CHALMERS
UNIVERSITY OF TECHNOLOGY

Defining Characteristics of Sustained Competitive Advantage in Machine Learning Systems

Master's Thesis in the Master's Programme
Entrepreneurship and Business Design

Lisa Rythén Larsson
Christina Thulin

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2016
Report No. E 2016:083

MASTER’S THESIS E 2016:083

Defining Characteristics of Sustained Competitive Advantage in Machine Learning Systems

LISA RYTHÉN LARSSON
CHRISTINA THULIN

Tutor, Chalmers: Bowman Heiden
Tutor, PARC: Markus Larsson

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2016

Defining Characteristics of Sustained Competitive Advantage in Machine Learning Systems

LISA RYTHÉN LARSSON, CHRISTINA THULIN

Master's Thesis E 2016: 083

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Telephone: + 46 (0)31-772 1000

Chalmers Reproservice
Gothenburg, Sweden 2016

Abstract

In the wake of the advances in ICT, the rules of competition have changed. Customers have more choices and supply alternatives are more transparent (Teece, 2010). The rules of competition are particularly unstable in high-technology environments (Christensen, 1997; D'Aveni et al., 2010), such as Machine Learning. In recent years, Machine Learning has transformed from being a technology field dominated by data scientist and a few organizations into a billion dollar market open for all type of organizations (Dambrosio, 2016). In fact, Machine Learning is believed to be the next wave of digital disruption where simply buying off-the-shelf Machine Learning tools will not suffice for firms to be competitive (Merrett, 2015). The control of strategically important resources is central for a firm to stay competitive over time. Intellectual Property Rights have traditionally been deemed as an effective mechanism for preventing competitive duplication, however, many argue that Intellectual Property(IP) legislation has failed to keep pace with advances in technology (Allison et al., 2007; Davies, 2011).

The purpose of this study has been to provide a starting point for firms to design sustained competitive advantage for Machine Learning Systems (MLS). This was done by identifying strategically important resources and determine how these can be effectively controlled. A comparative design has been used where valuable resources and effective control mechanisms have been identified empirically through interviews with Machine Learning researchers, business practitioners as well as industry-and IP experts. The findings from interviews were validated in case studies of Netflix and Amazon, which are companies recognized for their leverage of MLS as well as long-term commercial success.

A main conclusion from this study is that delivering a superior user experience is a requisite for competitiveness. Although superior user experience is the result of a combination of factors, providing the user with cognitive relief is central for MLS applications. Trust, personalization, solving a customer problem and easy-to-use have all been identified as vehicles for cognitive relief, and subsequently the resources creating these utilities are valuable. Sustainability is most effectively created for MLS by using technical, contractual or secrecy although a dynamic and combinatorial approach to control has been emphasized. Finally, Data and Hypotheses set, Machine Learning algorithms and Validation technologies have been identified as resources that score high, both in terms of value and control, and thus likely the most important sources of sustained competitive advantage.

This study builds on the theory of the Resource-based View e.g. Barney (1991), Rumelt (1997) as well as theories provided by Petrusson (2004, 2015) and Osterwalder and Yves (2010), yet expanding the theories by deconstructing characteristics of sustained competitive advantage for MLS in particular. However, the characteristics of MLS and the disruptive environment create challenges for proving sustainability of competitiveness, and future validation is therefore suggested.

Acknowledgements

This master thesis is a result of an internship at PARC, a Xerox company in Palo Alto, California, U.S., conducted during the spring of 2016. We would like to express our gratitude for the opportunity to work in this environment and side-by-side with world class researchers and IP specialists. Additionally it has been an invaluable experience to see an organisation in the front-line of Intellectual Capital Management.

There are many we wish to thank for helping us with our thesis, without whom we could not have accomplished this research.

Firstly, we would like to thank employees at PARC, who continuously have contributed with inputs and insights. A special thanks to our supervisor at PARC, Markus Larsson, who introduced us to the research problem and continuously supported us throughout the process of the thesis. Additionally, we want to thank Jeremy Clark for his guidance and support in preparing this thesis.

Interviews have been central for the empirical investigations and we would therefore like to direct a special thank you to all parties that have participated and iterated with us.

Additionally, without the guidance and feedback from our faculty supervisor, Bowman Heiden, this thesis would never have come together the way it did - thank you.

Finally, we would like to thank friends and family for love and support during this spring.

Lisa Rythén Larsson & Christina Thulin, Gothenburg, June 2016

Nomenclature

Artificial Intelligence (AI) - is the field of study and development of computer systems where intelligent behavior is created that normally would require a human to perform. Examples of tasks include visual perception, speech recognition or translations between languages.

Application Program Interface (API) - specify how software programs should interact. API is a set of routines, protocols and tools that are used when programming graphical user interface components.

Big Data - relates to large sets of data that can be analyzed using computers reveling patterns, trends and associations.

Cloud Computing - is a way of storing, managing, and processing data using a network of remote servers hosted on the Internet, instead of a local server or personal computer.

Intellectual Property Rights (IPR) - legal rights assigned through patents, copyright and trademarks allowing the holder to execute monopoly for a specific period of time for the defined item.

Information & Communication Technology (ICT) - includes any communication device or application, both products and services, that will store, retrieve, manipulate, transmit or receive information electronically in any digital form.

Intangible Assets - are assets such goodwill, brand recognition or IPRs that do not have a physical form.

Intellectual Assets - are investments in brands, design, technology or creative works that can be collectively named intellectual property.

Intellectual Capital Management - is the management of valuable intangible assets of a business.

Intellectual Property (IP) - are results of creative work or inventions to which one can have a right in form of patents, copyright or trademark (IPRs).

Machine Learning - is a sub-field of AI where advanced programs give computers the ability to learn without being explicitly programmed.

Non-Disclosure Agreement (NDA) - is a contract by which one or more parties agree not to disclose confidential information that has been necessary to share in order to do business.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents a brief background of the field of study, the research problem and an introduction to Machine Learning Systems. Additionally the purpose of the study, research questions, scope and limitations, the disposition of the study and reading instructions are described.

1.1 Background

Recent advancements in Information and Communication Technology (ICT) have driven the creation of new industries, industry roles and opportunities for growth (Chaudhuri, 2012). The market for Machine Learning is an illustrative example of an industry that has grown radically as a result of advances in ICT (Dambrosio, 2016). In recent years, Machine Learning has transformed from being technology field dominated by data scientists and a few organizations into a $5-10 billion dollar market open for all type of organizations (Dambrosio, 2016). In fact, Machine Learning is considered to be the next wave of digital disruption where simply buying off-the-shelf Machine Learning tools will not suffice to be competitive (Merrett, 2015).

In the wake of the advances in ICT, the rules of competition have changed. Customers have more choices and supply alternatives are more transparent (Teece, 2010). The rules of competition is particularly unstable in high-technology environments (Christensen, 1997; D'Aveni et al., 2010), such as Machine Learning. This instability has had implications on the sustainability of competitive advantage; recent studies have suggested that sustainable competitive advantage is in fact rare and declining in duration (D'Aveni et al., 2010). This is reflected by that industry leaders are more frequently over-throned (D'Aveni et al., 2010). There is a growing body of research arguing that these high velocity and disruptive environments never reach maturity, but simple innovate, cannibalize and selfreproduce and by doing so recreating the initial stages of different waves of industry and product life cycles (Christensen, 1997).

The core of value creation and sustained competitive advantage in any economy is control (Petrusson, 2004; Wang et al., 2009). Sustained economic performance of a firm is created by its inimitable and non-substitutable resources which prevent rivals from replicating the value (Wang et al., 2009). Hence, the sustainability of a firm's competitive advantage is directly related to the strength of various control mechanisms (Petrusson, 2004; Rumelt, 1997). It then follows that in the knowledge economy, sustained competitive advantage is not created by a firm's knowledge or innovations per se, but rather the control of knowledge and innovations (Petrusson, 2004).

Patents, trademarks and other type of Intellectual Property Rights (IPRs) are one type of control mechanism that is commonly employed to prevent rivals from imitating knowledge, innovations and other intangible assets without financially compensate the creator (Rumelt, 1997). However, the IP regime was created in an era dominated by physical embodiments of knowledge, whereas today, more and more firms embed their value propositions in virtual products and services (Petrusson, 2004). Intellectual property law is intended to reward cutting edge technology, however, many argue that legislation has failed to keep pace with advances in technology (Allison et al., 2007; Davies, 2011) .

An example of technology field that has created debates is Machine Learning. The main difference between Machine Learning and other software embedded technologies is that after implementation, the software generates its own code, i.e. "learns". There are multiple aspects to Machine Learning that have made practitioners and researchers questioning the effectiveness of IPRs as mean to creating control and subsequently appropriate long-term value. First of all, Machine Learning is software-based, a field that have been controversial for decades (Allison et al., 2007). Software is not considered a patentable invention in all jurisdictions, and the scope and quality of software patents are often questioned (Allison et al., 2007). Recent empirical studies suggest that software IPRs are increasingly used strategically, creating webs of cross-licensing agreements (Allison et al., 2007). It then follows that established firms have access to the majority of technologies, and compete through other means of control (Allison et al., 2007). Furthermore, the patent process is very long relative to the development and life cycle of software (Cockburn and MacGarvie, 2011). Thus, when a patent finally is granted, the code may be obsolete. Copyright and database protections are other IPRs that are used for software (Petrusson et al., 2010). These are generally considered to provide effective protection against copying but relatively easy for rivals to invent around (Allison et al., 2007). In addition to the general objections to software, the "learning" feature of Machine Learning adds further complexity. IPRs are viewed as products of the human mind, however, in the case of Machine Learning the computer is a co-innovator or a co-author to the final product (Davies, 2011). Historically, the dilemma of authorship/inventorship in AI and Machine Learning have been solved in court by deeming the computer to be a "tool" (Davies, 2011). However, as Machine Learning technologies become increasingly intelligent, it is only a matter of time until the validity of granted Machine Learning IPRs are questioned in court.

1.1.1 Research Problem

Machine Learning has evolved from being a high-technology field mastered by a few, to a technology field employing lots of people to create additional value. It is clear that the ability to innovate is central for companies to compete on the Machine Learning market. It is also clear that control mechanisms are requisites to sustained competitive advantage. It then follows that the control of strategically important resources related to Machine Learning Systems are crucial for long-term success. The IP regime has traditionally been deemed an effective control mechanism and mean to appropriate value, however, IP legislation and administration have failed to keep pace with the rapid advances in technology. Recent progress in ICT in general, and Machine Learning in particular, have evoked debates on the effectiveness of IPRs in this technology field. However, the problem has mainly been addressed from an ethical, judicial and administrative perspective. Little attention have been given to outline:

how different control mechanisms, including IPRs, can be used to establish sustained competitive advantage for Machine Learning Systems.

1.2 Introduction to Machine Learning Systems

Machine Learning technologies are a sub-field of Artificial Intelligence (AI) and includes one of the most challenging factors of AI, namely to implement the capabilities of the learning in computers (Carbonell et al., 1984). A reasonable question to ask is why machines should learn at all, why not simply design the machine to perform as desired in the first place (Nilsson, 1998)? If a system can learn and adapt to changes, a system designer does not need to foresee and provide solutions for all possible situations (Alpaydin, 2014). Hence, if a task is difficult to define, characteristics are unknown, information is scarce or added constantly as the environment changes over time, the concept of implementing learning in the system becomes inevitable (Nilsson, 1998). Arthur Samuel, a pioneer in the field of AI, defined Machine Learning in 1959 as the *"field of study that gives computers the ability to learn without being explicitly programed"* (Simon, 2013, page 115). The ultimate goal was to someday build machines as intelligent as humans, where Machine Learning is the mean to train computers to do things that are impossible to program in advance (Kosner, 2013). Today Machine Learning algorithms are already implemented in a number of applications such as self-driving cars, writing and publishing sports match reports or to identify terrorist suspects (Marr, 2016a).

Machine Learning was created conceptually in the 1950s (Marr, 2016a) when Alan Turning created the *Turning Test*. This test was developed to determine if a computer has real intelligence. The test is successful if a computer has a conversation with a human and that human believes that he/she is talking to another human (Marr, 2016a; Ball, 2015). To achieve this became an important milestone for Machine Learning research as it was a measurement that the computer could actually learn by communication. The test was further developed to be applied on other tasks such as playing games. Aurthur Samuels studied the game of checkers and showed in 1952 that a program would be able to play better in the course of time, thus it would learn to improve its skill and eventually beat a human opponent(McCarthy and Feigenbaum, 1990; Ball, 2015). Arthur Samuels's prediction came true many years later in 1997, when Deep Blue won over the World Chess Champion Garry Kasparow (Ball, 2015).

The development of Machine Learning has been explored with different emphasis in different periods of time (Beyer, 2015). In 1957, Frank Rosenblatt designed the first neural network for computers, the Perceptron, that simulated the thought process of the human brain (Marr, 2016a). The method uses a large number of interconnects (neurons) and solves in parallel a specific problem which is learned from an example (Dimitrios, 2016). The research on neural networks was not proceeded in the 1960s as difficulties with solving boolean functions occurred (Marr, 2016b). However in the 1980s it was further pursued and today it is widely applied for example in pattern recognition (Dimitrios, 2016). An other method for pattern recognition became recognized, the "nearest neighbour" algorithm (Marr, 2016a). It was developed in 1967 and is conceptually simple and and could be used for example to map a route for traveling, starting at a random city but ensuring visit to several cities during a tour (Marr, 2016a).

Additional important inventions within Machine Learning is the *Stanford Chart* developed in 1979, that can navigate obstacles in a room on its own (Marr, 2016a). In 1981 Gerald Dejong introduced the concept of Explanation Based Learning (EBL)[1] and in 1985 Terry Sejnowisk invented NetTalk, a program which learnt to pronounce words the same way babies do (Marr, 2016a). In 1990 the direction of Machine Learning development shifted from using a knowledge-driven approach to a data-driven approach (Marr, 2016a). A data driven approach requires large amounts of data that are used to draw conclusions and to learn from. The basic belief is that behind the seemingly complex and voluminous data, there lies a simple explanation where patterns can be detected and then used to uncover patterns to predict future data or perform other kinds of decision making under uncertainty (Murphy, 2012; Alpaydin, 2014). Thus, Machine Learning uses the theory of statistics and probability when building mathematical models detecting patterns and processes, to make interference from a sample which then can be used to drive business decisions (Alpaydin, 2014; Chaudhuri, 2012).

During the 21st century, rapid advancements of key capabilities in other technology fields have accelerated the introduction of Machine Learning in new application fields. This includes the capabilities of making large amounts of data available, reducing the cost of storing and receiving data, increase in computing power and memory, and development of new methods for performing Machine Learning (Horvitz and Mitchell, 2010;

[1] computer analyses training data and creates a general rule it can follow by discarding unimportant data (Marr, 2016a)

Hemsoth, 2015). Machine Learning can be divided into three different classifications. The classification depends on the nature of the "signal" or "feedback" available and the way the system learns (Kotsiantis, 2007). In turn, learning approach determines what methods are applicable to solve the problem (Stuart and Norvig, 2013). The three categories are:

- *Supervised learning*: Learning a general rule that maps inputs to outputs by using example inputs and desired outputs provided by a "teacher".
- *Unsupervised learning*: The learning algorithm finds structures in the input itself without given labels.
- *Reinforcement learning*: A computer program interacts with a dynamic environment in which it must perform a certain goal without being told what action to take. It must discover which actions yield the best reward.

Deep Learning is a new category of Machine Learning, which was introduced in 2006. Deep Learning consists of two key aspects; (1) the models consist of multiple layers of stages of nonlinear information processing and (2) methods for supervised or unsupervised learning use feature representation at successively higher more abstract layers (Li Deng and Dong Yu, 2014). Gefforey Hinton was initially developing this new concept when he developed algorithms that let computers *see* and distinguish objects and text in images in videos (Marr, 2016b). Other advances in the development of Machine Learning technologies have been made for example by companies like Microsoft, IBM, Google and Facebook. The Microsoft Kinect, was introduced in 2010, and allowed people to interact with a computer via movements and gestures (Marr, 2016b). Watson a question answering computer system developed by IBM, could in 2011 beat human competitors at Jeopardy and Google's algorithms could in 2016 beat a professional player at the Chinise game Go, which is considered the worlds most complex board game (Marr, 2016b).

It is today possible to use Machine learning to make computers see, understand and interact with the world around them (Marr, 2016b). The development of new features is growing at a remarkable rate as the quantities of data increases (Marr, 2016b; Hamilton, 2015). It is clear that Machine Learning is implemented in new products and services every day impacting and transforming most industries to gain advantage over competitors (Marr, 2016b; Hamilton, 2015; SAS, 2016). Machine Learning is applied in several sectors such as financial services, government, health care marketing and sales, transportation and in the oil and gas industry (SAS, 2016). Examples of applications are to use Machine Learning to prevent fraud, identify investment opportunities, increase efficiency in governmental organisation with large sets of data, improve diagnoses and treatment of patients or to analyze transportation patterns to improve routes (SAS, 2016).

1.3 Purpose of the Study

Considering the increased application of Machine Learning across industries, simply using Machine Learning is not a source of sustained competitive advantage. Both practical and theoretical problems have been identified related to how to control Machine Learning technologies in order to create long-term profits from Machine Learning System (MLS). However, since the market in a phase of emergence, all pieces of the puzzle are still to be determined. Hence the purpose is:

To provide a starting point for firms to design sustained competitive advantages for MLS.

In addition, the authors of the report also intended to provide interesting and accessible information to the public.

1.4 Research Questions

The research questions are constructed with the aim to fulfil to the purpose of the study. They are also considered guidance when conducting the research (Bryman and Bell, 2015). A main research question is formulated based on the purpose and the research problem. In order to answer the main research question, three sub-questions have been constructed.

1.4.1 Main Question

Despite that the Machine Learning market is still shaping, waiting until the disruption has settled and imitate the winners may be a risky strategy. By understanding likely sources of sustained competitive advantage today, firms are in a good position to mitigate the risks of being leapfrogged in the future. Thus the main question is:

MRQ: What are the most likely sources of sustained competitive advantage for Machine Learning Systems?

1.4.1.1 Sub Questions

A reasonable assumption of the study is that technology plays a significant role in the competitiveness of MLS. Hence, an understanding of the fundamental components of a MLS and the main areas in which MLS may be differentiating is a first step to answer the main research question. The first sub-research question is therefore:

SRQ1: What characterize Machine Learning Systems from a technology perspective?

A pre-requisite for sustained competitive advantage is that there first exists a competitive advantage. A relative comparison between specific MLS is outside the scope of this study as the main question specifically targets sources that are most likely to result in long-term success. Hence the second sub-research question is:

SRQ2: What are potential sources of competitive advantage for Machine Learning Systems?

In order to design sustained competitive advantage, firms must not only understand what makes MLS competitive but also how they can effectively control those aspects. It then follows that the third sub-research question is:

SRQ3: What mechanisms are effective for controlling Machine Learning Systems?

1.5 Scope and Limitations

The concept of success is in this study restricted to the attainment of profit, thus only comprising financial performance. Additionally, the research is limited to existing Machine Learning technologies and does not take into account possible developments of the technology in the future. Furthermore, the study is limited to U.S. legislation and norm of business conduct. This study will approach the subject from a business perspective only, not including ethical, legal or other perspectives unless there is a clear connection to financial performance. It should be emphasized that although the study is closely anchored in technology, any technical immersion is solely for the purpose of understanding the context and the reader should not expect a deep dive in current technical advancements in Machine Learning. Two case studies of Netflix's and Amazon's business models have been investigated. The purpose of this study is not to investigate specific businesses per se, hence, the business models have not been analyzed in its entirety but instead been for identifying strategically important resources for MLS.

1.6 The Disposition of the Study

The disposition of the study is set up to include seven chapters and the bibliography and appendices with the following content;

The first chapter includes the *Introduction* to the study. This entails the background, introduction to MLS, the purpose of the study, the research questions, the scope, the delimitation and a reading guide.

The second chapter, entails the *Theoretical Framework* for the study. In the theoretical framework, key concepts and theories are described culminating into a construction of a framework for the research.

The third chapter includes the *Method* for the study. This chapter presents the research strategy, research design and research methods applied and discusses the parameters considered in regard to the quality of the research conducted.

The fourth chapter includes the *Empirical Results* of the study. In this chapter, the empirical research conducted is presented including data gathered from interviews, literature review and case studies.

The fifth chapter contains the *Analysis* of the study. In this chapter the findings from the empirical result are examined, compared and scrutinized in detail.

The sixth chapter presents the key findings and includes the *Conclusion* of the study. Additionally, suggestions for future research are outlined.

Finally, the last chapter presents the *Discussion* of the study. The main takeaway is presented along with suggestions for future research.

1.7 Reading Guide

This study is directed at business professionals in the ICT industry in general and Machine Learning industry in particular. For business practitioners, chapter four, five and six are recommended.

The study is also directed at academia, including students, to inspire and contribute to future research in the intersection between high-technology, business and intellectual capital management. Readers from the academia is advised to start with chapter one and continue with relevant chapters of interest.

Chapter 2

Theoretical Framework

This chapter presents the key concepts and theories that construct the theoretical framework for this study.

2.1 Key Concepts & Theories

Concepts are the building blocks of theory and represent the structure for the empirical results and analysis (Bryman and Bell, 2015). For this study, two key concepts have been identified; sustained competitive advantage (SCA) and control mechanism. Four different theories have been used to construct the theoretical framework. This section outlines the characteristics of each concept and theory, including underlying assumptions and applicability for the study.

2.1.1 Sustainable Competitive Advantage

The concept of sustained competitive advantage is an important piece of the theoretical framework used in this study. In general, competitive advantage is a business concept that has been widely discussed by prior researchers (Huang et al., 2015). There are two main perspectives that address this concept, the inward looking Resource-Based view (RBV) and the externally oriented Industrial organization theory (IO) (Peng et al., 2008; Huang et al., 2015).

2.1.1.1 Two Perspectives on Sustained Competitive Advanatge

The RBV and the IO concurs in that competitive advantage is the competitive position a firm establishes as a result of superior profit compared to its competitors (Huang et al., 2015; Grant, 1996). The fundamental difference between the approaches is *how* superior profit is created, namely the sources of sustained competitive advantage. The RBV perceives that superior returns and the competitive position of a firm stems from its idiosyncratic resources (Grant, 1996; Barney, 1991) . In contrast, the IO asserts that superior profit is attained by a firm's stronger market position in an industry compared to its competitors (Huang et al., 2015). In the latter, research tends to focus on describing external conditions that drive firm performance and Michael Porter's "five forces" is a leading theory originating from such thinking (Barney, 1991). There are two underlying assumptions that differentiate the perspectives. Firstly, the IO view assumes homogeneity of firms, meaning that all firms control the same strategically important resources and pursue the same strategies (Barney, 1991). In contrast, the RBV assumes that firms in an industry may be heterogeneous with respect to the strategic resources they control (Barney, 1991). Secondly, the IO perceives resources and strategies as mobile across firms in an industry. The perfect mobility of resources means that the existence of resource heterogeneity, caused by e.g. a new entry, can only be temporary and therefore not a source of sustained competitive advantage (Barney, 1991). However, the RBV assumes that resources may not be perfectly mobile across firms an may in fact result in a long lasting heterogeneity (Barney, 1991).

In recent years, efforts have been made to reconcile perspectives in order to create a holistic framework where sustainable competitive advantage of a firm is a function of favorable market conditions as well as efficient deployment of unique resources (Hooley et al., 1998; Huang et al., 2015). This study has employed the RBV as a theoretical starting point for analyzing sustained competitive advantage. IO theories are not included because the MLS industry is an emerging industry characterized by constant new market entries and heterogeneity of firms. IO theories, such as the Porter's "five forces" are due to its fundamental assumption of resource homogeneity less appropriate. Furthermore, IO theories tend to focus on the current state of the industry and not taking into consideration future competitors (Barney, 1991). These theories are often more suitable for analyzing mature markets where the rate of new entrants is low.

2.1.1.2 The Resource-Based View (RBV) & the Knowledge-Based View

In this section, the RBV will be further described. The section will also outline the fundamental concepts of the Knowledge-Based View (KBV), which may be perceived as an extension of the RBV (Grant, 1996). Together the RBV and KBV will form the theoretical framework for the concept of sustainable competitive advantage.

As previously accounted for, the RBV suggests that the sustainable competitive advantage of a firm stems from the idiosyncratic resources that it controls (Grant, 1996; Barney, 1991). Numerous RBV theorists have attempted to define firm resources. (Barney, 1991, page 101) defined resources as "all assets, capabilities, organizational processes, firm attributes, information, knowledge etc. controlled by a firm that enable the firm to conceive of and implement strategies that improves its efficiency and effectiveness". Wernerfelt (1984) saw resources as anything that can be viewed as strength or a weakness of a particular firm. In literature, a common distinction is the notion that resources may be divided into two categories; assets and capabilities. In this study, an asset has been defined as anything intangible or tangible that a firm can use in its processes for creating, producing or offerings its goods or services to a market (Wade and Hulland, 2004). In contrasts, capabilities are defined as repeatable actions that that transform inputs into outputs of greater value (Wade and Hulland, 2004). Examples of assets are networking hardware, GPUs, patents and software. Capabilities are often manifested in skills such as technical ability, creativity, or ability to collaborate. In this study, the concept of resources is therefore the collective assets and capabilities of a given firm.

