

Steiber, A., Alänge, S., Ghosh, S., & Goncalves, D. (2020). Digital transformation of industrial firms: an innovation diffusion perspective. *European Journal of Innovation Management, 24*(3), 799-819.

Susman, G. I. (1983). Action research: a sociotechnical systems perspective. *Beyond method: Strategies for social research, 95*(113), 95.

Teesce, D. (2014). The Foundations of Enterprise Performance: Dynamic and Ordinary Capabilities in an (Economic) Theory of Firms. *The Academy of Management Perspectives, 28*(4), 328–352.

Tyystjärvi et al., T. (2022). Automated defect detection in digital radiography of aerospace welds using deep learning. *Welding in the World, 66*(4), 643-671.

Wang, Y. e. (2008). Detection of line weld defects based on multiple thresholds and support vector machine. *Ndt & E International, 41*(7), 517-524.

Wang, Y., Shi, F., & Tong, X. (2019). A welding defect identification approach in X-ray images based on deep convolutional neural networks. *Intelligent Computing Methodologies: 15th International Conference, ICIC 2019, Nanchang, China, August 3–6, 2019, Proceedings, Part III 15.* Springer International Publishing, 2019.

Wen-ming, G. U.-F. (2019). Welding defect detection of x-ray images based on faster r-cnn model. *Journal of Beijing University of Posts and Telecommunications, 42*(6), 20.

Wu et al., X. (2021). A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. *Mechanical Systems and Signal Processing, 149*, 107327.

Yang et al., D. (2021). Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence, 35*(15), 1237-1249.

Yang et al., L. (2021). Inspection of welding defect based on multi-feature fusion and a convolutional network. *Journal of Nondestructive Evaluation, 40*, 1-11.

Yaping, L., & Weixin, G. (2019). Research on X-ray welding image defect detection based on convolution neural network. *Journal of Physics: Conference Series. 1237.* IOP Publishing.

Zeba, G., Dabic, M., Čičak, M., & Daim, T. (2021). Technology mining: Artificial intelligence in manufacturing. *Technological Forecasting and Social Change, 171*.