Another important distinction of the RBV is that not all resources have the potential of sustained competitive advantages (Barney, 1991). There are four main attributes that must hold true in order for a resource to qualify: a) valuable, in the sense that it exploit opportunities or offsets threats in the external environment, b) rare, with respect to the firm's current and future competition c) imperfectly imitable, and d) non-substitutable, meaning that there cannot be equivalent substitutes that are valuable but neither rare nor imperfectly imitable (Barney, 1991).

If the resource is valuable and rare, it may be a source of competitive advantage (Barney, 1991). However, in order for a firm to obtain sustained competitive advantage, its valuable and rare (attribute a and b) resources must be difficult for competitors to imitate or to find substitutes for (c and d) (Barney, 1991). The inability of competitors to duplicate the benefits of a value creating strategy is what distinguishes sustained competitive advantage from (temporary) competitive advantage (Barney, 1991). Hence, sustained competitive advantage is not dependent on a specific period of calendar time during which a firm enjoys competitive advantage, but rather the possibility of competitive duplication (Barney, 1991).

In this study, MLS is considered a product of the knowledge economy, where the primary firm resources are intangible or intellectual by nature. Following that reasoning and the RBV, knowledge is the most strategically important resource of a firm to obtain sustained competitive advantage. This notion of knowledge as the most strategically important resource of a firm is the core of the Knowledge-Based View (KBV) (Grant, 1996). The KBV suggests that due to its specific characteristics, knowledge can only be owned by people (rather than organizations) and most knowledge can only be exercised by the people that possess it (Grant, 1996). This implicates a) the importance of a firm's human resources for sustained competitive advantage b) the role of the firm as a knowledge integrator and coordinator (Grant, 1996).

2.1.2 Business Model Canvas

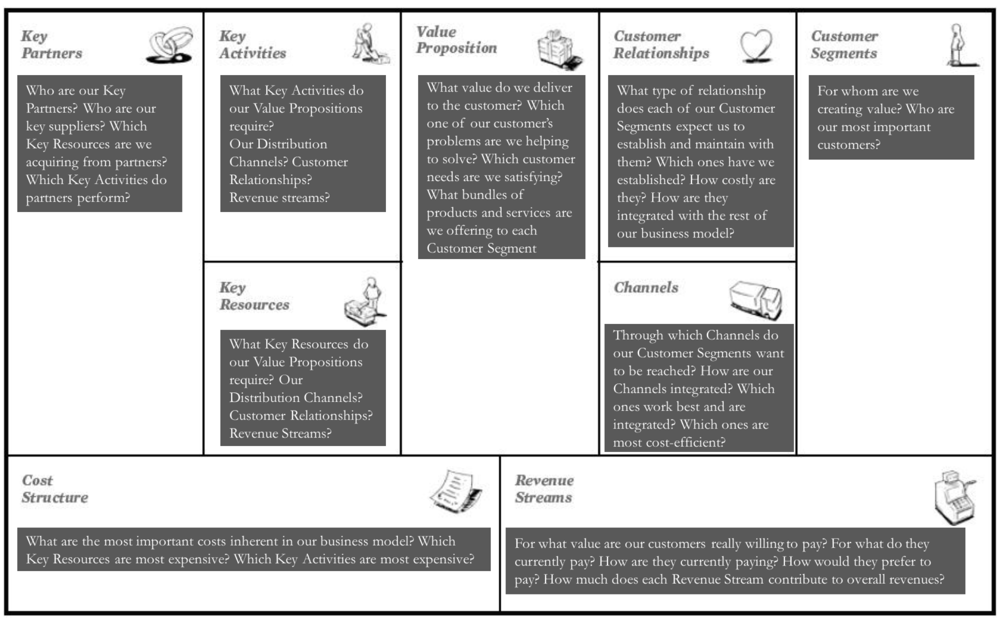
A business model describes the design or structure of a firm's value creation, delivery and capture mechanism (Teece, 2010).

The Busines Model Canvas (BMC) is a concept to describe and think through the business model of any organization (Osterwalder and Yves, 2010). Nine different building blocks are used for describing a business and how it intends to make money (Osterwalder and Yves, 2010). Each building block answers a set of questions, illustrated in Figure

2.1. The purpose of the Business Model Canvas is to create a converging understanding and an easy overview of your business model, allowing for creation of new strategic alternatives and successful innovation (Osterwalder and Yves, 2010).

The BMC is praised for its simplicity, however critics point out the lack of analyzing competitors and excluding strategic decisions such as objectives, missions and visions (Prof. Hong and Clemens, 2013; Kraaijenbrink, 2012). Furthermore, the model mixes the level of abstraction making the canvas unbalanced (Kraaijenbrink, 2012). However, since this study does not address competitive positioning nor specific business models, the critics of the model were not considered to compromise the usefulness for this study. Instead, BMC has been used for understanding the key resources and how they are linked to creating value for the customers (i.e. value proposition). Although not all building blocks are analyzed in detail, understanding all parts of the business model was considered necessary, as key resources and value proposition cannot be fully understood in isolation.

Figure 2.1: Business Model Canvas from Osterwalder and Yves (2010)



2.1.3 Intellectual Asset Mapping

Intellectual Asset Mapping (IAM) is a framework that provides processes and supporting tools for evaluating a organization's knowledge utilization, covering inward-oriented processes as well as external processes (Petrusson, 2015). The model was originally constructed for the academia, and specifically focusing on supporting the utilization of research result from a societal perspective (Petrusson, 2015). Since IAM is constructed to evaluate the performance of research organizations, the processes and tools are constructed for organizations which resource base and output mainly consist of knowledge. Hence the general premises of the model makes it well suited for the technology researched in this study. However, the model is developed for evaluating the utilization of knowledge from a societal perspective, rather than a commercial perspective which render some of the processes and tools obsolete in this study. Furthermore, since the purpose of this study is to investigate sources to long-term *commercial* success, the utilization of knowledge will be defined to *commercialization* of knowledge.

The IAM framework consists of four key processes: 1) Capturing 2) Positioning 3) Deciding 4) Managing (Petrusson, 2015). The capturing process (1) is characterized by supporting tools for identifying, analyzing and capturing intellectual assets. This is the most developed process of the four, and also the process that is most useful for this study. The positioning process (2) is characterized by supporting tools for positioning of research environment and intellectual asset portfolios. The deciding process (3) consists of supporting tools for considerations and decision-making in intellectual asset utilization. Finally the last step (4) outlines the managing of the intellectual assets identified.

In this study, the IAM framework was used for determining the characteristics of MLS as well as for identifying strategically important resources. Hence, only the first process of capturing intellectual assets was applied. The capturing process described by Petrusson (2015) focus on tools to capture knowledge assets, i.e. technical assets, which correlates well with the topic of this study. According to Petrusson (2015), knowledge assets can be further divided into eight different categories, which are described in Table 2.1.

Table 2.1: Knowledge Asset Categories

Knowledge Asset categories	**Examples**
Data	Measurement or test data, results, experiments, notes or journals
Database	For example Excel or Access files, matrices
Data correlation	Optimizations, trends, ranges, dependencies, connections
Theoretical framework	Models, theories, understandings, realizations, abstractions, schemes
Technical solution	Methods and processes, devices, units, compositions, designs, configurations, systems, technologies, inventions and solutions
Visualization and simulation	Designs, drawings, sketches, prototypes, diagrams, graphs, photos, simulations, models and demonstrations
Instruction	Algorithms, routines, procedures, guidelines, manuals, recipes and recommendations
Software	Systems, suites, platforms, programs, applications, drivers, plug-ins engines, clients/servers, GUIs, libraries, algorithms and script

2.1.4 Control Mechanisms

Following the Resource-Based View, the *sustainability* of competitive advantage depends on the possibility of competitive duplication (Barney, 1991). This means that if a firm takes measures to reduce the possibility of competitors to duplicate their strategy, they are in a better position to obtain sustained competitive advantage. These measures have been referred to in literature as *building blocks for structural control* (Petrusson, 2004) or *isolating mechanism* (Rumelt, 1997). For the purpose of this study, a combination of Petrusson's (Petrusson, 2004, page 136) structural control framework and Rumelt's (Rumelt, 1997, page 141) framework of isolating mechanisms have been used as theoretical starting point for control mechanisms. However, neither of the frameworks provides detailed definitions of the mechanisms, but rather examples of what they may contain. Therefore, the control mechanisms in this study have been defined by generalizing and interpreting examples provided in the two frameworks. In addition to the control mechanisms described in by Petrusson (2004) and Rumelt (1997), a firm's business model was identified in interviews as a sixth category of control.

2.1.4.1 Right Based Property

According to Petrusson and Heiden (2008), assets can only be considered as a property if it is trusted as an object of a commercial transaction. In order for a commercial transaction to take place, there must exists a system of well-established property rights, which can be validated in court and which is accepted by market actors and the society in general (Petrusson and Heiden, 2008). According to Alchian (2016), a property right is *"the exclusive authority to determine how a resource is used, whether that resource is owned by government or by individuals"*. A private property right, which is of most interest in this study, has three basic elements (Alchian, 2016):

1. exclusivity of rights to choose the use of a resource
2. exclusivity of rights to the service of a resource
3. rights to exchange the resource at mutually agreeable terms

Right based property can hence be viewed as an asset or resource that market actors and/or society is willing to financially compensate the owner in exchange for usage.

Although the underlying resource may be tangible, such as computer hardware, the right to use the computer hardware can only be viewed as intangible. Petrusson (2004) exemplifies Right Based Property with Intellectual Property Rights (IPRs), such as patents, copyrights, design rights, trade secrets and trademark rights. Also Rumelt (1997) recognizes *patents and trademarks* as isolation mechanism. For the purpose of this study, Right Based Property will be defined as Intellectual Property Rights. Understanding the scope and differences between different IPRs is essential for evaluating how at can be applied in different contexts. For this reason, IPRs will be treated as a separate key concept in Section *Intellectual Property Rights*.

2.1.4.2 Technical Control

Petrusson (2004) exemplifies technical control with a virtual product that cannot be copied because of a safety system. Also, Rumelt (1997)'s *specialized assets* can be mapped under this category. Technical control may not only be manifested by technical barriers for competitors to access resources, but also as a mechanism for retaining customers. Through standards, firms can make it technically difficult for its user to adopt competing products or services and thus creating a lock-in effect (Cusumano, 2010). Technical lock-in is only effective if the perceived value of the product or service exceeds the perceived value of interoperability. Technical protection services are another technical protection mechanism commonly applied by firms with products and services that are dominantly virtual (Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000). Technical protection services are software that is specifically designed for protecting a certain asset or IP (Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000). There are also technical protection for hardware, particularly for special purpose devices (Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000).

Physical security systems is another layer of technical control. Software and hardware can form very strong mechanisms for protecting IP in digital form. However, no technical protection can guarantee perfect control (Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000) as they, like any other invention, are subject to design and implementation errors which may be exploited by intruders (Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000). According to the Committee on Intellectual Property Rights in the Emerging Information Infrastructure (2000), the right technology ingredients must be weaved together into an end-end technical protection system in order to achieve effective protection. For the purpose of this study, technical control is defined as phenomena that technically hinder external entities from imitating benefits from a firm, or phenomena that technically hinders users from switching to another product or service.

2.1.4.3 Secrecy

Keeping the resource secret is another way of controlling a resource (Petrusson, 2004). This may be comparable to Rumelt (1997) *special information*, which no competitors posses. Secrecy does not only refers to what intellectual assets to keep secret, but also what information that should be disclosed. Secrecy is often employed for phenomena that cannot be protected by copyrights, patents or trademarks or if the control generated from the IPR is weak[1].

2.1.4.4 Contractual Control

Contracts generate rights of control that are enforceable in a court of law, provided that the agreement was made under fair conditions (Moore, 2009). Contracts opens up for flexibility as the terms of the contracts are completely up to the contracting parties (Moore, 2009). However, the flexibility of contracts also has a down-side. This is because contracts govern activities or transactions that will occur in the future, however, the agreement is designed based on the knowledge the contracting party has today (Tirole, 2009). Hence, the contract may be incomplete or in worst case even wasteful (Moore, 2009). According to Petrusson (2004), an example of contractual control is the transfer of property between to parties. Contracts may legally restrict a party from entering a market or using a resource and are therefore comparable with Rumelt (1997)'s *legal restrictions on entry*.

Contractual arrangements are only effective if all contracting parties understand and respect the agreement[1]. Although a party may receive remedies for a breach of contract, the damage may be more severe that immediate financial loss. For example if a trade secret is revealed, then that information can never be controlled again (Hallenborg et al., 2008). The enforceability and compliance of contracts vary by jurisdiction[1], and fostering relationships may be more effective in terms of control rather then the legal rights of contracts [1,2]. Informal agreement stemming from relationships are self-enforcing as long as the contracting parties find that the value generated from the relationship exceeds the consequences of non-compliance (Halac, 2012). Hence, strong relationships strengthen the control generated by contracts, and vice versa (Halac, 2012).

In this study, Contractual Control has been defined to control imposed through contracts between a firm and its external environment.

[1] Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05
[2] Sam Funnell, Stratified Medical, Interviewed 2016-04-19

2.1.4.5 Market Power

According to OECD (2002), Market Power refers to *"the ability of a firm (or a group of firms) to raise and maintain price level above the level that would prevail under competition"*. So given the same offering, what allows a firm to charge and maintain a higher price? Rumelt (1997)'s *reputation and image* fall into this category.

Successful firms have greater chances of sustaining superior performance over time if they possess good reputation (Roberts and Dowling, 2003). Reputation effects can be linked to brand loyalty, because if customers perceive a brand as reputable, they express a higher level of brand identification and loyalty (Kuenzel and Halliday, 2010). Brand loyalty may create a strong control position as it generates entry barriers to competitors, better ability to respond to competitive threats and it makes customer less sensitive to marketing efforts of competitors (Kurt et al.). Generally, trust is considered the core value a strong brand provides to customers, and a driving mechanism that lock customers to specific products, services or even firms (Kurt et al.). However, there are no legal rights generated from reputation effects or brand assets.

Another factor to consider when discussing market power is the concept of first mover advantage. Entering a market as pioneer is easier than breaking down a resource position barrier and replacing someone else (Rumelt, 1997). According to Lieberman and Montgomery (1988), there are three main mechanisms behind first mover advantage: (1) Technological leadership, (2) Preempting of scarce assets, and, (3) Buyer switching costs. Technological leadership may render control as a result of learning curve or experience, where costs falls with cumulative output (Lieberman and Montgomery, 1988). By pioneering a market, the firm may also achieve a control position from securing exclusive access to scarce assets. The third mechanism of first mover advantage refers to barriers arising from the additional costs new entrants may have to allocate to convince customers of the pioneer to switch from the product or service of the pioneer (Lieberman and Montgomery, 1988). This can be related to Rumelt (1997)'s *switching and search costs.*

2.1.4.6 Business model

In literature, business models have been described from a resource perspective as well as a mean for controlling intellectual property (Teece, 2010; Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000, see). For the purpose of this study, business models haven been approached as a control mechanisms. In Matteo's[3] opinion, a business model dictates the internal structures of the business as well as the externally-oriented structures. The model must be dynamic, meaning it needs to be designed to provide viable options in case the business fails on one market[3]. If designed correctly, a business model may compensate or even exploit properties of the business that is otherwise problematic from a control perspective (Committee on

[3]Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05

Intellectual Property Rights in the Emerging Information Infrastructure, 2000). The digitization has fostered a number of new business models as a result of the challenge of pricing information and the ease of customers to access the information without paying, i.e. copying (Teece, 2010). An example is free digital newspapers that monetize on the demographic data generated by its users instead of subscription revenues (Committee on Intellectual Property Rights in the Emerging Information Infrastructure, 2000). According to Teece (2010) the inherent complexity of the business model can create barriers to imitate. Also, commitment to customers, suppliers and key partners may hinder competitors from imitating another business model (Teece, 2010) Business models cannot be assessed without its context; sustainability and effectiveness of control can only be determined against the specific business environment in which the firm operates (Teece, 2010).

2.1.5 Intellectual Property Rights

According to World Intellectual Property Organization (WIPO), Intellectual Property Rights *"allow creators, or owners, of patents, trademarks or copyrighted works to benefit from their own work or investment in a creation."* (World Intellectual Property Organization (WIPO), 2013, page 3). There are two main reasons for countries having laws regulating the rights to intellectual property (Hallenborg et al., 2008). The first reason is to reward the creators with moral and economic rights to their creations and the right of the public to access their work. The second reason is to create incentives and promote the creation and dissemination of knowledge, and to encourage fair trade that contribute to both social and economic development. Intellectual Property (IP) is divided in five major categories: (1) patents, (2) copyrights, (3) trade secrets, (4) trademarks and (5) designs (Hallenborg et al., 2008). In this section, the legal premises, scope of rights and strengths and weaknesses will be covered for each of the five categories.

2.1.5.1 Patents

Patents are used for protecting an invention, meaning a product or a service that provides a new solution to a problem (World Intellectual Property Organization (WIPO), 2016). They are often considered the most powerful IPR as they protect both against deliberate and good-faith misuse and can serve as legal barriers for companies to capture returns on inventions (Hallenborg et al., 2008). Patents are used as means to make mergers, acquisitions and partnering proposals more attractive and generally enhances business reputation, increase invention prestige and protect shareholder value (Hallenborg et al., 2008). Patents give the owner of a patent a time-limited right to exclude others from using, producing, and commercialising an invention in a specific territory (Hallenborg et al., 2008). It is important to understand that patents do not give the patent holder a right to use its invention, simply the right to exclude others from copying, adapting, selling and making certain other uses of the protected invention (Hallenborg et al., 2008; World Intellectual Property Organization (WIPO), 2016). The trade-off between exclusivity and disclosure is also important to understand. If the the risk of not being able to exercise the exclusive rights exceeds the risk of making the invention publicly available, then the effectiveness of patents as mean of control is diluted [4]. In general, firms need to understand what market position they want to take and which patentable inventions that should be patented versus kept secret[4]. The territorial and time-limited properties of a patent means that patent filed in the U.S. is only valid in the US jurisdiction and in general 20 years from filing date (Hallenborg et al., 2008). The terms and conditions for a patent are governed by national laws which means that there are variations in terms and conditions (Lindmark, 2006). In the U.S., patents are issued by United States Patent and Trademark Office (USPTO), and there are three types of patents in the US patent law (Hallenborg et al., 2008):

1. **Utility Patent**, which may be granted *"to anyone who invents or discovers any new and useful process, machine, article of manufacture, compositions of matter, or any new useful improvement thereof"* (USPTO, 2014, "What is a patent?").
2. **Design Patent**, which may be granted *"to anyone who invents a new, original, and ornamental design of manufacture"* (USPTO, 2014, "What is a patent?").
3. **Plan Patents**, which may be granted *"to anyone who invents or discovers and asexually reproduces any new variety of plant"* (USPTO, 2014, "What is a patent?").

Of these types, utility patents are the most common (Hallenborg et al., 2008). In addition to fulfil the conditions of patentable subject matter, a utility patent is only issued for inventions that are absolute novel to the world (with one year grace period exception), useful and non-obvious for a person skilled in the area of technology related to the invention (Hallenborg et al., 2008; USPTO, 2014). Since 1998, business methods are considered patentable subject matter. The owner of a patent does not have to be human but is often a judicial entity, eg. a company (Davies, 2011).

In contrast to utility patents, design patents aim to protect *"ornamental design embodied or applied to an article of manufacture"* (USPTO, 2016, "What is a patent?"). Similarly to utility patents, it permits its owner to exclude others from making, using or selling the design(USPTO, 2016). In contrast to a utility patent, the term of the right extends to 15 years (USPTO, 2016). Another difference is that the design does not have to be useful or non-obvious, however it must fulfill the criteria of novelty and originality(USPTO, 2016).

Plant patents are not relevant for this study and will therefore not be covered in depth.

2.1.5.2 Copyrights

Copyrights protect "original work of authorship", including literary, dramatic, musical, artistic and certain other intellectual works, both published and unpublished (USPTO, 2014). Certain other intellectual work may be the source code of a computer program or a database as long as it is the result of human creativity and fulfills the criteria of

[4] Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05

"originality in expression" and "fixation" (Hallenborg et al., 2008; U.S. Copyright Office, 1997). The latter is by definition achieved since "fixation" means that the copyrighted work must be fixed on copies (Hallenborg et al., 2008). "Originality" refers to that the creator did not copy the work from somewhere else, and should not be confused with the stricter criteria of novelty for patents (Hallenborg et al., 2008). The scope of the protection does not include protection against identical or other similar works created independently (Hallenborg et al., 2008). Copyright can only be obtained for work that is the product of human authorship (Davies, 2011). This means that literary or artistic work produced by a mechanical process, or any non-biological artifact, without any contribution by a human author is not registrable (Davies, 2011). Similarly to patents, copyrights are negative rights giving the owner the exclusive right to exclude others from reproducing the copyrighted work, prepare derivative works, to distribute copies, to perform or to display the copyrighted work publicly (USPTO, 2014). Like patents, the exclusive right is time-restricted and territorial (Hallenborg et al., 2008). In US, copyrighted work is protected for the life of the creator plus 70 years (Hallenborg et al., 2008). Copyrights are afforded to the creator upon completion on work, as long as the work fulfil the criteria of originality and fixation (Hallenborg et al., 2008). This means that obtaining copyright protection does not require formal registration (Hallenborg et al., 2008). However in case of an infringement law suit, a registration in the US Copyright Office before the infringement will increase the remedies (Hallenborg et al., 2008). In relation to software, copyright has a very narrow protection as it only protects the exact formulation, and the same result can easily be achieved by minor alterations[5,6]. The requisite of creativity has resulted in various approaches to add creativity to enhance the copyright protection. Two main approaches are employed to increase originality of databases; the first is to add copyrightable texts such as annotations, abstracts etc. The second approach is to incorporate a more subjective selection of data or a more creative arrangement (U.S. Copyright Office, 1997). Despite these enhancements, copyright protection for databases is still very weak according to the U.S. Copyright Law (U.S. Copyright Office, 1997).

2.1.5.3 Trade Secrets

The purpose of trade secrets is to protect commercially valuable proprietary information (USPTO: Office of Policy and External Affairs, 2016). The Uniform Trade Secret Act defines trade secrets as *"information, including a formula, pattern, compilation, program, device, method, technique or process that both (i) derives independent economic value, actual or potential, from not being generally known to, an not being readily ascertainable by proper means by other persons who can obtain economic value from its disclosure or use; and (ii) is the subject of efforts that are reasonable under circumstances to maintain its secrecy"* (National Conference of Commissioners on Uniform State Laws, 1985, page 5). Hence, a trade secret can cover information that is subject to other Intellectual

[5] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29
[6] Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05

Property, for example patents and copyrights. The value of a trade secret stems from the secrecy itself, and disclosure means complete loss of protection (Hallenborg et al., 2008). This means that obtaining trade secret protection involves no registrations as such registration would require disclosure of the secret information (Hallenborg et al., 2008). However, to be subject to trade secret protection, the owner must use ”reasonable efforts” to keep the information secret (Hallenborg et al., 2008). In contrast to patents and copyrights, trade secrets provide protection against misappropriation, and not copying or reverse engineering (Hallenborg et al., 2008). Also, the term of the protection is unlimited (Hallenborg et al., 2008).

2.1.5.4 Trademarks

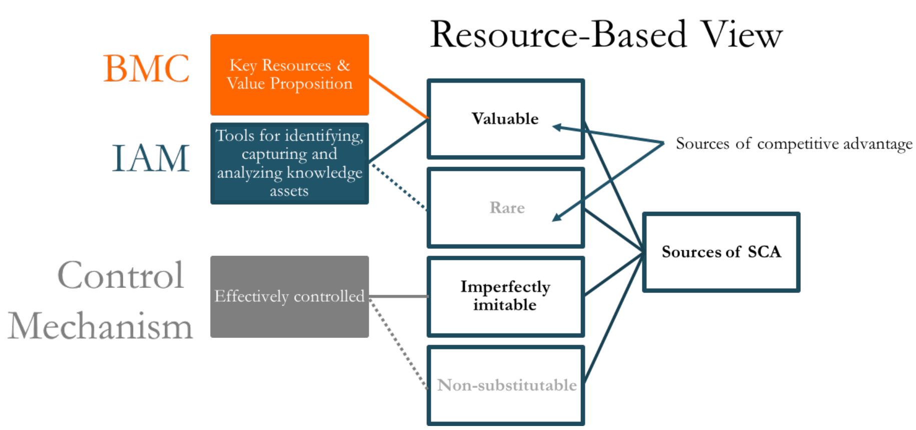
The term ”trademark” is often used in common language for two separate legal artifacts; trademarks and service marks (USPTO, 2014). A trademark is a word, name, symbol or device that is used in trade to indicate the source of goods and to distinguish the goods from competing products (USPTO, 2014). A service mark is equivalent to a trademark except it refers to a service rather then a product (USPTO, 2014). Trademarks are a subset of an overall brand concept making it closely connected to the brand strength (Greene and Wilkerson, 2013). Trademarks strongly influence the purchasing behaviour of consumers and have a vital impact on business (World Intellectual Property Organization (WIPO), 2009). There is no requirement on absolute novelty, however the trademark must be original in the sense that no other legal entity has prior rights to the same or confusingly similar mark for goods or services of the same type and in the same territory (Hallenborg et al., 2008). There are also other restrictions on trademarks, eg. it cannot be descriptive, generic, immoral, deceptive or simply a surname (Hallenborg et al., 2008). Trademarks can be obtained either by registration to USPTO or by establishment, although registration is usually considered to provide more legal benefits (Hallenborg et al., 2008). Trademarks give the owners the right to prevent others from using a confusingly similar mark for similar products or services in a certain territory (Hallenborg et al., 2008). The more established, i.e. recognized, a trademark is, the broader the scope of protection (Hallenborg et al., 2008). For example, a business or individual that is using the wording ”Coca Cola” in course of business is likely to be found guilty of infringement despite geographical location or type of products. Similarly to trade secrets and in contrast to patents and copyrights, trademark protection can last indefinitely, provided that it continues distinguishing the product or service (Hallenborg et al., 2008).

2.2 Constructed Framework

The Resource Based View (RBV) constitutes the theoretical foundation of this study and a framework for identifying and analyzing important resources of MLS. Additional theories and models have been used to support the identification of resources that meet the attributes dictated by RBV. Figure 2.2 illustrates how concepts from the Business Model Canvas (BMC), Intellectual Asset Mapping (IAM) and Control Mechanisms have been combined with RBV to construct the framework applied in this study.

BMC and IAM have mainly been used for identifying and analyzing valuable resources. Similarly, control mechanisms have mainly been used for identifying and analyzing imperfectly imitable. The attributes rare and non-substitutable have been approached from a hypothetical perspective and together the framework provide a basis for identifying and analyzing sources of sustained competitive advantage.

Figure 2.2: The theoretical framework constructed for this study



Chapter 3

Method

This chapter describes the methodology of the study including an outline of the research strategy, research design, research methods and an discussion on the quality of the research.

3.1 Research Strategy

The research strategy is based on the nature of the research and its link to theory (Bryman and Bell, 2015). The research strategy was ultimately anchored in the purpose of this study as well as how the research relates to theory.

3.1.1 The Relationship between Research and Theory

Relating to the purpose of the study, the aim was to generate new theoretical models rather than confirming existing theories. Due to the disruptive environment and the complexity of sustained competitive advantage, it was difficult to develop propositions from current theory that were testable in the real world (deductive reasoning) (Dubois and Gadde, 2002). The complexity of the researched field also imposed difficulties in generating the empirical data necessary for theory-building, which is the foundation of inductive reasoning (Bryman and Bell, 2015). Instead, the objective of this study was to select the 'best' explanation based on existing theories and empirical data (Bryman and Bell, 2015) collected from firms operating on the market as well as industry experts and leading researchers. Hence, the mode of reasoning employed in this study is dominantly abductive. According to Dubois and Gadde (2002), abductive reasoning is fruitful when the purpose is to find new things. Also, since Machine Learning is a field undergoing rapid development, it was beneficial not to be unnecessarily constrained by previously developed theories but instead allow for modification as discoveries were made (Dubois and Gadde, 2002).

3.1.2 Ontological and Epistemological Considerations

Ontological and epistemological considerations are anchored in the inference logic accounted for in *The Relationship between Research and Theory*. Furthermore, assumptions about the nature of knowledge (ontology) and how knowledge should be validated (epistemology) fed into the formulation of purpose and research questions of this study (Bryman and Bell, 2015).

From on ontological perspective, the key concepts researched in this study, *sustainable competitive advantage*, *control mechanisms* and *Machine Learning Systems* are phenomena which existences are the results of human creations. This type of ontological subjectivity implicated the adoption of a constructivist position, where social phenomena are produced through social interaction and in a constant state of revision (Bryman and Bell, 2015). Similarly, an interpretivist approach was adopted from an epistemological perspective as strategies from natural science would not be appropriate to answer the research questions. For example, in order identify sources of sustained competitive advantage, an understanding of the organization and the human actions that governed it was necessary rather than a scientific explanation. According to Bryman and Bell (2015) this aspiration to *understand* human behavior instead of *explaining* it is one of the key differences between interpretivism and positivism, which are the two most recognized position in epistemology. Following this reasoning of ontological and epistemological considerations, and taking into consideration that abduction stems from pragmatism, a pragmatic perspective was employed in this research. An pragmatic approach allowed freedom from mental and practical constraints imposed by strictly embracing one philosophical extremity, and simply focus on solving practical problems in ”the real world” (Feilzer, 2010).

3.1.3 Qualitative and Quantitative Research Considerations

Qualitative and quantitative research can be viewed as two distinctive clusters of research strategies, which foundations builds on the connection between research and theory as well as ontological and epistemological positions (Bryman and Bell, 2015). Following the positions taken in *The Relationship between Research and Theory* and *Ontological and Epistemological Considerations*, qualitative research methods were found to be most appropriate for this study.

3.2 Research Design

The study used a comparative research design where more or less identical methods were used across multiple cases (Bryman and Bell, 2015). The design entailed comparing observations from interviews with Machine Learning experts and business practitioners with the observations from two different businesses that are recognized as market leaders in MLS. In that sense, the design can be viewed as a multiple-case study approach.

The reason for adopting a comparative design was based on the logic that social phenomena can be understood better when they are compared to two or more meaningfully contrasting cases and situations (Bryman and Bell, 2015). Following this reasoning, similarities between different situations and cases were used as a starting point for analysis and conclusion. The design selection was influenced by the choice of conducting qualitative research as some designs are more appropriate for this type of research (Bryman and Bell, 2015). Comparative design following a multiple-case study approach is relatively common for business and management research (Bryman and Bell, 2015). A general objection towards multiple-case study is the risk of losing contextual sight (Bryman and Bell, 2015). This risk was mitigated by not only collecting observations from specific firms, but also collect observations from industry experts and practitioners.

3.3 Research Methods

Research methods are different techniques for collecting data (Bryman and Bell, 2015). This section outlines the data which was considered required for answering the research questions based on theoretical framework. The section also include outlines of the research process and how the data was collected.

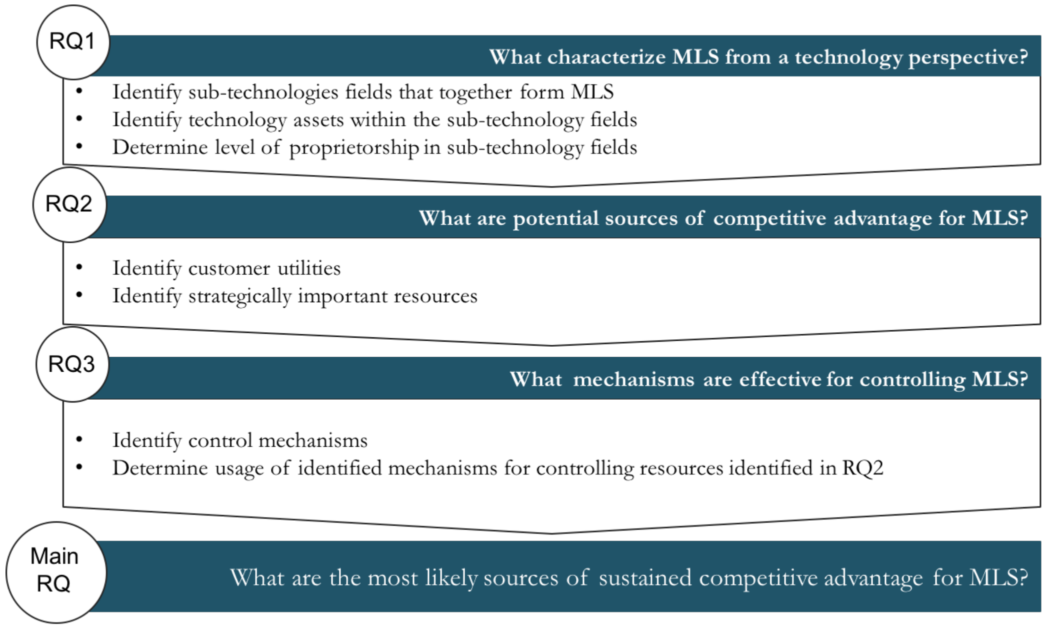
3.3.1 Required Data

In order to answer what characterize a MLS from a technology perspective, it was considered necessary to understand what different sub-technologies that constitute the backbone of MLS. Furthermore, following the theoretical framework, technical assets and their function were identified and described. In order to understand how one MLS may technologically differ from an competing MLS, the level of proprietorship was also investigated. Although this study never went as far as proving rarity, the level of proprietorship was used for reasoning around the uniqueness of different technical assets.

The second research question is centered around sources of competitive advantage. Following the Resource-based View, the most strategically important resources needed to be identified. A reasonable assumption was that there is a correlation between the strategic importance of a resource and the value that resource create for the customer. Hence, by identifying and linking customer utilities to the resources identified, the most likely sources of competitive advantage could be identified.

Relating to the third sub-research questions, a general understanding of control mechanisms was necessary as well as their applicability on MLS. Additionally, in order to identify which control mechanism are *effective*, data on the actual usage of control mechanisms in the MLS industry needed to be collected. The data required for all three sub-research question has been summarized in Figure 3.1.

Figure 3.1: Data required for this study



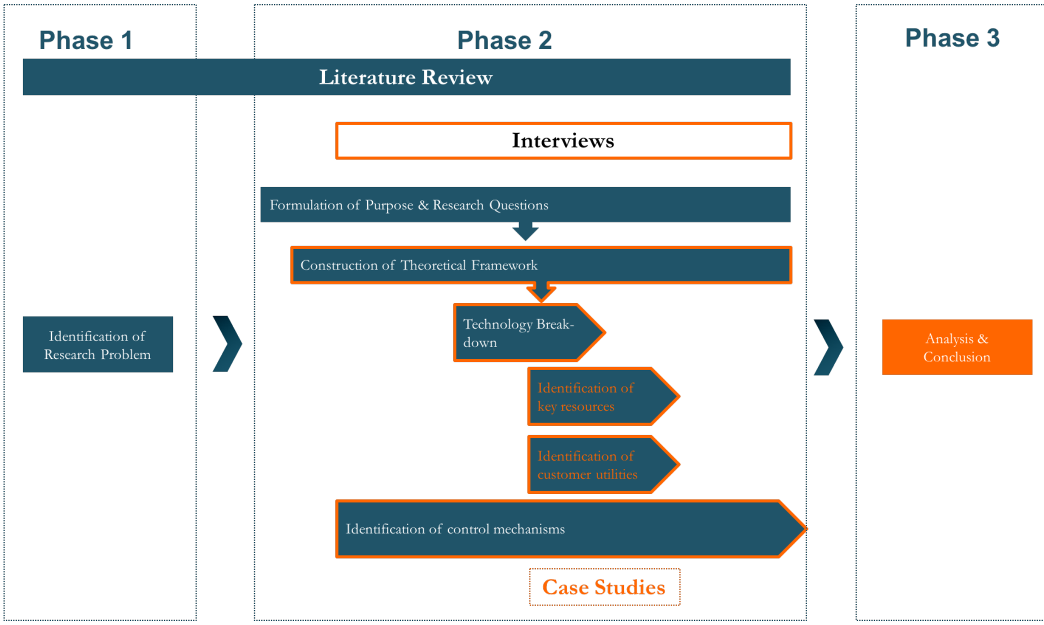
3.3.2 Research Process

The research process includes all the major steps conducted throughout the research. Figure 3.2 illustrates the sequence of the process used in this study, and the colouring of each step illustrates the data collection method used. The process consisted of three main phases which are denoted 1,2,3 respectively in Figure 3.2. In accordance with the selected research approach, several of the steps in Phase 2 stretch over the entire phase allowing for empirical findings to be incorporated.

In the initial phase, a practical problem was identified by a commercial actor in the researched industry and shared with the researchers. By reviewing prior research it was also validated that the problem had not yet been addressed in literature. Following an abductive approach, the practical problem served as initial hypothesis.

The second phase constituted the majority of the research process and was highly iterative in order to facilitate for new hypotheses and theories to be discovered and formulated. The second phase started with the formulation of purpose and research questions based on the research problem and the general interest of the researchers. Next, a theoretical framework was created from scanning relevant theories and models for the key concepts in the research questions. The theoretical framework governed the data collected as well as the analysis. Adjustments of the theoretical framework was made as new theories were revealed in interviews. For example, business model was added as a control

Figure 3.2: Research Process



mechanism after it was highlighted in interviews. Next, MLS was broken down into constitutive technologies, using a technology tree (see Petrusson, 2015, page 346). The process of constructing a technology tree provided general insights and understanding of the researched field (Petrusson, 2015). The technology break-down was a result of both interviews with MLS experts as well as literature. During this step, data on proprietorship was also collected. The third step in the second phase consisted of identifying the assets and capabilities that were linked to competitive advantage. This identification was done in several parallel steps. Examining two successful MLS companies' business models provided data on key resources and how these resources were linked to the value propositions. In parallel, interviews with MLS businesses, investors and industry experts provided additional observations on key resources and how MLS created customer value. Since general control mechanism were identified during the construction of theoretical framework, the next step consisted of interviewing business practitioners and IP experts on how MLS are controlled today.

In the final and third step of the research, the findings from all previous steps were interpreted and analyzed. Findings from the technology break-down were combined with findings on key resources and value creation to determine which resources that most likely would result in competitive advantage. These resources were then compared with findings on control mechanism to arrive at a final conclusion that satisfied the main research question and the overall purpose of the study.

3.3.3 Data collection

Thre main methods were used for data collection: (1) literature review, (2) interviews, and (3) case studies. The following section describes each method and considerations in regard to each method.

3.3.3.1 Literature Review

Literature review was used in several steps of the research process (see Figure 3.2). Goolge Scholar and the database of Chalmers University Library were mainly used when searching for relevant publications. Due to the time constraint only a limited amount of references could be collected, however an extensive search was conducted trying to find the most relevant literature. Publication year and the number of citations were considered when selecting literature to safeguard the quality of research. When searching for literature for theoretical framework and key concepts the amount of citations were favored over more recent publication year. In contrast, for literature related to Machine Learning and developments in control mechanisms, recently published data was considered more important to collect. For that reason blogs and forums were reviewed when the data required could not be found in scholar publications. This was particularly the case for data related to Machine Learning and case studies where publications were either considered outdated or did not exist.

3.3.3.2 Interviews

Interviews were one of the main methods to collect data for the empirical result and to gain an understanding of the field. The selection of the interviewees is most relevant to set the scene of the research. Although availability often results in convenient sampling (Bryman, 2016), effort was made to ensure a selection of different organizations to collect a broader perspective, thus, both mature organizations and start-ups where included in the sample. To only sample organizations or experts that have been or are active within the researched field was also deemed important. A theoretical sampling approach was applied where theoretical reflection on data was used as a guide to whether or not more data was needed (Bryman, 2016).

When conducting qualitative research, the information gathered during the interview should cover both a factual and a meningeal level (Teijlingen, 2014). Three main interview techniques can be used which are presented in Table 3.1.

Table 3.1: Interview Techniques

Structured	Predetermined questions with fixed wording.
Semi-Structured	Predetermined questions. Order, wording and questions can be modified.
Unstructured	No predetermined questions, general area of interest to guide the conversation.

In qualitative research, the main task when interviewing is to understand the meaning of what the interviewees say (Peterson, 1997). Thus, a semi-structured approach was chosen for the interviews as it allowed for modification and follow-up questions. An interview template (see *Appendix A: Interview Template*) was constructed to ensure a certain level of comparison of interview results. The design of the template reflects the data specified for answering each research sub-question. To improve efficiency, the questions were divided into three sections where each section reflected a part of the investigation. This structure enabled the interviewers to better adapt the interview to the interviewee(s), and only ask questions relevant for the interviewee(s). For example, an IP specialist was mainly asked questions related to control mechanisms, whereas a Machine Learning researcher mainly was asked questions related to the technology.

Interviews can be executed in three different ways: (1) face-to-face, (2) calling or (3) using the Internet. Face-to-face interviews have been found to generate greater amount of information(Opdenakker, 2006) and were therefore employed to the extent possible. However, due to geographical restrictions, phone interviews were also conducted. The interviews were not recorded as this might compromise the content of the answers. However, notes were taken by both of the interviewers during the interview.

3.3.3.3 Case Studies

The choice of case studies is based on interview findings where experts highlighted Amazon and Netflix as MLS market leaders and pioneers in their field. The assumption is that the valuable resources of these companies can be used to compare the resources identified as valuable in interviews and literature review. The case studies where designed to investigate a limited part of the companies, namely to gain in-depth knowledge of the defined part. Data was collected using secondary sources such as articles, books, blogs, technical forums and annual reports. The limitation to only investigate certain parts of the companies created an imminent risk of bias. Hence, no conclusions were inferred solemnly from case study findings.

3.4 Quality of Research

The quality of research can be established and assessed thorough different criteria. Bryman and Bell (2015) suggests *reliability*, *validity*, *trustworthiness* and *authenticity* as metrics for assessing quality of research. This section includes an assessment of these criteria in comparison to this study.

3.4.1 Reliability

The term reliability can be divided into two different parts, *external* and *internal* reliability. External reliability refers to the degree to which a study can be replicated, whereas the internal reliability concerns whether or not, one or more observers can agree upon what they see and hear (Bryman and Bell, 2015). An implication with conducting qualitative research is that a social setting with certain circumstances is impossible to replicate exactly. Nonetheless, in this study, steps have been taken to enable future studies with similar results. A template for the interviews ensures a certain amount of reliability in combination with with meticulous documentation of the research process. Furthermore, a theoretical framework was constructed to enable similar observations of the researched field, where observations can be obtained in a likewise manner. Regardless, bias of authors can not be exclusively excluded.

3.4.2 Validity

Similarly to reliability, validity can be divided into*external* and *internal* validity. The external validity refers to the degree to which findings can be generalized across social settings, whereas the internal validity refers to whether or not there is a good match between researchers observations and the theoretical ideas they develop (Bryman and Bell, 2015). In this study, a comparative research design was used and performed on a selected sample. Thus, the results presented give insights to the specific field and may be used as guidance for future research. However, it is not suited for conclusion of a larger group or field. A challenge when selecting the sample was the availability of objects making the sample restricted, which limits the validity. The internal validity tends to be the strength of qualitative research as the researchers participate over a long period of time which allows assurance of a high level of conformity between concepts and observations (Bryman and Bell, 2015). In this study, an iterative process was applied allowing for systematic and continuous validation of data.

3.4.3 Trustworthiness

Trustworthiness is made up of four criteria: (1) credibility, (2) transferability, (3) dependability, and 4) confirmability (Bryman and Bell, 2015).

Credibility relates to the degree of which the research is consistent with reality. In order to establish credibility of findings, the research must be carried out in accordance with good practice and be validated with social reality (Bryman and Bell, 2015). In this study, credibility is achieved by using a comparative design and triangulation, were several sources are compared to validate findings. The usage of blogs or forums, which may be driven by hidden agendas or goals, may compromise the credibility of this study. However, for some data, these were the most reliable sources that could be found. To compensate for the suspected bias, triangulation was used.

If the findings of the study can be applied in another context, high *transferability* is achieved. The transferability of this study is difficult to assess as the research is restricted to a defined field. However, in order to provide transferability, extensive descriptions of the field is provided, making it adequate for an individual judgment (Bryman and Bell, 2015).

Dependability is created if the findings can be repeated (Bryman and Bell, 2015) which is difficult for this study as several parts of the research is based on interviews that are impossible to replicate in the exact same manner. Nevertheless, by showing the process of the research and describe each step, the best possible conditions have been set.

Complete objectivity ensures *confirmability* (Bryman and Bell, 2015) which is very difficult to attain. To avoid bias, frameworks and methods were applied, for example when conducting interviews to circumvent leading the direction. Motivation and reasoning behind selections of frameworks as well as restrictions and challenges further strengthen the objectivity of the research.

3.4.4 Authenticity

The authenticity of the research measures if values and perspectives in the study are represented in a fair and balanced way, which is typically determined by five criteria: (1) fairness, (2) ontological authenticity, (3) educative authenticity, (4) catalytic authenticity and (5) tactical authenticity (Bryman and Bell, 2015). In this study, interviews were conducted with individuals who were experts in their field, thus the study reflects this viewpoint and should not be applied to all individuals in an organization. To enhance authenticity, different perspectives within the field were gathered.

Chapter 4

Empirical Results

This chapter presents the empirical results from the data gathering of the research and is structured in three different sections. The first section concludes the empirical results defining the *MLS Characteristics*. The second section presents the empirical data gathered for *Sources of Competitive Advantage for MLS*. The final section contains the empirical result found for *Control Mechanisms* in the researched field.

4.1 MLS Characteristics

This section will initially describe characteristics of MLS from an application perspective, and then drill down describing constitutive sub-technologies from a system perspective. The empirical result is based on literature and interviews with experts in an iterative process. Details about the interviews are found in *Appendix B: List of interviewed Persons*.

4.1.1 Machine Learning System from an Application Perspective

Machine Learning is commonly categorized by the learning structure of the model(see *Introduction to Machine Learning Systems*). Another way to categorize MLS is in relation to application field. Figure 4.1 illustrates the span of applications for MLS. Bob Price[1], Research Fellow at Parc with extensive experience in the field, states that depending on application, MLS range from Transactional Processing Systems to Embedded Systems. Transactional Processing Systems are characterized by real-time data processing and training of models. Examples of Transactional Processing systems are recommendation system, e.g. Netflix's content recommender. Embedded Systems do not rely on real-time data for training, but instead these models are pre-trained and are embedded within a system to perform a specific task. Image -and video recognition is a common application area for Embedded Systems. The system and performance requirements differ depending on application, however Embedded Systems tend to build on more advanced Machine Learning models, such as deep learning. Search engines, such as Google's, is an example

[1] Bob Price, Parc, Interviewed 2016-03-18

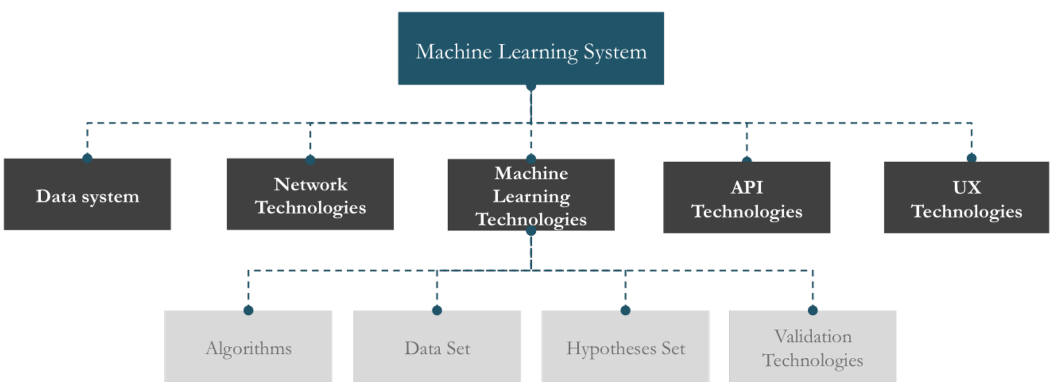
of a MLS type that is characterized by more advanced Machine Learning models compared to Transactional Processing System, but more simple than an Embedded System[2]. Most applications today only require transactional processing, however the amount of Embedded System applications are increasing[2].

Figure 4.1: Machine Learning Application Categorization

4.1.2 Machine Learning System Break-down

This section presents the structure of a Machine Learning System (MLS) for the purpose of understanding the characteristics of MLS as well as provide the reader with an overall understanding of the technology. Five sub-technology fields have been identified, which are illustrated in Figure 4.2; (1) Data System, (2) Network Technologies, (3) Machine Learning Technologies, (6) API Technologies and (5) User Experience (UX) Technologies. Each of these categories are extensive technology domains in themselves. However, considering the scope of this study, Machine Learning technologies are covered more in depth.

Figure 4.2: Machine Learning System



[2]Bob Price, Parc, Interviewed 2016-03-18

4.1.2.1 Data System

The Data System may include technologies for cleaning, integrating, storing and processing the data required to power the model. Data cleaning involves detecting and removing errors and inconsistencies from data in order to improve the quality of data (Rahm and Do, 2000). When the system relies on data from multiple data sources the need for data cleaning increases (Rahm and Do, 2000). Depending the characteristics of the data, such as amount of data and sources, technologies can be more or less manual[3]. There are several off-the shelf products and services that specialize in data cleaning[3]. When multiple data sources are used, there is also a need for transforming the data that it is represented uniformly in the storing system. The data needs to be integrated in a correct and systematic way to ensure ease of use and efficiency. This transformation process is managed by data integration technologies which employs software to extract, match and integrate schema (Rahm and Do, 2000). The schema is then implemented on some kind of storing system, such as database or database warehouse (Rahm and Do, 2000).

The data system also includes technologies for processing and perform analytics on the data stored in the database. Many companies employ extensive Database Management Systems (DBMS) to handle the storing, organizing and processing of information (Wodehouse, 2016). The aim is to create a system that is fast and energy efficient while maintaining low costs (Rubens, 2014; Goodwins, 2015). A DBMS consists of software which controls the storage, retrieval, deletion, security and integrity of data within a database (Wodehouse, 2016). In addition to software, the data system also includes hardware such as processors, memory and storage. Data storing technologies are undergoing dramatic evolution, where the overall efficiency and flexibility have become crucial (Rubens, 2014; Kieun, 2011). For data storing, the development is primarily focused on increasing capacity, performance and optimizing the physical size of the storage media (Pinola, 2015). Examples of storage technologies are helium filled drivers, Shingled Magnetic Recording (SMR) and Ethernet hard drives, and 60TB heat-assisted magnetic recording (HAMR) drives (Pinola, 2015; Rubens, 2014).

The choice of hardware depends on the application of the MLS[4]. For example, if the MLS is an Embedded System for image or video analytics, then GPU is the preferred processor technology as higher speed is required. However, if the MLS deals with transactional processing, such as a recommendation system, then CPU is the dominating technology for processor[4]. Although software is vital for the data system, the importance of hardware is not to be overlooked. For example with the development of GPU, the neural net training is 10-20 times faster than with CPUs (Jones, 2015). Historically, recommendation systems are most commonly employed to create application using Machine Leaning technologies and there are several open-source platforms and frameworks

[3] Marzieh Nabi, Parc, 2016-03-08

[4] Bob Price, Parc, Interviewed 2016-03-18

for dealing with large amount of transaction data[5]. These frameworks employ distributed computing in which a cluster of autonomous computers communicate with each other in order to achieve a goal (U.C. Berkeley, 2011). By distributing the data processing on multiple units the processing time is reduced and the reliability and scalability of the system increases[5]. Apache Hadoop is an example of an open-source software framework that is commonly used by MLS developers in which the user is presented with a virtual database system, but in reality the data is processed and stored in different locations on commodity hardware[5] (ApacheHadoop, 2016). From interviews with researchers and commercial MLS actors it has become clear that although there are examples of companies that construct and design their own data systems from scratch, most MLS use data systems that are built on top of existing solutions.

4.1.2.2 Network Technologies

Networking technologies enable communication and foremost access between two or more devices with the purpose to share data (Mitchell, 2016). For a MLS, this implies that the supplier computing resource can be shared with the customer computing resource[6]. To enable computer networks, a combination of computer hardware, cabling, network devices and computer software is put together (ComputerNetworkingNotes.com, 2016). On a high level, networks consist of hosts, network devices, links, protocols, applications and humans and agents (Park, 2016; ComputerNetworkingNotes.com, 2016). This high-level definition spans an entire system and can therefore include Data System, APIs, and UX technologies. For the purpose of this system break-down of a MLS, the definition of network technologies have been narrowed to specifically cover the technology domains that are not covered by the other sub-technologies, namely network devices, network links and network protocols. By combining the components in different ways the following characteristics are attempted to optimize: availability, costs, speed, scalability, topology, security and governance (Park, 2016; ComputerNetworkingNotes.com, 2016).

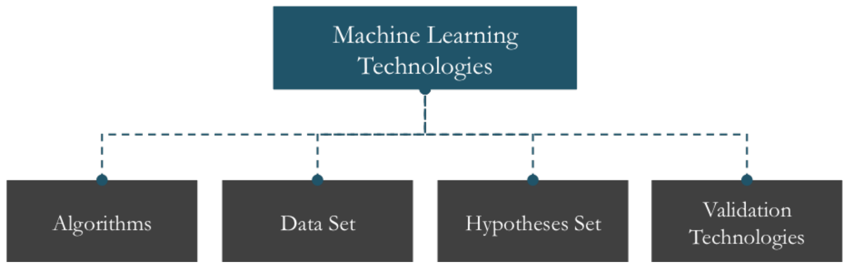
Together the network technologies create a network system. Commonly used network systems are cloud networking, enterprise networking or host server configurations (Microsoft, 2016). A MLS does not necessarily require networking technologies, the Data System, Learning Model, APIs and UX can all potentially be incorporated in one device[5]. However, MLS generally requires large storing and processing capacity and a single device, such as a laptop, seldom meet the requirements on scalability, reliability and processing time[5].

[5]Bob Price, Parc, Interviewed 2016-03-18
[6]Matthew Shreve, Parc, Interviewed 2016-03-09

4.1.2.3 Machine Learning Technologies

Figure 4.3 illustrates how Machine Learning technologies can divided into Algorithms, Data Set, Hypotheses Set and Validation Technologies.

Figure 4.3: Machine Learning Technologies



Danny Bobrow[7], Expert within AI with over 100 patents, states that the characteristics of Machine Learning technologies, also referred to as Machine Learning model, depend on the problem that to be solved. The model is often described as a blackbox that solves a problem, leading to different results (Ho, 2012). The result depends on multiple factors such as model selection, type and amount of data, the features prepared, but also the framing of problem as well as selection of objective measures used for estimating the accuracy (Brownlee, 2014). Examples of models are binary classification models, multiclass classification models or regression models (AmazonWebServices, 2016). To create a learning model, technologies for algorithms, data set and the hypotheses set are employed (Yaser, 2012). Furthermore, validation technologies are incorporated to measure the result of the model(Ho, 2012).

An algorithm is defined as a procedure or a formula with a set of rules to be followed for solving a problem (OxfordDictionaries, 2016). The algorithm will thus highly influence the performance and outcome of the model. Algorithms, which may also be referred to as classifiers, are the very core of the learning model[8]. Matthew Shreve, Machine Learning Expert at Parc, further explains that although some developers might develop their own algorithm, many algorithms can be found off the shelf. There are hundreds of different Machine Learning algorithms which can be applied to create a large amount of models solving different problems (Ambati, 2015). Algorithms can be clustered based on how they function, for example if they use tree-based methods or neural network methods (Brownlee, 2013). Alternatively they can be grouped depending on how the model learns, for example using regression or classification models (Brownlee, 2013).

The learning model does not only include algorithms, but also a hypotheses set. The hypotheses set is a collection of candidate formulas construed for a given data set (Yaser,

[7]Danny Bobrow, Parc, Interviewed 2016-03-09
[8]Matthew Shreve, Parc, Interviewed 2016-03-09

2012). The hypotheses set formulates the questions that solve the problem of the Machine Learning model[9]. Marzieh Nabi[10], Machine Learning Expert at Parc, considers the hypotheses set to be one of the most differentiating factors for a model, as it directly impacts the algorithm selection and overall design of the model.

There are typically three different types of data sets employed in a model; training data, test data and verification data[9], [10]. The training data set is used to build the model and determine its parameters and will tell if the model learns what it is suppose to learn (Keller, 2016). The challenges are to find the most relevant features for representing the data and to select the most relevant examples to drive the learning process (Blum and Langley, 1997). This process of getting the best results from the data for your algorithms is referred to as feature engineering (Brownlee, 2014). When adding features and combinations thereof, the amount of training examples have to increase to reach a desired level of accuracy (Blum and Langley, 1997). As Machine Learning algorithms learn from data, it is vital to chose the right data for the specific algorithms you apply (Brownlee, 2015). Good data is data that is adapted to address the question or problem that you are trying to solve (Brownlee, 2015) which varies depending on the hypotheses and selection of algorithms. An Embedded System with a neural network approach requires very large amount of data to get adequate result[9]. However a Transactional Processing System, such as the Netflix movie recommendation system may not require as high volumes of data, but rather the right type of data. In fact, it has been shown that data from ten new movie ratings is more valuable to Netflix's recommendation system than meta-data (Pilaszy, 2009).

The test data set is primarily used for measuring the performance of a model (Keller, 2016). Many different testing data sets are needed for creating an optimal model and testing is a highly time-consuming process[10].

The data validation set is is used to tune the model, one example could be for pruning a decision tree (Keller, 2016). Additionally, the model is validated by using non-computerized technologies, such as experts, to determine whether the result generated is applicable. For example, when constructing a recommendation system for medical treatment an expert(i.e. a physician) would be consulted to determine if the model is recommending the right treatment[10].

4.1.2.4 Application Program Interface (API) Technologies

Application Program Interface (API) is code that enables different software programs, such as operating systems and other applications, to communicate with each other (Rose and Li, 2004; Orenstein, 2000). The characteristics of API technologies vary. In some instances it may take the form of a library with specifications on routines, data struc-

[9]Matthew Shreve, Parc, Interviewed 2016-03-09
[10]Marzieh Nabi, Parc, Interviewed 2016-03-08

tures, and other variables defining the application (Authorize.Net LCC, 2015). In other instances an API includes a set of standardized requests or implemented function calls that defines the proper way to request services for specific applications (Rose and Li, 2004; Orenstein, 2000). In simplified terms, API technologies can be viewed as doors to a building, where the building represent an application, and the external surrounding of the building represent the external surrounding of the application (Orenstein, 2000). In MLS, APIs facilitate performance on many levels. First of all, API technologies may abstract the inherent complexity of Machine Learning algorithms and facilitate management of the data system and the heavy infrastructure that is often required to enable learning (Martin, 2015). Additionally, APIs add traceability and repeatability to Machine Learning operations and tasks (Martin, 2015). From a business perspective, APIs may also drive adoption and even further improvement of Machine Learning powered products and services as it may serve as a platform for external developers (Martin, 2015). Furthermore, business interest in API technologies has increased as a result of increasing demand for cloud computing services since these requires integration of the cloud provider's service with on-premises systems (Rose and Li, 2004). The close connection between cloud services and Machine Learning APIs is demonstrated by that the largest cloud service providers, such as Google, Amazon, Microsoft and IBM, also top the list over the most frequently used Machine Learning APIs (Yegulalp, 2016).

4.1.2.5 User Experience (UX) technologies

Despite the growing interest in User Experience (UX), there is yet no universal definition of the concept or the technologies covered (Law et al., 2009). In this study, the definition of UX has been based on the ISO definition stated:

"[a] person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service" (International Organization for Standardization (ISO), 2010)

Hence, UX is not only confided to the actual consumption of the service, but also the users' emotions, perceptions and responses before and after using the product. This means that that experience is not only a function of the performance or presentation of the service, but also a consequence of brand image, the user's skills and context of use (International Organization for Standardization (ISO), 2010). UX technologies and design serves to improve utility, ease of use, and pleasure from the human interaction with a product or service in order to increase customer satisfaction and loyalty (Kujala et al., 2011). A broad interpretation of the ISO definition would then include the entire MLS, however, in this section it has been narrowed to the technologies behind the immediate interaction with the user.

According to Mike Kuniavsky, UX Designer & Principal Scientist at Parc[11], in order to provide a good user experience, the service or product powered by the MLS must

[11] Mike Kuniavsky, Parc, Interviewed 2016-03-22

be perceived as meaningful, trustworthy and seamless. UX from a value creation perspective will be covered in depth in section *Sources of Competitive Advantage for MLS.* From a technology perspective, the integration of the learning model in an application is fundamental to create a trustworthy and seamless MLS. Applications are software that is designed for performing specific tasks, operations and activities for the benefit of the user (PCMag, 2016). Examples of common applications are word processing programs and web browsers. In MLS, the Machine Learning algorithms can be viewed as a set of instructions expressed in programming language which properties are not optimized for hardware or an overall good user experience[12]. Hence, application technologies translate and integrate the Machine Learning algorithms into several computer programmes that in addition to running the algorithms more efficiently also perform other tasks specifically designed for the intended use. For Transactional Processing Systems, the application software also has the technical function of collecting user data[13]. One of the distinctive technical feature of Machine Learning is its ability to learn and adapt to specific user behaviors and by doing so providing a cognitive relief (Kuniavsky, 2016). However, if the application collects data in a way that compromise the perception of the service being seamless, then this may lead to circle of evil where the user will use the product of service less, resulting in less training data, resulting in reduced accuracy of the model and overall trustworthiness[14]

On top of the applications is another technology layer, which are technologies that the user can comprehend with their senses. Depending on if the MLS is embedded in a physical product, or delivered as a service or virtual product this domain includes Product and Service Design or Graphical User Interface (GUI). These technologies are characterized by layouts, colours, geometrical shapes, buttons, and materials which function may be both technical as well as aesthetic.

Another technology domain that have been deduced from interviews with researchers and practitioners is user training technologies. The MLS and all its technical functions and utilities cannot be unleashed without empowering the users with the skills required to operate the services or products. There are multiple ways of educating the user of an system e.g. documentation, on-line tutorials, in-person demonstrations etc.

[12]Matthew Shreve, Parc, Interviewed 2016-03-09
[13]Bob Price, Interviewed 2016-03-09
[14]Mike Kuniavsky, Parc, Interviewed 2016-03-22

4.2 Sources of Competitive Advantage for MLS

In this section, factors that can be linked to success or competitive advantage are investigated. The purpose is to identify the resources that are most valuable and strategically important for success. The empirical results related to this section stems from interviews with Machine Learning researchers, commercial actors and industry experts. A complete list of interviewed and interview template can be found in *Appendices*.

4.2.1 Data

When asked for sources of competitive advantage, all interviewees mentioned the importance of data or rather access to the right data. There are two main aspects stated in the interviews to why data is important for success. First of all, the performance and accuracy of a learning model always depend to some extent on the data quantity as well as the data quality[15]. If you do not have the data to train and test the system, the quality of algorithm does not matter. The second reason mentioned in interviews is that the outcome (eg. the movie recommendation in the case of Netflix) is a result of the data set which the system learned from[16]. Practically, this means that one and the same algorithm, trained on two different data sets, will generate two different outcomes. This property of MLS will be discussed further in Section *Technical Assets*.

Many of the interviewed also mentioned the ability to collect data cleverly[15,17,18]. Data may be difficult to access for several reasons. It may be proprietary, but it may also be that collecting the data requires cognitive ability or a level of human intelligence that machines currently cannot live up to[15]. Using crowd sourcing platforms such as the Amazon American Turk (mturk)[19] is today frequently used for collecting large and unique sets of data cheaply (Winter and Siddharth, 2012). Another example of smart data deployment, especially if the access to data is scarce, is to use off-the-shelf and pre-trained models that are then customized by training with unique data[16].

4.2.2 Human Resources

Another resource that is highlighted in all interviews is human resources. Oliver Downs, PhD., Chief Scientist and CTO at Amplero[20] stated that having the right composition of people and skills is essential for long term success.

[15] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29

[16] Matthew Shreve, Parc, Interviewed 2016-03-09

[17] Shivon Zills, Beta Bloomberg, Interviewed 2016-03-10

[18] Mike Kuniavsky, Parc, 2016-03-22

[19] mturk is an web-based marketplace launched by Amazon in 2005. The mturk was originally intended to be used for *human computing tasks*. Human computing tasks are micro-tasks that are very difficult or impossible for computers to perform, for example filtering adult content etc. Today, the mturk has grown into a platform with hundreds of thousands of workers, both human and software supported, making it a suitable platform for almost any type of data collection (Winter and Siddharth, 2012).

[20] Oliver Downs, Amplero Inc, Interviewed 2016-03-05

"[The company needs] to have the ability to turn an idea into an algorithm and make it work"

- Oliver Downs, PhD., Chief Scientist and CTO for Amplero, Interviewed 2016-03-05

In order to so, you need people that not only can understand customer needs, but also people that have the skills to transform those needs into accurate models and algorithms as well as integrating models into attractive services or products. Hence, according to Downs, PhD.[21], interdisciplinary and well-functioning teams spanning business, mathematical abstraction, mathematical modeling, software engineering and design are a key resource for success.

Similarly, Shivon Zills, Partner & Founding Member of the investment company Bloomberg Beta[22] who has specialized in machine intelligence and data driven companies, highlighted the importance combining people that understand customer needs with truly excellent User Interface designers and the right Machine Learning minds.

David Rose, CEO Ditto Labs Inc.[23], and Michael Sollami, Chief Scientist Ditto Labs Inc.,[23] also highlighted the importance of human resources to success. In their opinion, people that are talented in modeling and algorithm creation is of particular importance. According to Rose and Sollami[23] as there are off-the shelf solutions for Machine Learning models and algorithms the requirement on knowledge and skills have shifted from being able to create a model to being able to build *on top of* open-source frameworks and solutions. These skills are excluded the top researchers in the field, which are relatively few in numbers compared to the demand. In Rose's and Sollami's[23] opinions, the ability to build customized models on top of commodity models is the reason for why larger corporations are so actively recruiting top Machine Learning researchers. Another indication of the importance of human resources in general, and Machine Learning capabilities in particular, is the trend of leading MLS companies, such as Google, to release their Machine Learning APIs and algorithms[24]. Rose and Sollami[23], believe this is a strategy for organizations to identify and recruit Machine Learning talents.

4.2.3 Technical Assets

In the scenario when access to data is not a differentiating factor, for example if the data is public, then technical assets becomes more important for differentiation[23,25]. For

[21] Oliver Downs, Amplero Inc, Interviewed 2016-03-05
[22] Shivon Zills, Bloomberg Beta, Interviewed 2016-03-10
[23] David Rose & Michael Sollami, Ditto Labs Inc., Interviewed 2016-03-02
[24] Google open sourced its latest Machine Learning software library, TensorFlow, 2015-11-09 (Dean and Monga, 2015). In 2016-01-25 Microsoft followed and announced the open source of CNTK, a deep learning toolkit (Linn, 2016)
[25] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29

smaller companies where Machine Learning is the very core of the product or service, the know-how and innovations related to the learning model may become a key differentiating factor[26]. Innovative solutions related to subjective classifiers and dynamic validation of hypotheses are technical assets that have been brought up in interviews as key for long-term success[26,27]. Another finding is that multiple learning methods are often used, where different algorithms address different aspects of the problem[28]. However, technical assets related to back-end technologies, i.e. Data System, Network Technologies are often commodity solutions[29,30,27]. There may be value in how different solutions are combined, or if it solves a specific customer need. For example, Amplero's customers wanted the service to be delivered as SaaS, but using conventional cloud computing solutions was not an option for security reasons[28]. Hence, Amplero developed proprietary technology, creating a private virtual cloud to solve a customer problem[28].

Although many of the interviewed stressed the importance of technical assets for success, they also emphasize that the customer seldom notices improvements in system performance[28,31,26]. However, the customer does care about the ease of use, flexibility, and what the product or service enables them to do[26].

4.2.4 Data Driven & Innovative Over Time

Two capabilities that often were mentioned in conjunction are the ability to be innovative over time and being data driven[32,28,26]. Being data driven means that the organization has a culture and/or processes for basing its strategy and business decisions on data insights. The capability of being innovative over time is also related to company culture and/or internal processes, and is described by Downs[28]as the ability to develop and integrate new knowledge. According to Downs[28], in order to be innovative over time, the organization must develop and manage internal processes as well as external processes for knowledge acquisitions. Down[28] further emphasized the importance of establishing close relationships with prominent universities and research institutes for acquiring new knowledge, either through collaborations or recruitment.

4.2.5 User Experience (UX)

Another success factor that has been stated during interviews is the importance of creating a superior User Experience (UX). UX was describes from a technology perspective in Section *User Experience (UX) technologies*, and is a more holistic concept compared to individual resources and will therefore be treated as a separate entity. The majority of the people interviewed for this study state that the foundation of competitiveness starts

[26] David Rose & Michael Sollami, Ditto Labs Inc., Interviewed 2016-03-02
[27] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29
[28] Oliver Downs, Amplero Inc, Interviewed 2016-03-05
[29] Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05
[30] Matthew Shreve, Parc, Interviewed 2016-03-09
[31] Mike Kuniavsky, Parc, 2016-03-22
[32] Shivon Zills, Bloomberg Beta, Interviewed 2016-03-10

with solving a real customer problem. According to Kuniavsky (2016) the sophistication of technology has reached a level in which connectivity or smartness can be embedded in almost any artifact or service, the difficulty lies in solving a real problem for the user. Furthermore, Kuniavsky (2016) claims that Machine Learning technology has the capacity to create value for the user through cognitive relief, reducing the burden of making decisions. Although Machine Learning has the technical capacity of solving real customer problems, identifying the problems and formulating relevant Machine Learning hypotheses are the true challenges[33,34,35].

Another aspect of customer experience that has been emphasized is designing how the system learns. Examples of designs that have failed on this are applications that probe the user for answers, resulting in cognitive burden rather than a cognitive relief[36]. If designed poorly, there is an evident risk of a catch 22 situation, where the product or service requires usage to reach sufficient performance level, but users will not use the product or service because its performance is inferior or inadequate[37]. According to Kuniavsky[36], in order to provide cognitive relief, the job of training the system should not be on the user.

Building trust and reducing uncertainty is another aspect to user experience that is brought up in relation to success. There are very few, if any, MLS that are 100% accurate, and the user tend to judge the product or service on the things it gets wrong rather than all the results that are accurate(Kuniavsky, 2016). A few inaccurate results may shatter the users confidence and stop them from further consuming the product or service(Kuniavsky, 2016).

Lastly, intuitive and appealing design[36] as well as educating the user are key factors for creating a positive user experience[38]. Both Downs[50] and Shreve[39] view data visualization as a differentiating factor and a source to competitive advantage. Nest, market leader in smart thermostats, can be used as an illustrative example of the impact of appealing design to success. The Nest thermostat was neither the first nor the most accurate predictive thermostat, but the physical appearance and the GUI were more aesthetically appealing compared to competing solutions [36]. Downs[50], Rose and Sollami[40] also stressed the importance of guiding the user, helping them develop the skills required to maximize the value of the product or service.

[33]Hoda Eldardiry, Parc, Interviewed 2016-02-22
[34]Shivon Zills, Beta Bloomberg, Interviewed 2016-03-10
[35]Danny Bobrow, Parc, Interviewed 2016-03-09
[36]Mike Kuniavsky, Parc, Interviewed 2016-03-22
[37]Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29
[38]Oliver Downs, Amplero Inc, Interviewed 2016-03-05
[39]Matthew Shreve, Parc, Interviewed 2016-03-09
[40]David Rose & Michael Sollami, Ditto Labs Inc., Interviewed 2016-03-02

4.3 Control Mechanisms

This section includes the results from interviews and literature on how different control mechanisms are used to prevent imitation, and also, what benefits and drawbacks that are connected to each control mechanism for Machine Learning Systems (MLS). The data collection has been structured around the control mechanisms defined in theoretical framework (see Section *Control Mechanisms*).

4.3.1 IPRs

According to Sam Funnell[41], IP Manager at Stratified Medical with 25 years of experience in working with IPRs in the ICT industry, there is no real difference between using IPRs for conventional software compared to Machine Learning software. However, there are several ways that IPRs are used in MLS. IPRs are getting harder to obtain for MLS as human skills are needed to justify the intellectual rights[41]. Shivon Zills, Partner & Founding Member of the investment company Bloomberg Beta, IPRs are not an effective mechanism of control for MLS businesses, but rather the right data and relationships [42]. However, Oliver Downs, CEO of Amplero states that unique IP, like patents or trade secrets, are employed as protection for competitive duplication, but that continuous development and progression is more important for long-term success[43]. Eran Kahana[44], experienced IP lawyer at Maslon LLP and Research Fellow at Stanford University Law School CodeX, stated that patents and copyright still are effective tools for protecting current state of MLS. However, as Machine Learning technologies advances and applications becomes more autonomous, there may be legal implications related to infringement[44].

4.3.1.1 Patents

It has become increasingly difficult for companies to use patents as means of control for software implemented inventions[41]. The vast majority of the people interviewed in this study agree on that the the commercial value of patents covering Machine Learning technologies and software in general are questionable. Many algorithms and models are publicly available, thus patents have receded in significance[45]. Relating to the hardware, MLS typically use commodity solutions and only the really big organizations innovates and patents hardware for MLS[45]. For the technology fields surrounding the Machine Learning model, software tools are important for improving technology, collecting data and building an ecosystem[45]. The tools might be patentable but the effectiveness of such patents for control is questionable[45]. However, patents covering hardware are still considered to be effective since infringement are easier to detect and prove[45]. Patents

[41] Sam Funnell, Stratified Medical, Interviewed 2016-04-19
[42] Shivon Zills, Beta Bloomberg, Interviewed 2016-03-10
[43] Oliver Downs, Amplero Inc, Interviewed 2016-03-05
[44] Eran Kahana, Mason LLP & Stanford Law, Interviewed 2016-04-04
[45] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29

and design patents covering user interface features are by some of the interviewed considered effective[46], and Funnel[47] reported that she has seen an increased presence of design patents for on-line product and services. There are however testimonials cotradicting the usefulness of design patents for control[48]

One explanation to the reduction in effectiveness of patents for MLS is that software patents are increasingly difficult to enforce in the court[47]. According to Funnell[47] there are two main reasons to why enforcement is challenging. Firstly, detecting infringement is very difficult, especially for technology that is running in the back-end of the system. Secondly, there is a declining rate of patents actually standing up in court. *Alice v. CLS Bank* is an influential court case where the eligibility of software patents undoubtedly was restricted (Quinn, 2015).

The underlying difficulty with patenting Machine Learning technologies and software in general, is that they consist of mathematical models and abstract ideas, and subsequently at greater risk of rejection or invalidation[47]. Business methods face similar challenges[47]. In U.S., there is no requirement on further technical effect which means that more abstract inventions may be patentable[47]. According to Funnel
footrefsam1, Apple's patent on overscroll bounce[49] covers an invention that would not fulfil the criteria of further technical effect. Another explanation to low commercial value of software patents are that they are often easy to invent around[48]. The risks of invent-around and invalidation highlight another complication with using patent as means of control, namely the trade-off between exclusivity and disclosure[47,50,51]. As described, inability to exercise the exclusive right may refer to multiple scenarios in MLS; invalidation, invent-around or reverse-engineering and patent rejection[47,50,51]. Downs[50] believes that the risks associated with disclosing a Machine Learning invention has resulted in declining use of patents.

Another factor that may hamper the ability to exercise the right for organisations applying MLS, is the costs for enforcing the right in court[50]. Costs associated with patents are generally high, and patenting Machine Learning inventions and other software based inventions may be hard to motivate unless you have substantial financial means[52]. However, there are other benefits with obtaining patents. For example, Rose and Sollami[53] stated that patents are most useful for flagging for investment or acquisition, and in fact Ditto Lab rarely goes through with the patents they file for. Additionally, patents may

[46] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29

[47] Sam Funnell, Stratified Medical, Interviewed 2016-04-19

[48] Mike Kuniavsky, Parc, Interviewed 2016-03-22

[49] Apple's overscroll bounce patent covering "List scrolling and document translation, scaling, and rotation on a touch-screen display"(Apple Inc., 2007) has been debated in media and its validity was attested by the US court in the lawsuit between Apple and Samsung.(Essers, 2014)

[50] Oliver Downs, Amplero Inc, Interviewed 2016-03-05

[51] Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05

[52] Eran Kahana, Maslon LLP; Stanford Law School, Interviewed 2016-04-04

[53] David Rose & Michael Sollami, Ditto Labs Inc., Interviewed 2016-03-02

be useful as leverage in disputes[54].

4.3.1.2 Copyrights

Although source code generally is considered copyrighted work, there may be implications for code generated by model as it does not fulfil the requirement of human authorship [55]. There is well known case where a monkey takes a picture when a photographer left its camera and where the photographer claims copyright. However, as the author is not human, the court ruled that the picture not to be copyrighted work[55]. According to Kahana[55], the same reasoning would apply if for code created by a machine. According to Funnell[56] there is a balance of fair use when it comes to copyright for Machine Learning, especially in regard to data. Although data itself is not copyrightable, databases used in MLS may be considered copyrighted work (U.S. Copyright Office, 1997), provided that they meet the requisite of creativity. Copyright is considered a good protection in the sense that it is straight forward and can be applied on product or service instructions[57]. However, the lack of court cases indicates that the usage of copyright as a defensive mechanism for MLS is low[56].

4.3.1.3 Trade Secrets

For a Machine Learning inventions some of the most valuable know-how is kept as trade secrets[58,54]. Such know-how is rather abstract and thus often difficult to control through other mechanisms[57]. Examples on trade secrets is the know-how related to the combination different technologies in MLS[58] or the know-how of good hypotheses to solve a specific problem[59]. Also, the know-how related to validation, testing and the selection of learning rates are often protected by trade secrets (Kumar, 2016). Secrecy and trade secrets for MLS are often controlled through Non-Disclosure Agreements[56]. Section *Secrecy* and Section *Contractual Control & Relationships* further describes secrecy.

4.3.1.4 Trademarks

According to Funnel[56], it is important to understand the difference between branding and technical vocabulary to obtain a strong brand from a commercial perspective, and a strong trademark from a legal perspective. Trademarks are primarily used for brand assets such as logo, name of product and services and elements related to the user interface, e.g. website in MLS[56,57].

[54] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29
[55] Eran Kahana, Mason LLP & Stanford Law, Interviewed 2016-04-04
[56] Sam Funnell, Stratified Medical, Interviewed 2016-04-19
[57] Oliver Downs, Amplero Inc, Interviewed 2016-03-05
[58] Bob Price, Interviewed 2016-03-09
[59] Hoda Eldardiry, Parc, Interviewed 2016-02-22

4.3.2 Technical Control

Technical control constitutes and important control mechanism in MLS, both for internal and external control. Internally, technology is frequently used for implementing trade secret programs by making sure that employees only can access code or information that is relevant for their job and position[60]. The importance of controlling access to information is demonstrated by that there are a number of companies specializing in these types of access services, eg. Github, Softology[60]. Technology can also be employed for establishing external control. According to Funnell[60], in the modern ITC environment there is really no need from a usage point of view for the software provider to disclose source code. Hence, technical barriers are used for hindering users to access the source code to application programs or other software[60]. Another example of technical barrier is the usage of cloud computing. Ditto Lab develop, train and run their Machine Learning models on private clouds, hindering the user from ever getting in contact with the algorithms[61]. Also, technical barriers are used for web and data security[60]. According to Funnell[60] web and data security is especially important in MLS in which personal or sensitive data is collected. An example of how technical control can be used to ensure data privacy is Apple. In 2014, Apple specifically modified its software to ensure that it cannot unlock customer phones and decrypt customer data (Zetter, 2016).

API and software lock-ins have in some instances been found to have a negative effect on competitive advantage in the Web 2.0 era (O'Reilly, 2005). However, standards related to user interface features, such as Apple's overscroll bounce feature, may create very effective lock-in effects and overall control for MLS[62]

4.3.3 Secrecy

Secrecy as means of control for MLS is often manifested in trade secret programs or in contracts such as Non-Disclosure-Agreements [60,63]. According to Matteo[63] defensive publication can be a powerful weapon for hindering competitors from obtaining patents in a field. Another control aspect related to defensive publication is that it contributes to positioning the company as a technology leader and create interfaces to academia[63]. Other aspects of secrecy are further described in Sections *Trade Secrets* and *Contractual Control & Relationships*.

4.3.4 Contractual Control & Relationships

Despite the inherent risks of breach and imperfect information Funnell[60] experiences contracts to be the most effective control mechanism for software-based businesses, such as MLS. Contracts are used both internally and externally. Internally, contracts are

[60]Sam Funnell, Stratified Medical, Interviewed 2016-04-19
[61]David Rose & Michael Sollami, Ditto Labs Inc., Interviewed 2016-03-02
[62]Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29
[63]Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05

often used to control the employees[64]. In the ICT sector, the turnover rate on employees is high and there is a need for firms to make sure that people do not bring confidential material with them to their next assignment[64]. It is also important to ensure that new employees do not exploit intellectual assets from a former employment in an unfair, illegal or simply disadvantageous manner. For that reason, software engineers and programmers may be asked to sign a contract stating that they commit to produce fresh code[64]. Externally, contracts may be used to control the usage of the product or service, i.e. through a software license. Contractual agreements are also used for securing access to resources that the company does not posses. For MLS, contracts are commonly used to control and secure access to data and technology[64]. Contracts may also be used to extend protection of a company's IP, i.e. through IP licensing[64]. However, according to Matteo[65] understanding and aligning interests is key to establish strong contractual control. For Amplero, relationships with academia and research institutes has been very important for knowledge acquisition as well as public recognition[66].

4.3.5 Market Power

The findings from interviews suggest that having a strong brand generates control for MLS in several ways. Firstly, having a reputation or brand signaling technology leadership is a way to secure recruitment of prominent researchers and talents[67]. Secondly, Machine Learning and AI are still experienced as intimidating and unsafe for many users, and trust and reputation are key components to secure consumption[64]. Thirdly, a strong brand can generate control in the form of a physiological lock-in. Customers' purchasing decision is affected by which brands they are aware of, and positioning the brand on the top of the consumers mind can therefore block the customer from identifying substitutes[64,65].

4.3.6 Business Model

Damon Matteo, CEO of Fulcrum Strategy and Advisory Board Member at the Stanford Hoover Institute[65], stated that the business model itself may be a source of sustained competitive advantage in MLS. Management and execution are two related concepts that have been mentioned by several of the interviewed related to sustained competitive advantage[68,64].

[64] Sam Funnell, Stratified Medical, Interviewed 2016-04-19
[65] Damon Matteo, Fulcrum Strategy, Interviewed 2016-04-05
[66] Oliver Downs, Amplero Inc, Interviewed 2016-03-05
[67] David Rose & Michael Sollami, Ditto Labs Inc., Interviewed 2016-03-02
[68] Industry Expert, over 40 years of experience in AI, Interviewed 2016-03-29

4.4 Case Studies

In interviews, both Amazon and Netflix have been exemplified as companies that achieved long-term success and where Machine Learning is a core element of their business. It can therefore be argued that MLS have contributed to their long-term success and that a deeper understanding of their technology and business model will give valuable insights to potential sources of competitive advantage. This section includes a short description of each company's history, MLS characteristics and the business model.

4.4.1 Amazon

Amazon is considered one of the largest, if not the leading, retailer in the world, with more than 304 million active customer accounts worldwide (Amazon, 2014)(Noren, 2013a). Amazon.com offers millions of unique products either through their mobile websites or apps along with local services, computing services and digital content (Amazon, 2014). Amazons business can be divided into three different parts, the Marketplace, Prime and Amazon Web Services (AWS) (Amazon, 2014). The Marketplace refers to the web- page where third-party sellers compete against Amazons retail category products (Amazon, 2014). Amazon Prime is an annual membership program that includes unlimited free shipping of millions of items, unlimited instant streaming of movies and TV-shows and access to hundred thousands of books to borrow and read on Kindle devices (Amazon, 2014). AWS is a platform for on demand cloud computing services (AmazonWebServices, 2016). Machine Learning technologies can be detected in multiple elements of the Amazon business. Amazon rely on predictive analytics, i.e. Machine Learning, to improve forecasts of customer needs that governs the entire replenishment activity (SourceMedia, 2016; Metz, 2015). Amazon has created a large ecosystem with interlinks between their different products and services thus, Machine Learning technologies can be detected in all the different segments.

4.4.1.1 Amazon's History

Amazon was founded in 1994 by Jeff Bezos and launched on on the web in 1995 (Amazon, 2016). Initially, it was a book store which rapidly grew reaching one-millionth customer only two years after opening (Amazon, 2016). Amazon's vision is to provide shoppers with anything they might want to buy on-line, both products and services. In 2002, Amazon opens up its technology platform, where developers can build applications and tools and incorporate the features of Amazon.com into their own websites (Amazon, 2002). This would have a large impact as it did transform the whole ecosystem around developers (Stratecery, 2016). Several years later, in 2006, user and costumer offerings were enhanced by enabling small businesses to grow via *Fulfillment by Amazon* and *WebStore by Amazon* (Amazon, 2016). In 2007, Amazon launched a new part of their business, the wireless e-reader Kindle. This concept was developed one step further in 2011, introducing Kindle Fire, a multimedia tablet offering more than 18 million movies, TV shows, songs, books, magazines, apps and games (Amazon, 2016). Only two years later, Amazon releases their first original TV series, *Alpha House and Betas* (Amazon, 2016). In 2013, Machine Learning solutions, such as a camera-based identification and tracking system, were implemented in the fulfillment centers (Metz, 2015).

The Amazon Prime was launched in 2005. Originally it was designed as an all-you-can-eat free and fast shipping program (Amazon, 2014). The foundation of Prime is the inventory retail business where a large-scale system automates much of the inventory replenishment, inventory placement and product pricing (Amazon, 2014). Further the worldwide network of fulfilment centers allows a precise delivery-date promise to its customers (Amazon, 2014). This has enabled Prime Now, which is available in a selected number of cities in the U.S., and offers a free two-hour delivery on tens of thousands of items at a low cost. In 2011, Amazon Prime was further expanded adding offerings such as Prime Instant Video, a benefit where movies and TV episodes are available for unlimited streaming (Amazon, 2014).

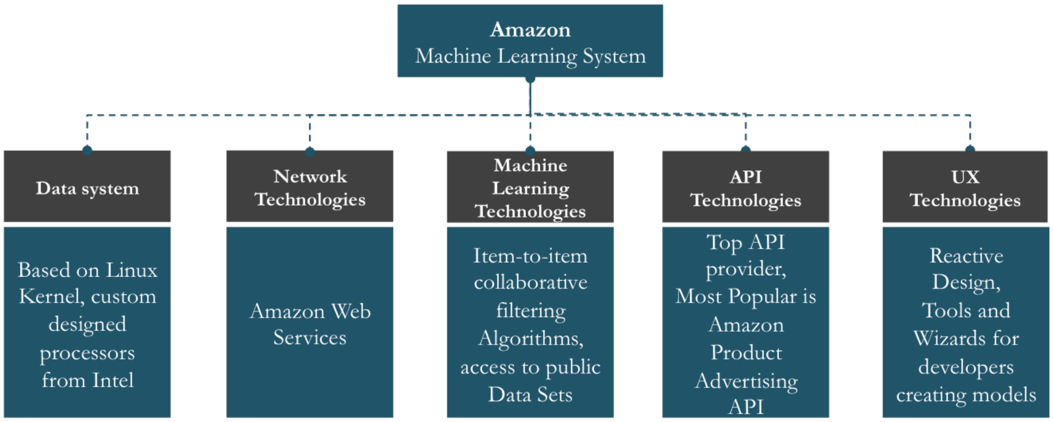
Amazon Web Services (AWS) is another radical idea that increasingly has grown since its launch in 2006 (Amazon, 2014). The focus is different from the other business as the approach is to help companies manage their IT better and faster at an attractive price (Amazon, 2014). Initially AWS was a service infrastructure for the development team within Amazon, however the potential of building an *infrastructure service for the world* was realized and opened up for the world (Clark, 2012). In 2015, AWS announced the launch of a Machine Learning platform, where anyone can create data driven applications that quickly can be built and validated (Miller, 2015). Today, all sizes of companies and organizations uses AWS in every imaginable business segment reaching more than a million active customers (Amazon, 2014).

4.4.1.2 Amazons's MLS

Amazon's recommendation system customizes the browsing experience to create a personalized shopping experience for each customer (Linden et al., 2003). The MLS is based on massive amounts of data that is continuously collected (Konstan and Riedl, 2012) and immediately responds, regardless of the number of purchases and ratings (Mangalindan, 2012). Amazon's MLS is further applied in other parts of the Amazon business, such as the AWS platform or the fulfillment centres. The following section describes the Amazon MLS technologies. An overview of the most distinguishing technologies for each sub-technology in the MLS, can be found in Figure 4.5.

The core of Amazon's Data System migrated to Linux in 2003 and 2004 and the technology architecture handles millions of back-end operations everyday as well as quires from third party sellers (Vanghan-Nichols, 2012). It is believed that Amazon had close to half-a-million servers running on a Red Hat Linux variant in 2012 (Vanghan-Nichols, 2012). Amazon builds its own proprietary servers and racks, and its processors are custom-designed together with Intel as the market standard does not fulfil the requirements it need for its system (Kassner, 2014). Amazon applies open source technologies

Figure 4.4: Summary of Amazon's Machine Learning System



when using the Linux Kernel which is further applied in the AWS services using for example Xen and MySQL (Clark, 2014). The different services for AWS further apply different data systems, one example is Amazon Elastic MapReduce which based on the open-source Hadoop framework (AmazonWebServices, 2016a). Security is a major concern for Amazon as hundreds of thousands of people register their credit card numbers within the data system every day (Layton, 2005). Amazon applies the Netscape Secure Commerce Server using SSL protocol to store credit card numbers in separate databases that are not Internet accessable (Layton, 2005).

The network-intelligence start-up DeepField estimated in 2012 that one-third of all daily Internet usage accesses a site running on AWS which during the past years most likely has increased looking at the growth of AWS (Burrington, 2016). Amazon contracted an Original Design Manufacturer (ODM) to make custom networking gear as it could not adapt the off-the shelf networking gear and current protocols to meet load demands of the businesses (Kassner, 2014).

The algorithms for Amazon's recommendation system builds on the principle of *item-to-item collaborative filtering*. They are based on several simple elements such as what the user has bought in the past, which items they have in their virtual shopping cart, items they have rated and liked, and what other customers have viewed and purchased (Mangalindan, 2012). These algorithms are developed by Amazon as no off-the-shelf solutions are compatible with the massive data sets and scale of Amazons MLS (Linden et al., 2003). Machine Learning algorithms are also used in other parts of Amazons business such as the Amazon Machine Learning platform. Developers can access some of the Machine Learning models developed by Amazon to process new data and generate predictions for their own application (AmazonWebServices, 2016). The Machine Learning models available are binary classification, to predict a binary outcome, multiple classifications to generate predictions for multiple classes or regression models to predict a numeric value (AmazonWebServices, 2016c). AWS further provides a repository of public data sets within biology, chemistry, economics or encyclopedic available in two formats, Amazon EBS snapshots or Amazon S3 buckets, that can be accessed for your own applications (AmazonWebServices, 2016b).

For developers, the most valuable parts of AWS are the services and APIs built on top of the infrastructure (Burrington, 2016) which makes it possible to create applications with capabilities such as pattern recognition and prediction (Wagner, 2015). AWS has gained a lot of popularity which have turned Amazon into one of the top the API providers in the world (Wagner, 2015). Using the Amazon APIs, anyone can build applications that feature for example fraud detection, content personalization or document classification (Wagner, 2015).

Amazon uses a flowable or fluid page design unlike many sites which enables it to make the most of real-estate on-screen[69] (Chaffey, 2014). Amazon's Machine Learning services are said to be more complicated than for example Google Prediction or Microsoft Azure ML, however they are providing visualization tools and wizards that help users with the process of creating the models (Wagner, 2015). The Amazon User Interface (UI) applies a reactive design, where Amazon is able to react to a customer's engagement with personalized attention (Prunty, 2014).

4.4.1.3 Amazon's Business Model

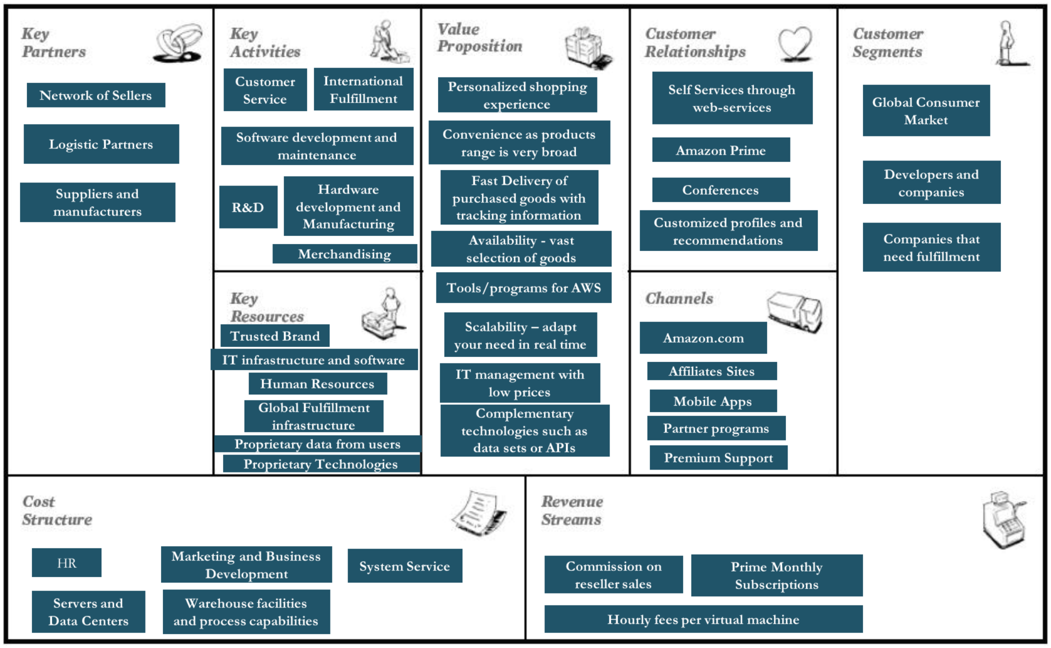
This section provides an overall picture of Amazons business model. Figure 4.5 illustrates the model and each box is further explained with emphasis on the elements that are relevant to this study.

Value Proposition

Amazon's value proposition consists of several core values which are consistent throughout their business. Initially, the vast selection of products combined with a great price and great customer experience was the unique offering that attracted 1.5 million customers between 1995 and 1997 (Amazon, 2014). Today, in addition to cast selection, Amazon provides a customized shopping experience facilitated by a Machine Learning based recommendation system (Mangalindan, 2012). By allowing sellers to offer their products side-by-side, Amazon has created a one-stop on-line shopping experience with easy to access products with large availability (Amazon, 2014; Mochari, 2016). Thus, the foremost value for Amazons customers is the vast selection along with the customized product offering where customers are able to shop everything from toilet paper to shoes on one market place.

[69]Screen Real Estate refers to the amount of space available on a display for an application to provide an output where the aim is to have as much data and as much control as possible visible on the screen to minimize the need for hidden commands and scrolling (UsabilityFirst, 2014)

Figure 4.5: Amazon Business Model Canvas



AWS value proposition is to offer organizations improved and faster IT management to a low price (Amazon, 2014). Furthermore, the AWS offers scalability and adaptability to organizations and developers, as you only pays for what the business require at the moment (AmazonWebServices, 2016a). Complementary services such as tools and wizards are also available to improve user experience of AWS (AmazonWebServices, 2016). Additionally, Amazon has created an infrastructure providing complementary technologies such as data sets or APIs.

Another important value proposition is fast and reliable fulfillment with a friction-free delivery of your purchased good or services from the marketplace or AWS. Fulfillment by Amazon (FBA) is eliminating shipping fees and providing a timely customer service (Amazon, 2014). This is especially important for Prime customers as a core benefit is fast delivery of their products. As more sellers join FBA, more products become available and sellers increase their sales and the membership value of Prime increases (Amazon, 2014). Paid Prime memberships grew more than 50% in the U.S. and 53% worldwide in 2013 (Amazon, 2014). Another important element is the shipping support where Amazon was among the first companies to implement extensive use of tracking (Minsker, 2015). This enabled customers to be aware of the status of their package and Amazon can provide information and communicate with the customer before the problem was recognised by the buyer (Minsker, 2015).

Taking into account all the different offerings from Amazon it can be concluded that the most important for Amazon customers is the personalized shopping experience, the convenience of the product range, fast delivery and tracking information of goods, availability, scalability, complementary technologies and tools for AWS services, low prices and convenient IT management.

Customer Segment
The major segment for the e-commerce sales is the global consumer market, which includes all customers acquiring products or services from the Amazon web-pages where anyone could acquire products. Another large segmentis the companies using Amazon as a logistic partner by utilizing Amazons fulfillment centres. In the case of AWS, practically anybody with a credit card can be considered a costumer as it offers on-demand, pay-as-you-go cloud storage and compute resources, however it is mainly targeting software developers and companies (Amazon, 2014; van Eijk, 2014). From the beginning AWS was developed for the internal use of the development team, however, the concept evolved and succeeded as an unmet need for companies was identified. Today AWS is used by established organisations such as Pinterest, Dropbox and Airbnb (Amazon, 2014). Amazon's customer segments are summarized in the left side of Figure 4.5

Customer Relationships and Channels
Amazon's customers are self-served mainly through the retail website which can be accessed both directly on amazon.com or through mobile websites and apps (Amazon, 2014). AWS also employs a self-service direct model with delivery through APIs on a web user interface (van Eijk, 2014). The customers can also access premium support and there are extensive partner programs developed particular for SaaS providers (van Eijk, 2014). The Amazon Prime offering is another way to provide selected services for an additional fee. Although Amazon's primarily customer contact interface is their webpage, they also maintain relationships with its customer through conferences. In 2013 they hosted a re-invent developer conference that attracted 9000 visitors (van Eijk, 2014). Amazon strategy is based on a customer-centric culture that is employed by observing customers and noticing that things can always be better (Bishop, 2013). Thus, an extensive predictive analytics on the activity of customers on both the AWS platform and the websites for e-commerce sales is the very fundamental base that distinguishes Amazons relationship with its customers (van Eijk, 2014).

Revenue Stream and Cost Structures
The bottom part of the Ostwalder's Business Model Canvas (see Figure 4.5) consists of the cost structure and the revenue stream. In 2015, Amazon reached a business milestone by generating \$107 billion in revenue, surpassing \$100 billion (Roettgers, 2016). The e-commerce services mainly gain revenue by fixed fees, revenue share fees, per-unit activity fees or some combination thereof (Amazon, 2014). Thus, Amazon is taking a cut of every purchase simply for providing the channel and does not handle advertising or shipping (Noren, 2013b). Overlaying with the retail model, the Prime memberships

generates revenues in the form of annual fees and subscription for the "all you can eat" model cloudcomputing. Amazon takes pride in their selection, convenience and price and use economics of scale as they main strategy where low cost structures and high volumes are key factors. Furthermore, Amazon has a cost leadership strategy[70] with major warehousing facilities and processing capabilities driving economies of scale (Grundy, 2015). Other costs are marketing, technology research and development and payroll costs (Amazon, 2014).

The AWS revenue model differs from other IT service models as it has usage based rates rather than asset based rates (ClubCloudComputing, 2016). This mean that companies are charge hourly fees per virtual machine. Revenues from AWS increased with 69.37% year-over-year in the last quarter of 2015, closing at $2.4 billion in total year-of-year revenues (Novet, 2016). The cost structure in relation to AWS is mainly consistent in the elements of assets such as servers and data centers (ClubCloudComputing, 2016). There are also costs for system services, such as electrical power and telecommunications. HR costs for people developing and managing the systems is an additional expense (ClubCloudComputing, 2016; Amazon, 2014).

Key Activities

Amazon's key activities include retail interface design, back-end supply-chain management, and advanced technical innovation (Mochari, 2016). These activities can all be connected to analytics on the large amounts of data generated from their customers. A steady stream of automated machine-learned "nudges" are generated continuously which alert sellers about opportunities to avoid going out-of-stock, add selection that is selling or sharpen their prices to be more competitive (Amazon, 2014). Delivery is the key activity of AWS, and similarly to e-commerce activities, this activity is highly automated(ClubCloudComputing, 2016). However, oversight and resource planning are still high effort activities of AWS (ClubCloudComputing, 2016).

Amazon is highly involved in their international fulfillment service as well as their customer services. In 2012 they filed a patent titled "Method and system for anticipatory package shipping" which potentially could reduce shipping, inventory and supply chain costs (Ulanoff, 2014). This shows concretely how Amazon constantly is putting effort into optimizing their business and investing in RD activities.

Amazon invests heavily in new technology areas and business development and is known for betting on somewhat unconventional technologies. One example is the development of Flying robot drones to deliver products ordered from Amazon.com within 30 minuets (Strange, 2013). Another example is the announcement to build a "virtual reality experience within Amazon Video" which is a new technology segment which companies

[70] A cost leadership strategy is where similar or lower prices are charged compared to competition, but where the costs are certainly lower than competitors (Grundy, 2015),thus resulting in a higher profit margin

such as Netflix and Hulu also are exploring (Mason, 2016). Another investment is in voice recognition technology for the Echo device (Higginbotham, 2015). Amazon does not always succeed in their new technology launches, an example of that is the Amazon Fire phone launched in 2014 (Warren, 2014).

Like Google and Facebook, Amazon has developed its own equipment servers and data centers (Kassner, 2014). This includes networking gear, networking software, racks and processors which Amazon is developing together with Intel, to get processors even better than available on the market (Kassner, 2014).

Key Resources
The key resources for Amazon are the people working to develop and explore new technology fields and business opportunities for Amazon to invest in (Amazon, 2014). Vital is also the development of proprietary technologies, such as the recommendation system, that is the foundation for many of its value propositions (Mangalindan, 2012). The recommendation system is based on algorithms developed by Amazon, which makes Machine Learning technologies an important resource (Mangalindan, 2012). Additionally, the data is a requisite for many of the activities and Machine Learning technologies(Amazon, 2014; Konstan and Riedl, 2012). The installed base of customers, both for AWS but also the other businesses, has become one of the most valuable assets as it provides Amazon with constant data and continuous feedback[71]. Amazon has also succeeded in establishing a trusted brand. In 2015, the Amazon figured for the first time in the top 10 of Interbrandś ranking of the 100 most valuable brands (Sullivan, 2015). The global network of fulfillment centers, is another resource that has seen extensive growth, from 13 centers in 2005 to 109 in 2014 (Amazon, 2014). For AWS, hardware is a key resource (ClubCloudComputing, 2016). However this can not solely deliver the service, thus, extensive software and processes are also strategically important resources for Amazon (ClubCloudComputing, 2016).

Key Partners
For the AWS business Amazon has partnered with hardware vendors, i.e. Intel to provide the hardware required for its data centers and networking(ClubCloudComputing, 2016; Kassner, 2014). In addition, Amazon has founded a tiered Partner Network called APN where Technology partners or professional services firms gets access to various tools and support to more efficiently build their solutions (AmazonWebServices, 2016d). The tiered program basically builds on the principle of the more you contribute and engage with AWS, the more tools and support you will get access to. Technology partners include Independent Software Vendors (ISVs), SaaS, PaaS, Developer Tools and Management and Security Vendors (AmazonWebServices, 2016d). For Amazon's retail business, the most important partners are the sellers using Amazons channel for their products (Noren, 2013b). If the sellers provide the ”right” items resulting in increased sales, Amazon will benefit as well as the sellers. Amazon also has several logistics part-

[71]Bob Price, Interviewed 2016-03-09

ners to help providing the capacity that it needs (Bensinger, 2016).

4.4.2 Netflix

By the end of 2015, Netflix had over 74 million streaming members and an annual revenue from video streaming of over 7 billion USD (Netflix, 2016c). Netflix has throughout the years discovered the value in incorporating recommendations to personalize the streaming experience (Amatriain, 2013). In the end, everything is a recommendation for Netflix, which makes its Machine Learning technologies, manifested by the Cinematch recommendation system (Gallaugher, 2012) together with the surrounding technologies and architectures the very core of its value propositions.

4.4.2.1 Netflix's History

Netflix was founded in 1997 by Reed Hastings and Marc Rudoplh to offer online DVD rentals (Netflix, 2016a). In 1999 the company debuted with their subscription service, offering unlimited DVD rentals for one low monthly cost. In 2000, Netflix released Cinematch, a recommendation system using data mining and predictive analysis, i.e. Machine Learning, to recommend content for users based on user behavioral data (Kovacs, 2015; Reinsberg, 2009). Netflix early realized the potential in online streaming technologies, and in 2007 they introduced streaming as a free service bundled with its DVD rental subscription service (Kovacs, 2015). Despite having a very limited selection of streaming content, the subscription base increased from 7,5 million in 2007 to 12 million at the end of 2009 due to rapidly growing online television market (Kovacs, 2015). By 2011, the market had become increasingly competitive with actors such as Amazon, Hulu and Google threaten the position of Netflix with their own streaming services (Bushey, 2014).

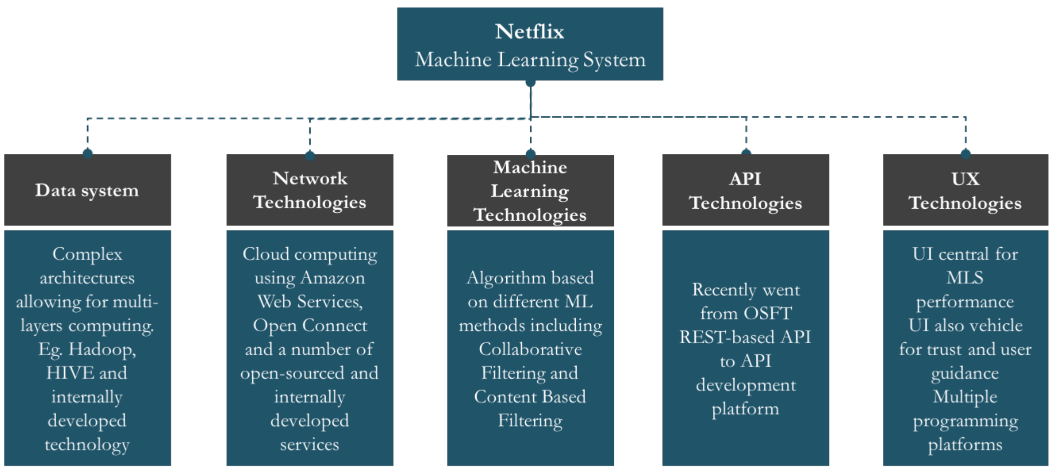
Netflix realized that its current business model was vulnerable to competition since 1) they depended on licenses from original content providers, which theoretically could be turned over to the highest bidder 2) The most popular streaming videos were sold to syndication at basic cable channels (Bushey, 2014). Netflix overcome this issue by analyzing television network data and their own customer data arriving at two main insights. The first insight was that the cable channels lost viewers on their serials since the viewers were forced to watch the show at certain dates and times. If a potential viewer missed a couple of episodes, data showed that they would not return or even begin to watch since they would not understand (Bushey, 2014). The second insight was that Netflix's own customer data revealed that their viewers enjoyed to binge-watch series. Netflix acted on this information and in 2013 they released whole seasons of two originally produced series (Cook, 2014).

Netflix still offers physical delivery of DVD rentals, although the streaming services have since long surpassed the DVD rental service. In 2014, the DVD rental business constituted 14% of total business and declined with 16% compared to 2013 (Netflix, 2015). For the purpose of this study results and analysis are solely covering the core business of Netflix, namely providing personalized streaming content.

4.4.2.2 Netflix's MLS

In this section the technologies constituting Netflix's MLS are described using the system break-down from Section *Machine Learning System Break-down*. A summary of the characteristics described in this section is illustrated in Figure 4.6.

Figure 4.6: Summary of Netflix's Machine Learning System



In order to enable real time recommendations, Netflix relies on architectures that allows it to combine complex offline batch processing with real-time data streams (Amatriain, 2013). The fundamental principle of the architectures is that processing is done at different layers (offline, nearline, online) whereas learning, features and model evaluation can be done at any level (Basilico, 2014). Netflix employs a mixture of open-source, off-the-shelf and proprietary technology for its Data System and Network Technologies (Netflix, 2016b; Basilico, 2014). For example, Netflix uses Amazon Web Services(AWS) as a cloud computing platform and Hadoop for distributed computing in. Other off-the-shelf technologies employed is Spark, a open-source cluster computer network for big data processing, and HIVE which is a open-source data warehousing solution (Basilico, 2014). Although the much of the technology foundation consists of open-sourced or licensed technology, Netflix also develops technologies on top on existing solutions to tune for Netflix use cases (Madappa, 2012). An example of such tuned technology development is EVCache, which is a caching solution optimized for AWS but tuned for Netflix usage (Madappa, 2012). Netflix has released EVCache, making it an open-source development platform (Madappa, 2012). Netflix also develops proprietary technology (Brodkin, 2014). One of Netflix's key technologies is the proprietary Open Connect content delivery network (Brodkin, 2014). Open Connect enables Netflix to deliver content to tens of millions of connected devices all over the world at any time (Brodkin, 2014). Without a sophisticated content delivery network, Netflix would not have been able to

handle data traffic during peak hours and Open Connect is therefore considered critical in creating a seamless user experience (Brodkin, 2014).
Netflix's Machine Learning technologies can be described as a ranking model that can use a wide variety of information to produce an optimal sorting of movies and shows (Amatriain, 2013). In order to present the most relevant recommendations for content, Netflix has constructed ranking algorithms that aim to optimize the ordering of a set of items for a user, within a specific context and in real-time (Amatriain, 2013). Netflix has not made its current Machine Learning algorithms and models available to the public, which means that the exact composition of algorithms remains unknown. However, Netflix has been public about that it is using a number of different Machine Learning approaches: from unsupervised methods such as clustering algorithms to supervised learning methods using classifiers (Amatriain, 2013). After the Netflix Prize competition in 2006, Netflix incorporated a number of different algorithms, all addressing different problems or challenges (Amatriain, 2013). Examples of algorithms that are known to be used in Cinematch after the completion of Netflix Prize in 2011 are Matrix Factorization (MF) and Restricted Boltzmann Machines (RBM) (Amatriain, 2013). Netflix combined these solutions with neighbouring Machine Learning methods and made adjustments to make them operational (Amatriain, 2013). In fact, the sophistication and accuracy of Netflix's recommendations stems from the combination of a vast number models, rather than specific classifiers or clusters (Paterek, 2012). In terms of training, testing and validation data Netflix uses a large number of different data sources, including ratings from members, global item popularity for ranking, search terms and queuing lists, content metadata, social data, external data (movie reviews etc.), demographics and impression data[72] (Amatriain, 2013). Netflix believes that its Machine Learning technologies depend on both its models as well as its data, meaning that large quantities of data can not compensate for poor models or vice verse (Pilaszy, 2009). User data, such as rankings, are much more valuable since this creates more accurate recommendations (Pilaszy, 2009). Also, the processes that Netflix has developed for testing the models are crucial for the overall performance (Amatriain, 2014)

In 2012, Netflix initiated an extensive redesign effort of its original REST-based API going from a traditional one-size fits all approach to a platform for API development (Christensen, 2013; Jacobson, 2012) building on JAVA. Before 2012, when the decision to redesign was made, the key audiences for the API were a small group of known developers which mostly consisted on internal Netflix UI developers (Jacobson, 2012). However, as Netflix have initiated more and more open source projects, the audiences have grown to also include a large set of unknown developers (Jacobson, 2014; Spyker and Meshenberg, 2015). Up until 2014, the Netflix API was available to public, allowing external app developers to access their data and software libraries (Jacobson, 2014; Bohn, 2013). However, as of November 14 2014, Netflix does not allow access to third party developers to their API program to reduce versioning (Jacobson, 2014).

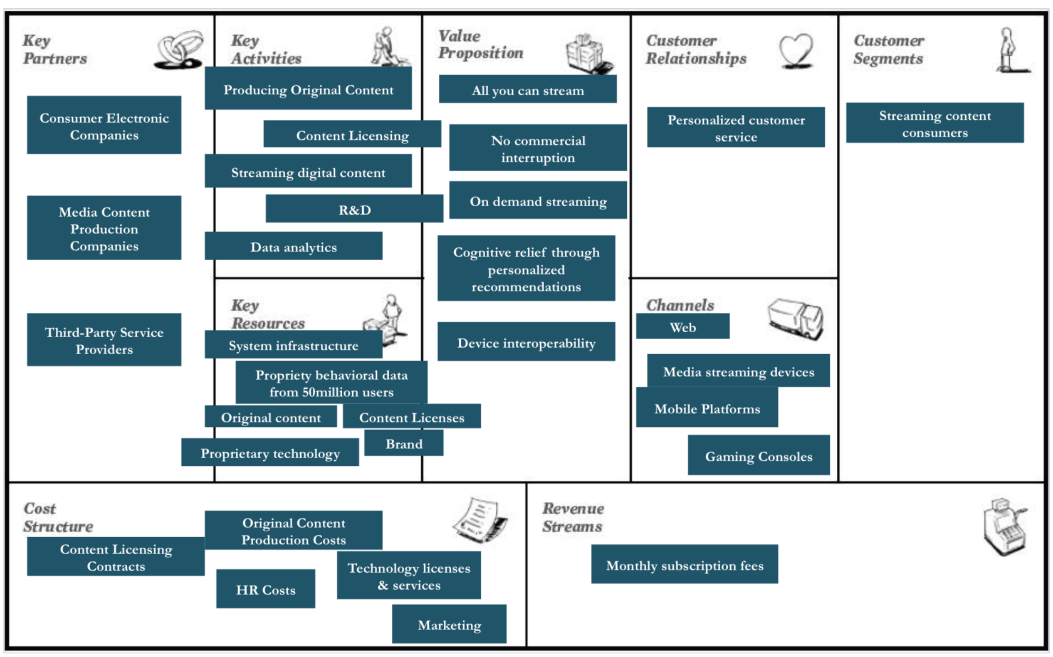
[72] Impression or presentation data is information related to prior actions given a certain event. For example, on what item did the user click after being recommended a certain item? (Amatriain, 2013)

In terms of UX technologies, Netflix uses open-source software integration tools, such as Jenkins, as well as develop their own software(Basilico, 2014). The User Interface (UI) is a very important technology field for Netflix as it generates the user feedback that feeds the ranking and recommendation algorithms (Amatriain, 2014). In 2015, Netflix launched a new user interface via their web site (Ingraham, 2015). The new UI was the result of years of research on how humans look for things to watch (Lowenshohn, 2015). The research included thousands of in person interviews, testing different designs on smaller audiences and email surveys to millions of users (Lowenshohn, 2015). UI design is also a vehicle for building trust and awareness of the system (Amatriain, 2013). According to Amatriain (2013), communicating why certain content is presented and recommended promotes trust as well as encourages users to give feedback that can be used for improving the models (Amatriain, 2013).

4.4.2.3 Netflix's Business Model

In this section, Netflix business model will be presented using the Business Model Canvas (see Osterwalder and Yves (2010)), and the model is illustrated in Figure 4.7.

Figure 4.7: Netflix's Business Model Canvas



Value Proposition & Customer Segments
Starting with value proposition, Netflix leverages several components to its value offering. First of all it provides online streaming to consumers (Netflix, 2016a). For a fixed

subscription fee, Netflix's customer may consume as much content as they want, without any interruptions from commercials (Gallaugher, 2012; Bushey, 2014). The fact that the service is free from non-Netflix advertisement has since the launch of the streaming services been a cornerstone of Netflix's value proposition. Originally, it differentiated Netflix from conventional television (Gallaugher, 2012). Commercial independence is still an important peice of the value proposition as well as a vehicle for trust. This may relate to that data privacy is becoming increasingly important for brand value and ultimately business operations. This can be demonstrated by the impact that the Prize I lawsuit had on Netflix's business in 2009. Netflix was sued after having released user data in connection with an open contest for machine-learning scientists, called the Prize I Program (Zenor, 2014). After the lawsuit was settled in 2011, Netflix experienced a drop in net profits with 14% in the final quarter (Zenor, 2014). Although the Privacy Terms of Netflix service is frequently questioned, Netflix maintain that they do not sell any of the user data to third-parties, such as advertisement firms (Popper, 2015). The majority of research published on Netflix, including Netflix own publications, view the personalized recommendations as the core of the service and a source of competitive advantage (Gallaugher, 2012; Kovacs, 2015; Cook, 2014; Amatriain, 2013). The utility of such recommendation can be described as a cognitive relief, where the customer does not have to search through the vast amount of content in the Netflix library to find something they may enjoy (Kovacs, 2015). Finally, Netflix offers value through device interoperability. Instead of a specific hardware, Netflix can be run on a variety of devices and platforms(Gallaugher, 2012).

Channels
These devices and platforms serve as the company's main channels, illustrated in Figure 4.7. In addition to Netflix's webpage, users may also access the service though hundreds of media streaming devices such as Apple TV, Google Chromecast, consumer electronic products from all the larger vendors (Gallaugher, 2012).

Revenue Streams
Netflix rely solemnly on revenues generated from from streaming customers in the form of monthly subscription fee(Netflix, 2015; Cook, 2014).

Customer Relationships
Regarding the customer relationships, Netflix almost exclusively has an online relationship with their customers (Basilico, 2014). However, Netflix have managed to differentiate themselves from main competitors by encouraging their support staff to be themselves (Stenovec, 2013) when chatting with customers. For example, Netflix received a lot of positive media attention after a support chat transcript was published at Reddit in 2013 (Stenovec, 2013). In the conversation, the Netflix representative impersonated his favorite tv-show character when helping the customer.

Key Activities
Looking at the key activities of Netflix, streaming and licensing digital content are two of the main activities (Kovacs, 2015). Another key activity is the production of original content (Kovacs, 2015). Two of Netflix's first original productions were the tv-shows *House of Cards* and *Orange is the new black* which by have been successful by almost any measures (Kovacs, 2015; Shaeffler, 2015). The decision to produce those particular shows was the result of advanced data analytics on customer data (**?**). Using data insights to select tv-shows to produce is only one of many data-driven decisions that Netflix have made throughout the years, making data analytics is yet another key activity of the business. Finally, although Netflix rely on several third-party service providers for their technology platform, they also develop their own technologies (Gallaugher, 2012). Hence, R&D is central for the Netflix's business model.

Key Resources
In terms of key resources, since the value proposition and revenue streams are created from streaming content, the access to content is a key resource (Gallaugher, 2012). As previously mentioned, Netflix have two sources of entertainment content. Firstly, exclusive licenses to content produced by television and picture networks and secondly originally produced content (Gallaugher, 2012; Cook, 2014). The system infrastructure and Machine Learning technologies are also key resources (Kovacs, 2015; Basilico, 2014). Key technology resources are described in Section *Netflix's MLS* A premises for MLS or any type of data analytics activity is data (Amatriain, 2013). The proprietary behavioral data from approximately 50 million subscribers (Netflix, 2016a) generating billions of unique data items (Amatriain, 2013) is another key resource of Netflix. Finally, the Netflix brand itself is an important resource for Netflix's business. According to findings from the 2014 Harris Poll EquiTrend®[73], Netflix's brand equity increased at the fastest rate among the top 100 measured brands between 2012-2014 (PRNewswire, 2014).

Key Partners
Netflix's key partners directly follow from the segments of the business model canvas already described. In order to deliver device interoperability (see *Value Proposition* in 4.7), Netflix has partnered with the leading consumer electronics and mobile platform companies, such as Apple, Android, Nintendo Wii and Samsung (Gallaugher, 2012). Another group of key partners is Media Content Production companies, such as the television networks and movie production companies. These partnerships are crucial since they give Netflix access to streaming content as well as original content (Gallaugher, 2012). Lastly, as previously mentioned, Netflix system infrastructure builds to a large extent on off-the-shelf technologies, making third-party IT Service Providers key partners to Netflix's business (Netflix, 2016b).

[73] Harris Poll EquiTrend is an annual brand equity study that evaluate the brand equity of more than 1,500 unique brands across 170 categories (PRNewswire, 2014).

Cost Structure

The eighth and final segment in the Business Model Canvas is the cost structure (Osterwalder and Yves, 2010). The cost structure is a reflection of the key partners, activities and resources of Netflix. Before Netflix started to produce its own content, the costs were dominated by multi-year licenses to television networks and other media content producers (Netflix, 2015). These costs are primarily fixed by nature, meaning that they are not tied to member usage or the size of the number base (Netflix, 2015). Since 2011, Netflix are devoting more financial resources towards production of original programming, where they are in a better position to negotiate the terms and conditions of the contracts and thus adding flexibility to the cost structure (Netflix, 2015; Gallaugher, 2012). Another key cost item is the cost for delivering streaming content which includes various software, infrastructure and platform licenses and services, eg. Amazon Web Services (Netflix, 2015). Netflix also states in its annual filing Netflix (2015) that Marketing and payroll (referred to as HR in 4.7) are key cost items in the cost structure. Although Netflix spend around 10% of their revenues on technology and development, the majority of the expenses arise from payroll or software licenses, which are included in *HR Costs* and *Tehnology licenses & services*. Only a minor fraction of Netflix costs are connected to tangible assets, such as hardware or real estate (Netflix, 2015).

Chapter 5

Analysis

THIS chapter presents the analysis of the study. The analysis aims to compare and interpret the empirical results using the theoretical framework. The analysis is divided into three sections connecting to the research questions of the study; *Machine Learning Characteristics*, *Sources of Competitive Advantage* and *Control Mechanisms.*

5.1 Machine Learning Characteristics

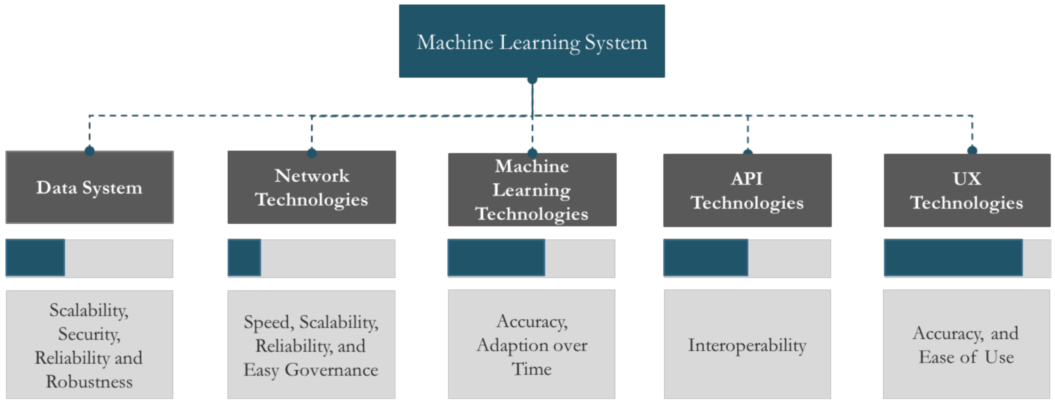
The evolution of Machine Learning teaches that the rapid diffusion and application of Machine Learning Technologies seen today is enabled by advances in other technology fields. The importance of taking a system approach has also been confirmed in interviews.

Although the investigation of MLS demonstrates high interdependency of the different technologies, a deeper analysis also indicates that some technologies are more vital than others. Figure 5.1 gives an overview of the five technology categories identified for MLS. Off-the-shelf solutions are used in essentially all MLS. However, there are also unique (proprietary) technologies within each system. The level of proprietary technology indicates differentiation which in turn indicates strategical importance. An overall evaluation of the level of proprietary technology is illustrated for each technology category in Figure 5.1. Additionally key performances or technical functions are identified and illustrated in the figure. A more comprehensive analysis of the findings for each technology category is presented in this section.

5.1.1 Data System

Data have been concluded to be a vital asset for MLS making the data system an irreplaceable component. According to experts, desired performance metrics are high speed, scalability, reliability, robustness and security. It is also important to allow for flexibility and optimization while maintaining low cost. The relative importance of different performance metrics vary depending on the MLS application. Findings from

Figure 5.1: Machine Learning Characteristics



case studies and interviews indicate high usage of open-source platforms and frameworks rather than unique and proprietary technology. It is rather the combination of data system technologies and its interoperability with the whole system that creates superior performance.

5.1.2 Network Technologies

Although Network Technologies are not required in theory, the trend of growing data sets and more complex algorithms makes Network technologies a practical necessity to meet the requirements on speed, scalability, easy governance, low costs and reliability. Specialized cloud computing solutions and open source framework dominate the market. However, to optimize the capabilities of the Networking Technologies, larger firms like Amazon develops its own networking gear adapted to their specific operational needs. However, Amazon should rather be viewed as an exception as it has developed an entire business around cloud computing. In comparison, Netflix customizes its technologies by developing proprietary solutions on top of existing solutions. Interviews with smaller firms confirm that off-shelf solutions are mainly used for Networking Technologies, the exception being if the system handles sensitive data. To conclude, Network technologies primarily constitute of off-the shelf solutions where layers of proprietary technologies are added for customization of system or to accommodate for specific customer needs and performance requirements.

5.1.3 Machine Learning Technologies

Accuracy and adaptation over time are the most important performance metrics for Machine Learning technologies. The technical challenges are essentially not to enable the program to learn, but rather to determine how to objectify what to learn and how to learn given the different limitations of the data. Another challenge is to compensate for lack of data or noisy data. There are plenty of off-the-shelf solutions, especially for algorithms. Most firms, including Amazon and Netflix, combine off-the-shelf solutions with their proprietary algorithms to customize and tune their models for the specific usage.

The need for proprietary algorithms diminish with access to high quality and high volumes of data. The quality of data is vital, and the type of data needed varies depending on the problem you are trying to solve. Another important factor influencing the performance of MLS is data volume. Different models require different amounts of data to reach desired accuracy. For example, Ditto Labs Inc. uses an Embedded System with unsupervised learning methods, where large amounts of data are necessary to make their learning model accurate. In other cases, access to the right data is more important.

Findings indicate that validation technologies are closely interlinked with the usage of a product or service. There are several validation technologies available off-the-shelf and they may even be integrated in other parts of the MLS. Typically each MLS requires a different set and combination of validation technologies, not only were technical solutions are used. The validation technologies must be customized and adapted to the problem formulation. Likewise, the hypotheses set is dependent on the initial problem where the challenge is to create good instructions giving accurate results. Conventional technologies may be used to realize the instructions and it is often the formulation of instructions combined with the accurate validation technologies that create superior performance.

5.1.4 API Technologies

The empirical results indicate that APIs can be a differentiating factor for MLS. Interoperability is a significant performance measurement for APIs as applications with different features need to work in the same system. Many APIs are available open source where large companies like Amazon, Google, Microsoft and IBM are the main providers. An interesting finding is that some MLS companies have refocused their business around APIs rather than consumer products and services, taking a more upstream position in the value chain.

5.1.5 UX Technologies

UX technologies are emphasised in interviews significant they constitute the interface towards customers. It is not only the design of for example a website that is important but rather what emotions, perceptions and responses you create in relation to your product or service. UI, design such as visualization of data or how you train the customer to use the product or service, are always customized to the specific application. Both Netflix and Amazon have invested extensively in their UI design. A reasonable theory is therefore that UI and the design of product and services play a significant role in differentiating from competitors.

5.2 Sources of Competitive Advantage

Several value drivers for competitive advantage have been identified. Findings from Sections *MLS Characteristics*, *Sources of Competitive Advantage for MLS* and *Case Studies* have been deconstructed into customer utilities which are then connected to the assets and capabilities identified as potential sources of competitive advantage. Due to their nature, some of the resources cannot be directly linked to customer utilities, instead they can be considered to contribute indirectly to all utilities through its impact on other resources. Table 5.1 combines findings from interviews with the result of the case studies of Amazon and Netflix. It should be noted that the list of resources and in Table 5.1 is not exhaustive, but it is rather an attempt to identify the most valuable resources based on the findings from the empirical results. For example, Intellectual Property, and financial assets are resources that often is mentioned in Resource-based View literature. However, since non of the interviewed in this study have mentioned IP or financial assets as valuable resource, these are not included in the list.

5.2.1 Customer utilities and the creation of a superior User Experience

One of the main findings from the interviews is the importance of superior user experience for success. However, superior user experience is not a resource in itself, but instead it can be viewed as the result of management and governance of several different resources. Hence, a first step is to analyze what customer utilities that constitute the core of a superior user experience. This is done by comparing the value propositions of Netflix and Amazon with findings from interviews.

The core of both Amazon's and Netflix's value propositions are that they solve customer problems. Reducing the costs and development efforts for businesses to adopt Machine Learning technology and reducing the cognitive burden of searching for content that you like are customer problems that are solved. Interview results also corroborates the importance of solving a customer problem in creating a superior user experience.

Personalization or customization are key components of both Netflix's and Amazon's value propositions. With vast selections of shopping items or streaming items, only presenting the customers with items that appeal to them has a positive impact on user experience as it reduces the cognitive burden of going through all items and making a decision.

Another finding from interviews is that the cognitive relief depends to what extent the user trusts the product, service or even the company delivering the service. If Netflix or Amazon fail to recommend items the customer likes, or if AWS customers are uncertain if their data is secure in the Amazon cloud, then this will likely stop them from consuming the the product or service.

Easy-to-use or seamless were identified both in interviews and case studies as another customer utility that would add cognitive relief. For example, expressions of easy-to-use in Netflix's value proposition is that its streaming service is not restricted to a particular device.

Table 5.1: Linking resources to value creation

Resources	Impact on Customer Utilities	Customer Utilities
Assets		
Data	Direct	Trust, Personalization
Machine Learning Technologies	Direct	Trust, Personalization, Solves a problem
User Experience Technologies	Direct	Trust, Easy-to-use, Solves a problem
MLS composition	Direct	Trust, Easy-to-use, Solves a problem
Human assets	Indirect	
Capabilities		
Data driven	Indirect	
Innovative over time	Indirect	

5.2.2 Assets

In this section of the analysis, assets are linked to customer utilities to analyze how each asset contributes to value creation. Ultimately the aim is to establish how valuable they are for MLS repectively.

5.2.2.1 Data

Data or data access is considered a key resource for both Amazon's and Netflix's business model. Also, data was often mentioned as one of the first strategically important resources in interviews. Access to data drives success for many reasons. Firstly, there is always a correlation between the accuracy of the machine learning model and the amount and quality of data. If the system fails in performance, for example not recommending a movie you want to see, the customer may not trust the product or service and the general experience of using it will decrease. The amount, type and quality of data that is required depends on the problem to be solved. For example, the core of both Amazon's and Netflix's value propositions is personalized experiences. Although Netflix and Amazon employ macro data in training their models, they would not be able to deliver a personalized experience without access to user data. On the other hand, if the product or service do not use Machine Learning for personalized experiences, i.e. using Machine Learning for teaching a self-driving car to understand different road signs, then the need for real-time user data decrease. Also, different algorithms require different types and amount of data. For example, deep learning require much larger data sets than supervised learning algorithms.

Although data is an asset that is valuable for all MLS, its contribution to competitive advantage depends on how rare it is. For example Netflix and Amazon have free access to unique data sets through users interaction with their web services. In contrast, Ditto Labs product is trained on public data which can be accessed cheaply by all its competitors. Although the argument of rarity holds true for all resources and business fields, the degree of scarcity of the data may be particularly powerful in Machine Learning industry due to the learning trait of the technology. Simplistically, the more unique data set, the more unique learning, and the the better position you are in to create a unique product or service. However, data can also be a risk for the business since it may have a negative effect on trust. Due to the business risk, companies may take technical measures to ensure data security.

5.2.2.2 Machine Learning Technologies

A common denominator between Amazon's and Netflix's Machine Learning Technologies is the usage of multiple algorithms and learning methods. Both companies use ranking models to recommend items to its users and both companies rely on huge data sets that consist of both macro data, such as demographic data, but also behavioral data such as ratings. Although specific details on validation technologies have not been identified, Netflix acknowledge that the processes that have been developed for testing its models are crucial for the performance of the system. The main findings from the interviews supports the findings from the case studies. Namely, the combination of learning methods addressing specific problems generates the highest accuracy. Connecting to utilities, high accuracy is perhaps the most important system performance metric for facilitating trust. Since trust is a key customer utility, superior Machine Learning Technologies will always be valuable. However, since the outcome and accuracy of an algorithm depend to a certain degree on the data it learns from, the importance of Machine Learning technology for competitiveness increases with decreasing rarity of data. In fact, unique Machine Learning technology (excluding data set) becomes extremely powerful if the data is public or constant for all actors on the market. In such scenario, formulating the best hypotheses, having access to the superior composition of algorithms and having the best processes and algorithms for testing the models, will have direct impact on the value and the rarity of the outcome, and meet the attributes of a source of competitive advantage.

5.2.2.3 User Experience Technologies

One of the main findings from interviews was the value of assets relating to the appearance and delivery of the the product or service, the User Experience technologies. Essentially, if the user do not understand how to use product or service, or if the product or service is unappealing or in any other way fails to solve a customer problem, it does not matter how superior the underlying technologies are. The success of Nest demonstrates that appealing and intuitive design may compensate for inferior Machine Learning technology and MLS. For this reason, UX technologies have been broken out and analyzed independently from the system.

The value of UX technologies is supported by that the users seldom notice improvements in system performance but only care about the ease-of-use and what the product or service enables them to do. Equipping the users with the right skills to maximize the perceived value is therefore considered crucial for long-term success. This finding can be supported by that Amazon provides free visualization tools and wizards for its Machine Learning services. Also, Netflix and Amazon have both been recognized for their good customer service. Hence, the assets equipping the users with the skills required are valuable and potential sources of competitive advantage.

Another approach to ensure maximization of value is to make the product or service easy to use through an appealing and intuitive user interface. Both Amazon and Netflix have directed efforts in the UI design of their products and service to make them intuitive and appealing. The motive to this appears to be two-folded. Firstly, since both Amazon and Netflix are dependent on behavioral data for providing personalized content, creating user engagement is vital for the overall performance of the recommendation system. Comparing to interview results, designing a seamless process for training the system was highlighted as crucial asset for creating superior user experience. However, the case studies neither confirm nor dismiss this statement. Also, user engagement for the purpose of training the system is only applicable on Transactional Processing MLS characterized by real-time computing of behavioral data sets. Secondly, if designed properly, UI can be a vehicle for trust. In addition to the technical features, the aesthetic and graphical features of UI, such as logos and colour themes, are brand elements that the firms use to distinguish their product or service from competitors.

Training tools and intuitive UI also contribute in solving a customer problem, because the lack of such assets will essentially add cognitive burden, instead of reducing it.

5.2.2.4 Machine Learning System Composition

Based on interviews and case studies it has become clear that a key success factor is the composition and design of the entire system. Analyzing Amazon's MLS, although employing external technology, it has developed proprietary technologies across all the sub-technology fields. In contrast to Netflix, Amazon have invested a lot of resources into developing and customizing the back-end of the system. Technical solutions related to cyber security can be found both in Amazon's MLS and from interviews. For Amplero, being able to provide secure cloud solutions was a pre-requisite for its customers to engage due to the sensitivity of the data stored and communicated in the system. Comparing Networking Technologies and Data Systems, Netflix focuses less on hardware technology development and has instead focused on architecture design and developing technology tuned for its streaming platform on top of existing technologies. Comparing with the findings from interviews, the back-end technologies are valuable because they provide the infrastructure for the product or service. However, most systems would use commodity hardware and software, and the rarity is created through the design of the architecture and the customization to meet specific customer requirements.

In terms of front-end technologies, Amazon and Netflix are both directing substantial efforts in developing proprietary API technologies. This resonate with the findings from interviews where most of the interviewed believed that knowledge differentiation stemmed from assets in front-end technology fields. In terms of APIs, Amazon and Netflix have different strategies for its technologies. Amazon's APIs is an open platform for machine learning developers and businesses. In contrast, Netflix withdrew public access to its APIs in 2014 and has instead directed substantial internal development efforts in order to optimize accessibility and flexibility for usage of different devices. This difference can be explained by that Amazon has turned its API into a business in itself, generating a steady stream of valuable data.

5.2.2.5 Human Assets

Many of the key activities and key resources of Amazon and Netflix are the result of humans applying their knowledge to already existing knowledge. For example, both Amazon and Netflix have R&D as a key activity, resulting in the creation of key resources like proprietary technology and IT infrastructure. The fact that one of the major costs of both Netflix's and Amazon's cost structures is HR and/or payroll affirms the linkage between human assets and other key resources. Although it is the capabilities of humans rather then the asset itself that translates into customer value, having access to the right people is one of the main findings of the interviews. Partnering with leading universities, participating in conferences, arranging competitions (i.e. Netflix Prize) and open-source initiatives are all activities that companies engage in in order to attract and recruit talents. In particular, prominent Machine Learning researchers and practitioners appears to be a source to competitive advantage. Partly because there is a greater demand than supply, but also because they have the knowledge and skills required to build on top of commodity solutions, and thus creating Machine Learning technologies that are both valuable and rare. Although the findings from interviews indicate that the right Machine Learning minds are the most valuable and rare category of human assets, there is also a need for entrepreneurs that can identify customer needs and transform them into products and services. User Interface experts is another category of human assets that has been singled out as important for success. Since an expert by definition posses knowledge and skills that others do not, the knowledge they bring into the firm will be rare. Although human assets cannot directly be linked to any customer utilities, the linkage between human assets and other key assets indicates that their value stems from enabling the creation of other valuable and rare assets.

5.2.3 Capabilities

In this section of the analysis, capabilities are linked to customer utilities to analyze how each capability contributes to value creation and ultimately determine their value and role in competitiveness.

5.2.3.1 Data driven

To use data insights for strategic and business decisions is identified as a key success factor in the interviews. This is corroborated by the the history and key activities of Amazon and Netflix which repeatedly show how business decisions are driven by data insights. Hence, to have a business strategy that is anchored in data insights enables better management of current resources as well as identification of opportunities to create value in the future. Based on the findings from interviews and case studies, there is support for that being data driven is a valuable capability. However, the findings from empirical results do not spread light on the rarity of the capability and therefore it can neither be determined nor dismissed as a source of competitive advantage.

5.2.3.2 Innovative over time

The case studies illustrate that both Amazon and Netflix at several occasions have demonstrated their ability to be innovative. Netflix organized a public competitions to improve its recommendation system or to develop APIs for device independence. The launching of AWS is in itself a testimonial to Amazon's ability to be innovative. In terms of knowledge acquisition, Amazon continuously acquires know-how and technology to implement in its business, such as the robots from Kiva Systems that have significantly improved Amazons fulfillment centres. Relating to both case studies and findings in interviews it can thus be stated that a firm's ability to be innovative over time is valuable and provided that its competitor is less innovative over time, a source of competitive advantage.

5.2.4 Other Considerations

The connection between customer utilities and capabilities are less direct compared to assets since capabilities by definition are resources that are applied on assets. With that being said, the weaker connection to customer utilities does not mean that they are not sources of competitive advantage. Not all success factors, resources or utilities found in the empirical results have been analyzed in this section. The reason for that being that there has not been sufficient empirical support. For example, the ability to collaborate cross-functionally was stated as a valuable capability in one interview, however, this statement could not be corroborated from other interviews or the case studies.

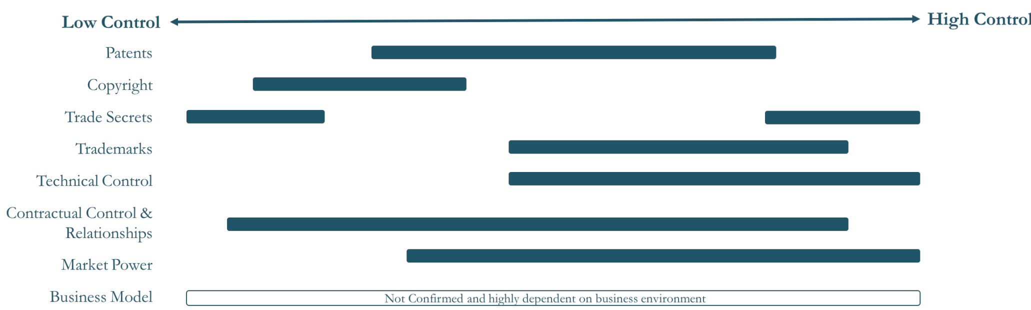
5.3 Control Mechanisms

The analysis of control mechanisms is divided into two sections. Firstly, the possibility of competitive duplication is assessed by ranking the level of control generated by each mechanism. Secondly, an analysis is presented of how the different control mechanisms are used for the identified potential sources of competitive advantage in MLS.

5.3.1 Level of Control

The level of control generated by a control mechanism depends on the context of which it is used. It is therefore important to understand the range of control for each mechanism in MLS. The ranges are illustrated in Figure 5.2, and are based on the advantages and disadvantages described in *Control Mechanisms* and specifically in the context of MLS.

Figure 5.2: Level of Control for the Control Mechanisms identified



The strength of a patent protection is determined by the scope of the patent. Many MLS inventions are on the boarder-line of patentable subject matter, which limits the effectiveness of patents for control. Furthermore, it is important to be able to enforce your right which is difficult for many patents within MLS due to the hardship of proving infringement. The effectiveness of patents for protecting MLS is assessed to medium depending on the invention and the intended use.

Although copyright is easy to obtain, it is not considered a strong mechanism for MLS related work due to the narrow scope of the right. Hence, the range is rated to be fairly low in 5.2.

Trade secrets i.e. secrecy can indeed generate high level of control. If you can prevent MLS technology or know-how to be spread, it will exclusively belong to you. However, if the knowledge is leaked the damage is complete and your protection is fully exhausted. Consequently, using trade secrets as control mechanisms for MLS implies either very high control or very low control.

Trademarks provide exclusive protection, implying high level of control for MLS. However, the right is not only obtained by registration, but determined by field of use. The strength of the brand also has high impact on the scope of protection, as they are closely connected. Hence, the level of control using trademarks for MLS varies from medium to high depending on the strength of the brand in the field of use.

Technical control through the usage of specifically designed software and hardware may generate very effective protection of intellectual assets in MLS. However, as technology in MLS changes rapidly, the control granted from a technical protection service is constantly challenged. Also, comparing to IPRs, technical protection only provides exclusivity over a given asset, and does not hinders competitors from developing and using similar assets themselves. Technical control for MLS in the form of customer lock-in can also add to the strength, provided that the lock-in effect does not compromise the perceived value of the product or service.

Similarly to IPRs, contracts generate rights that are enforceable in a court of law and can therefore provide high level control for MLS. Also, the terms of the contract can be designed to meet specific needs, which enable customization and bridge weaknesses of other control mechanisms used to control the MLS. The strength of control generated from contracts are often dependent on the strength of the relationship between the contracting parties. Contracts that are well-designed and where the relationship of the contracting parties is strong can generate high level of control. On the other hand, contracts that are poorly designed and where contracting parties have low incentives to comply may implicate relatively low level of control for MLS.

The control generated from market power has been described in interviews from a reputation or brand perspective as a common manifestation of market control for MLS. Furthermore, trust has been identified as essential for creating a superior user experience in MLS and the core value provided by brand is trust. The control generated from financial strength has not emerged during interviews, which may be an indication of low control effectiveness. The concept of first mover advantage can be effective for MLS in many situations. Experience from solving a particular customer/technical problem, securing exclusive access to limited data or switching costs are all reasonably effective to a certain extent for MLS. However, reputation effects or first mover advantage do not generate rights that are enforceable in court and require constant management to retain.

Although business models have been identified as a means of control, the findings from empirical results do not provide enough insight to analyze the level of control of a business model compared to other control mechanisms used for MLS. Therefore, the analysis related to business model will be restricted to acknowledging that business models can be a source of sustained competitive advantage, and be used as a mechanism of control, particularly when other control mechanisms fall short.

5.3.2 Control of Sources of Competitive Advantage

The underlying assumption of this study is that the most important elements to control are the most valuable and rare resources. Hence, this analysis presents how control mechanisms and combinations thereof are applied on the potential sources of competitive advantage identified in *Sources of Competitive Advantage*. Table 5.2 presents an overview over the resources and what control mechanisms are used in MLS, what potentially could be used and what is not applicable.

The analysis of control mechanisms is primarily based on interviews with experts within the field (see Section *Control Mechanisms*). Identifying how control mechanisms are utilized has been challenging as this information is usually kept secret. Thus, the case studies are not included in this analysis as the information would not be sufficiently comprehensive.

Table 5.2: Mechanisms used for controlling potential sources of competitive Advantage

Potential Sources of CA	Control Mechanisms							
	Patents	Copyright	Trade Secrets	Trademark	Technical Control	Contractual Control & Relationships	Market Power	Business Model
Data	N/A							
Machine Learning Technologies				N/A				
UX Technologies								
MLS Composition								
Human Assets	N/A	N/A	N/A	N/A	N/A			
Capabilities	N/A	N/A		N/A	N/A			

Control Mechanism used for MLS

N/A Not Applicable

Possible Control Mechanism

5.3.2.1 Data

Data or the access to data is controlled indirectly through the use of copyright on databases, in which data is structured to optimize access. Technical solutions for collecting the data may be fulfil the criteria of patentable subject matter, however these types of patents have not been emphasized in any of the interviews as a mechanism that is employed. Instead, data is mainly controlled technically through the use of specially designed software, restricting the access to data from both external threats, such as competitors and users, as well internal threats such as employees. The data is also controlled technically through the use of specifically designed hardware, such as server and network devices. A third layer of control of data that is commonly used is contractual control to the access of data. Through contractual agreements, the firm may secure rights to store and use data that is required to train and run its Machine Learning models. Depending on the data, contracting parties may be its customers but also other external entities that the firm has partnered with. In addition to be well-designed and build on strong relationships, the terms of the contracts must be exclusive in the field of use to meet the requisites of a source of sustained competitive advantage. If data is scarce, first mover advantage can be assumed to yield some level of control. As a pioneer, you do not have to compete for access to data in the initial stage and you are therefore in a better position to establish control, e.g. through relationships and contracts. Although there are inherent weaknesses with technical, copyright, contractual control and first mover advantage, the combination of control mechanisms provides an overall strong control position.

5.3.2.2 Machine Learning Technologies

As the control of data sets have already been accounted for, this section will analyze the control of hypotheses sets, algorithms and validation technologies. Although algorithms and validation technologies may be considered patentable subject matter, the findings from the interviews indicate that patents seldom are used as means of control. The reasons for this are mainly that detecting infringement is difficult, the patents are generally easy to invent-around and the enforceability in court is low. However, patenting of Machine Learning technologies do occur but mostly for larger companies that have the financial strength to withstand a law suit. Smaller companies have also been found to apply for patents for their Machine Learning technologies, but primarily as a mean to attract investors rather than as a mean of control. Instead, hypotheses sets, algorithms and validation technologies are mainly controlled through a combination of secrecy and technical control. Secrecy, through the usage of Non-Disclosure-Agreements and trade secrets is employed across all assets, but perhaps particularly controlling the know-how of what questions to ask and how hypotheses can be validated. Technical control, through the use of private cloud solutions and other technical solutions restricting access to the learning and testing algorithms are also employed. Additionally, the intrinsic properties of Machine Learning models themselves can be viewed as a technical barrier as even the inventors themselves cannot explain the outcome. This ”black-box” phenomena makes reverse-engineering almost impossible. Although not confirmed in interviews, the first mover advantage can generate control as the model will improve with learning over time. Although the programming code constituting the algorithms may be subject to copyright, there have been no court cases of copyright infringement and the findings from interviews indicates that copyright is not considered effective for protecting the crude code. Overall, technical control mechanisms and secrecy enables effective control over Machine Learning assets.

5.3.2.3 UX Technologies

Similarly to data and Machine Learning technologies, several of control mechanism are employed to control UX Technologies. Since many of assets are visible or publicly available, secrecy is not an option and the overall risk of imitation is high. On the other hand, the visibility also entails that infringement is more easily detected. Starting with the control mechanisms generating legal rights, trademarks and possibly design patents are used to create control. Trademarks can generate strong control over product and service names, as well as elements of websites. However, the level of control achieved by trademarks is highly dependent on the strength of the brand and less effective in itself. Design patents have been found to be used for controlling user interface features, however whether these patents generate strong control have not been unambiguously confirmed. Instructions, visualization elements and software are mainly protected by copyrights, which offers a very narrow scope of protection. Since the users interacts directly with many of the assets, these assets can be assumed to constitute the big part of the brand. Hence, lock-in effects rising from strong brand loyalty, brand image or top-of-mind-logic can be very effective to retain customers, but it will not hinder imitation. Conclusively, a comprehensive and strong control position is difficult to achieve for UX technologies.

5.3.2.4 Machine Learning System Composition

The composition of different technologies have been identified as one of the most valuable assets and this appears to be controlled through secrecy. Although some some assets may be public, the know-how of how the different technologies are combined are usually protected by trade secrets. Technical control restricting users and employees from accessing all parts of the system is also used. Furthermore, since many MLS are characterized by off-the-shelf technologies or open-source software for Data System, Network Technologies and APIs, license agreements becomes an important mechanism to ensure access to valuable technology. License agreements and contracts regulating terms of use are also employed to control usage of the product or service. Patents are used for protecting both software and hardware across all sub-technologies. However, the level of control yielded is generally higher for hardware patents as infringement is more easily detected. For the same reason, patents covering back-end technology are considered less effective. Another type of technical control that can be effective is opening up front-end technologies, such as APIs for the purpose of creating a standard which then drives usage of underlying infrastructure or to generate data. Based on the findings and analysis of MLS characteristics, it is reasonable to assume that the complexity of the system in itself creates substantial barriers to duplication. On the other hand, a too complex system may generate managerial and execution issues. Conclusively, the composition of MLS can be highly controlled, but at the expense of efficiency.

5.3.2.5 Human Assets

Human assets are mainly controlled by contracts and market power. Although contracts have the legal capacity to regulate unlawful use of know-how, it cannot prevent valuable employees from leaving neither can contracts control the access to them in the first place. Hence, contracts yield relatively weak level of control over human assets. Instead, reputation effects and brand appear to be much more effective in terms of attracting talents. However, the increased turnover of employees in ICT indicates that human assets are increasingly hard to control.

5.3.2.6 Capabilities

The ability of a firm to be data driven and innovative over time cannot be linked directly to any of the control mechanisms based on the findings from interviews. However, that does not necessarily mean that they cannot be sources of sustained competitive advantage. Firm capabilities are ultimately developed and governed by the overall model of the business, and a reasonable conclusion is therefore that the business model can be a mechanism of control. It may be that the challenges associated with developing and remaining data driven and innovative over time create barriers for competitive imitation. Additionally, the ability of being data driven and innovative over time is dependent on the control of other assets, such as data, technology for analyzing data and human assets.

Chapter 6

Conclusion

The main customer value created by Machine Learning System (MLS) is cognitive relief where the burden of decision-making or effort is reduced. Thus, creating superior user experience is a pre-requisite for competing. In order for the customer to experience a cognitive relief, the product or service must be trustworthy, perceived as personalized, easy-to-use and solve an actual problem. Failure of delivering these utilities will add cognitive burden and impair the user experience. Data, Machine Learning technologies, User Experience (UX) technologies, the composition of MLS, human assets, the ability to be data driven and the ability to be innovative over time are the most important resources for generating these utilities, and hence the most valuable resources for a firm to possess.

To be sources of sustained competitive advantage the resources must also be rare and difficult for competitors to imitate or substitute. Although human assets are rare by virtue, they are difficult to control. The combined usage of IPRs, secrecy, technical, and contractual control form a strong protection against duplication of the MLS composition, however, off-the-shelf and open source technologies are commonly employed compromising rarity. UX technologies are pre-requisite for success since these are the resources that ultimately communicates the customer utilities. In relation to other technologies in the system, these are characterized by the highest degree of proprietorship and reasonably unique to some extent. However, UX technologies are difficult to control and therefore more likely a source of competitive advantage, rather than *sustained* competitive advantage.

Although not explicitly confirmed in interviews, it is reasonable to assume that the difficulty for a firm to develop and maintain capabilities such as innovative over time and data driven automatically creates a certain level of rarity and barriers for duplication. Data and ML technologies are assets for which a high level of control can be achieved, primarily by technical, contractual and secrecy means. Provided rarity, data and ML technologies are concluded to be strong candidates to sources of sustained competitive advantage.

Figure 6.1: The valuable resources identified in this study mapped against potential rarity and control.

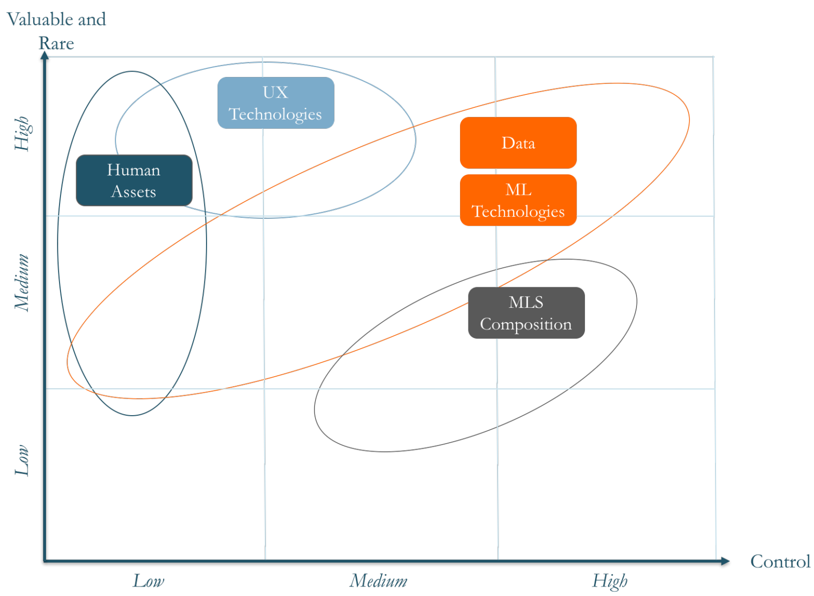


Figure 6.1 maps the most valuable resources and how the level of rarity and control affects a resource's potential to be a source of sustained competitive advantage. The resources in the top right corner meets the criteria of *sustained* competitive advantage, whereas resources in the top left corner likely only are sources of competitive advantage. Because the accuracy and performance of Machine Learning technologies depends on the data it learns from, two main scenarios of competition have been identified;

1. data is rare
2. data is public

In the first scenario, the firm has exclusive access to data sets resulting in unique outcome that is impossible to duplicate despite the usage of off-the-shelf algorithms or low level of control over technologies. By reducing the level of control and using a higher degree of commodity technology, the firm will likely be able to reduce costs without compromising revenues and potentially increase profitability. In the second scenario, data cannot be a source to sustained competitive advantage since it is not rare. In such scenario, superior accuracy depends exclusively on the hypotheses sets, algorithms and validation technologies. Thus, the importance to control these assets increase, and this will be the starting point for sustained competitive advantage. In this scenario, human assets will likely play a more important role for success and efforts should be taken to increase the control over these assets. In this case, costs associated with controlling data can be reduced. The dynamic act of optimizing the level of control and rarity is illustrated by the oval shapes in 6.1, starting from the upper corners and moving down towards the bottom left corner.

Finally, designing sustained competitive advantage for MLS requires an understanding of the interdependencies of the system and how different sub-technologies are affected by what problems are solved and what data sets are required. The interdependencies of the MLS suggests a dynamic approach rather than a static, which needs to be reflected in the business model. The business model should be dynamic and creative to mitigate or exploit control weaknesses. It also needs to employ multiple control mechanisms to not only prevent competitive duplication, but also enables access and development of valuable and rare resources. In fact, such dynamic model can in itself be both an effective control mechanism and source of sustained competitive advantage, provided that it is unique and difficult to replicate.

Chapter 7

Discussion

The technology field investigated in this study is evolving rapidly, resulting in a constant change of market dynamics. The adoption of Machine Learning in new industries and for new applications fuels the technology advancement, where new innovations and ideas constantly are added which will influence future society.

The major takeaway is the impact and role of data. Although technologies develop, the basic ingredient for all Artificial Intelligence is data. Data might be the most indispensable resource for companies in future where not only access but also management of data will be crucial. The ability to make sense of data, both externally and internally will be crucial across all business fields as enabling technologies mature. Creating trust by developing and promoting cyber secure solutions is increasingly important. We predict that data privacy, cyber security and ultimately trust will play a more and more significant role for success as the usage of Machine Learning, and AI in general, increase.

Suggestions for Further Research

In order to fully determine sustainable competitive advantage, the competitive landscape should be investigated to prove the attributes of rarity and non-substitution. Such research could be further extended to investigate different business models and their impact on sustained competitive advantage.

A more comprehensive sample could be selected to create a more extensive and general study. Also, a narrower scope focusing on i.e. MLS start-ups or a specific application field of MLS could allow for more in-depth insights. On the other hand, such restriction would compromise generalization. Contrl mechanisms for ICT industry is another topic that can be further researched. Although the abductive approached allowed for additional control mechanism to be discovered, there may be other mechanisms of control that play a role designing sustained competitive advantage.

Bibliography

A. A. Alchian. The Concise Encyclopedia of Economics, Property Rights, 2016. URL http://www.econlib.org/library/Enc/PropertyRights.html.

J. Allison, A. Dunn, and R. Mann. Software Patents, Incumbents and Entry. *Texas Law Review*, 85:1579–1625, June 2007. URL http://ssrn.com/index.html.

E. Alpaydin. *Introduction to Machine Learning, third edition.* The MIT Press, Cambridge, Massachusetts, London, England, 2014.

X. Amatriain. Big and Personal: data and models behind Netflix recommendations. In *The 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, Chicago, Illinios, USA, Aug. 2013. URL https://pdfs.semanticscholar.org/4b89/dfc8782bf0fc403b5b43e5e55ea1d4dc44c6.pdf.

X. Amatriain. 10 Lessons Learned from Building Machine Learning Systems, 2014. URL http://www.slideshare.net/xamat/10-lessons-learned-from-building-machine-learning-systems.

Amazon. Amazon.com Launches Web Services, 2002. URL http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=503034.

Amazon. *Amazon.com Inc, Form 10-K(Annual Report).* Amazon, 410 Terry Avenue North, Seattle, Washington 98109-5210, United States of America, 2014.

Amazon. Amazon History and Culture, 2016. URL http://www.amazonfulfillmentcareers.com/about-amazon/history-and-culture/.

AmazonWebServices. Amazon Machine Learning, 2016. URL http://docs.aws.amazon.com/machine-learning/latest/dg/types-of-ml-models.html.

AmazonWebServices. Amazon EMR, 2016a. URL https://aws.amazon.com/elasticmapreduce/.

AmazonWebServices. Using Public Data Sets, 2016b. URL `http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-public-data-sets.html`.

AmazonWebServices. What is Amazon Machine Learning?, 2016c. URL `http://docs.aws.amazon.com/machine-learning/latest/dg/what-is-amazon-machine-learning.html`.

AmazonWebServices. AWS Partner Network, 2016d. URL `https://aws.amazon.com/partners/`.

S. Ambati. Machine Learning: A practical introduction, 2015. URL `http://www.infoworld.com/article/3010401/big-data/machine-learning-a-practical-introduction.html`.

ApacheHadoop. What is Apache Hadoop?, 2016. URL `http://hadoop.apache.org/`.

Apple Inc. US7469381B2 - List scrolling and document translation, scaling, and rotation on a touch-screen display, 2007. URL `https://patents.google.com/patent/US7469381B2/en`.

Authorize.Net LCC. Merchant Web Services API. In *Customer Information Manager (CIM) SOAP Guide*. Authorize.Net LCC, 2015. URL `http://www.authorize.net/content/dam/authorize/documents/CIM_SOAP_guide.pdf`.

P. Ball. The truth about the Turning Test, 2015. URL `http://www.bbc.com/future/story/20150724-the-problem-with-the-turing-test`.

J. Barney. Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17:99–120, mar 1991. doi: 10.1177/014920639101700108. URL `https://business.illinois.edu/josephm/BA545_Fall%202011/S10/Barney%20(1991).pdf`.

J. Basilico. Lessons Learned from Building Machine Learning Software at Netflix. In *Talk from Software Engineering for Machine Learning Workshop(SW4ML)*, Montreal, Canada, Dec. 2014. Neural Information Processing Systems (NIPS). URL `http://www.slideshare.net/justinbasilico/lessons-learned-from-building-machine-learning-software-at-netflix`.

G. Bensinger. Amazon Hints at New Shipping Business, 2016. URL `http://blogs.wsj.com/digits/2016/01/29/amazon-hints-at-new-shipping-business/`.

D. Beyer. On the evolution of machine learning, From linear models to neural networks: an interview with Reza Zadeh, 2015. URL `http://radar.oreilly.com/2015/05/on-the-evolution-of-machine-learning.html`.

T. Bishop. Jeff Bezos explains why Amazon dosen't really care about its competitors, 2013. URL `http://www.geekwire.com/2013/interview-jeff-bezos-explains-amazon-focus-competitors/`.

A. Blum and P. Langley. Selection of relevant features and examples in machine learning. In *Artificial Intelligence*, volume 97. Elsevier, Dec. 1997. URL http://www.sciencedirect.com/science/article/pii/S0004370297000635.

D. Bohn. Netflix kills third party app platform, current apps will continue to work, 2013. URL http://www.theverge.com/2013/3/8/4081336/netflix-kills-third-party-app-platform-current-apps-will-continue-to.

J. Brodkin. Netflix's many-pronged plan to eliminate video playback problems, 2014. URL http://arstechnica.com/information-technology/2014/05/netflixs-many-pronged-plan-to-eliminate-video-playback-problems/.

J. Brownlee. A Tour of Machine Learning Algorithms, 2013. URL http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/.

J. Brownlee. Discover Feature Engineering, How to Engineer Features and How to Get Good at It, 2014. URL machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/.

J. Brownlee. How to Prepare Data For Machine Learning, 2015. URL http://machinelearningmastery.com/how-to-prepare-data-for-machine-learning/.

A. Bryman. *Social Research Methods*. Oxford University Press, 198 Madison Avenue, New York, NY 10016, United States of America, 2016.

A. Bryman and E. Bell. *Business Research Method*. Oxford University Press, 198 Madison Avenue, New York, NY 10016, United States of America, 2015.

I. Burrington. Why Amazon's Data Centres Are Hidden in Spy Country, 2016. URL http://www.theatlantic.com/technology/archive/2016/01/amazon-web-services-data-center/423147/.

R. Bushey. How the $400 Million loan Netflix just took out will help it undercut Hollywood, 2014. URL http://www.businessinsider.com/why-netflix-makes-its-own-shows-2014-2.

J. Carbonell, R. Michalski, and T. Mitchell. *Machine Learning: An Artificial Intelligence Approach*. Springer-Verlag, Berlin Heidelberg GmbH, 1984.

D. Chaffey. Amazon's business stretegy and revenue model: A history and 2014 update, 2014. URL http://www.smartinsights.com/digital-marketing-strategy/online-business-revenue-models/amazon-case-study/.

S. Chaudhuri. What next? a half-dozen data management research goals for big data and the cloud. In M. Krötzsch, editor, *Proceedings of the 31st symposium on Principles of Database Systems*, pages 1–4, Scottsdale, AZ, USA, 2012. ACM.

B. Christensen. Optimizing the Netflix API, 2013. URL http://techblog.netflix.com/2013/01/optimizing-netflix-api.html.

C. Christensen. *The Innovator's Dilemma: The Revolutionary Book That Will Change the Way You Do Business*. Harvard Business Press, Cambridge, MA, USA, 1997.

J. Clark. How Amazon exposed its guts: The History of AWS's EC2, 2012. URL http://www.zdnet.com/article/how-amazon-exposed-its-guts-the-history-of-awss-ec2/.

L. Clark. How Amazon Web Services Uses Linux and Open Source, 2014. URL https://www.linux.com/blog/how-amazon-web-services-uses-linux-and-open-source.

ClubCloudComputing. Business Model Canvas for Cloud Provicers, 2016. URL http://www.clubcloudcomputing.com/2014/08/business-model-canvas-cloud-providers/.

I. Cockburn and M. MacGarvie. Entry and Patenting in the Software Industry. *Management Science*, 57:915–933, Apr. 2011. doi: org/10.1287/mnsc.1110.1321.

Committee on Intellectual Property Rights in the Emerging Information Infrastructure. Protecting Digital Intellectual Property: Means and Measurements. In *The Digital Dilemma: Intellectual Property In the Information Age*, pages 152–198. National Academy Press, 2000. doi: 0.17226/9601. URL http://www.nap.edu/read/9601/chapter/7.

ComputerNetworkingNotes.com. Network Definitions Characteristics Components locations, 2016. URL http://computernetworkingnotes.com/network-security-access-lists-standards-and-extended/network.html.

C. I. Cook. Netflix: A Stepping Stone in the Evolution of Television, 2014. URL http://dspace.nelson.usf.edu/xmlui/bitstream/handle/10806/11681/Cook_ARP.pdf?sequence=1.

M. Cusumano. Technology strategy and management: The evolution of platform thinking. *Communications of the AMC*, 53:32–34, Jan. 2010. URL http://ebusiness.mit.edu/research/papers/2010.01_Cusumano_Technology%20Strategy%20and%20Management_272.pdf.

L. Dambrosio. Machine Learning as a Service: How Data Science Is Hitting the Masses, 2016. URL http://www.huffingtonpost.com/laura-dambrosio/machine-learning-as-a-ser_b_9548962.html.

R. D'Aveni, G. Battista Dagnino, and K. Smith. The age of temporary advantage. *Strategic Management Journal*, 31:1371–1385, Dec. 2010. doi: 10.1002/smj.897.

C. Davies. An evolutionary step in Intellectual Property Rights: Artificial Intelligence and Intellectual Property. *Computer Law & Security Review*, 27:601–619, Dec. 2011. doi: 10.1016/j.clsr.2011.09.006.

J. Dean and R. Monga. Tensorflow:Google's latest machine learning system, open sourced for everyone, 2015. URL `http://googleresearch.blogspot.com/2015/11/tensorflow-googles-latest-machine_9.html`.

S. Dimitrios. Why neurial networks?, 2016. URL `http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/ds12/article1.html`.

A. Dubois and L. E. Gadde. Systematic combining: and abductive approach to case research. *Journal of Business Research*, 55(7):553–560, July 2002. doi: 10.1016/S0148-2963(00)00195-8.

L. Essers. Apple's crucial overscroll bounce patent claim is valid, U.S. patent office says, 2014. URL `http://www.macworld.com/article/2042023/apples-crucial-overscroll-bounce-patent-claim-is-valid-us-patent-office-says.html`.

M. Feilzer. Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. *Journal of Mixed Methods Research*, 4(1):6–16, Jan. 2010. doi: 10.1177/1558689809349691. URL `http://jmmr.sagepub.com`.

J. Gallaugher. *Getting the most out of Information Systems, v.2.0.* Lardbucket.org, 2012. URL `http://2012books.lardbucket.org/books/getting-the-most-out-of-information-systems-v2.0/`.

R. Goodwins. The future of storage:2015 and beyond, 2015. URL `http://www.zdnet.com/article/the-future-of-storage-2015-and-beyond/`.

R. M. Grant. Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal*, 17:109–122, Winter Special Issue 1996. doi: DOI:10.1002/smj.4250171110.

T. D. Greene and J. Wilkerson. Understanding trademark strength. *Stanford Technology Law Review*, 16, 2013. URL `https://cyberlaw.stanford.edu/files/publication/files/understandingtrademarkstrength.pdf`.

T. Grundy. The source of Amazons competitive advantage, 2015. URL `http://www.accaglobal.com/us/en/discover/cpd-articles/business-management/amazon-flow.html`.

M. Halac. Relational Contracts and the Value of Relationships. *The Americal Economic Review*, 102:750–779, Apr. 2012. doi: http://dx.doi.org/10.1257/aer.102.2.750.

L. Hallenborg, M. Ceccagnoli, and M. Clendenin. Intellectual Property Protection in the Global Economy. In G. D. Libecap and M. Thursby, editors, *Advances in the Study of Entrepreneurship, Innovation and Economic Growth, Volume 18 : Technological Innovation : Generating Economic Results*, pages 63–116. Emerald Group Publishing Ltd, 2008.

J. Hamilton. The Stunning Scale of AWS And What It Means For The Future Of The Cloud, 2015. URL http://highscalability.com/blog/2015/1/12/the-stunning-scale-of-aws-and-what-it-means-for-the-future-o.html.

N. Hemsoth. Why the golden age of machine learning is just beginning, 2015. URL http://www.nextplatform.com/2015/10/20/why-the-golden-age-of-machine-learning-begins-now/.

S. Higginbotham. Amazon launches a $100 million fund to support the AI brain behind its Echo device, 2015. URL http://fortune.com/2015/06/25/amazon-alexa-fund/.

R. Ho. A Introduction to 6 Machine Learning Models, 2012. URL hhttps://dzone.com/articles/introduction-6-machine.

G. Hooley, A. Broderick, and K. Möller. The Resource Based View of the Firm. *Journal of Strategic Marketing*, 6:97–116, 1998. doi: 10.1080/09652549800000003.

E. Horvitz and T. Mitchell. From Data to Knowledge to Action: A Global Enabler for teh 21st Century, 2010. URL http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/From_Data_to_Knowledge_to_Action.pdf.

K. Huang, D. Dyerson, L. Wu, and G. Harindranath. From Temporary Competitive Advantage to Sustainable Competitive Advantage. *Brithsh Journal of Management*, 26:671–636, 2015. doi: 10.1111/1467-8551.12104.

N. Ingraham. The redesigned Netflix website is rolling out to users starting today, 2015. URL http://www.theverge.com/2015/6/15/8783761/netflix-web-interface-redesign-launch.

International Organization for Standardization (ISO). ISO 9241-210:2010(en): Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems, 2010. URL https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en.

D. Jacobson. Embracing the Differences : Inside the Netflix API Redesign, 2012. URL http://techblog.netflix.com/2012/07/embracing-differences-inside-netflix.html.

D. Jacobson. Top 10 Lessons Learned from the Netflix API - OSCON 2014, 2014. URL http://www.slideshare.net/danieljacobson/top-10-lessons-learned-from-the-netflix-api-oscon-2014.

S. Jones. How the GPU Is Revolutionizing Machine Learning, 2015. URL https://blogs.nvidia.com/blog/2015/12/10/machine-learning-gpu-facebook/.

M. Kassner. A look at Amazons world-class data-centre ecosystem, 2014. URL http://www.techrepublic.com/article/a-look-at-amazons-world-class-data-center-ecosystem/.

F. Keller. Connectionist and Statistical Language Processing. In *Evaluation*. Universität des Saarlandes, Apr. 2016. URL http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/lecture11_4up.pdf.

P. Kieun. Database Technology for Large Scale Data, 2011. URL http://www.cubrid.org/blog/dev-platform/database-technology-for-large-scale-data/.

J. Konstan and J. Riedl. Deconstructing Recommender Systems, How Amazon and Netflix predict your preferences and prod you to purchase, 2012. URL http://spectrum.ieee.org/computing/software/deconstructing-recommender-systems.

A. Kosner. Why is Machine Learning (CS 229) The Most Popular Course At Stanford?, 2013. URL http://www.forbes.com/sites/anthonykosner/2013/12/29/why-is-machine-learning-cs-229-the-most-popular-course-at-stanford/#273543c861ba.

S. Kotsiantis. Supervised Machine Learning: A Review of Classifications Techniques. In I. Maglogiannis, editor, *Emerging Artificial Intelligence Applications in Computer Engineering*, Nieuwe Hemweg 6B, 1013 BG Amsterdam, Netherlands, 2007. IOS Press.

G. Kovacs. An analysis of strategies by Netflix in the television market, 2015. URL http://pure.au.dk/portal-asb-student/files/86448002/Thesis_GaborKovacs_201208049.pdf.

Kraaijenbrink. Three shortcomings of the Business Model Canvas, 2012. URL http://kraaijenbrink.com/2012/07/shortcomings-of-the-business-model-canvas/.

S. Kuenzel and S. V. Halliday. The chain of effects from reputation and brand personality congruence to brand loyalty: The role of brand identification. *Journal of Targeting, Measurement and Analysis for Marketing*, 18:167–176, Nov. 2010. doi: 10.1057/jt.2010.15. URL http://www.palgrave-journals.com/jt/journal/v18/n3/full/jt201015a.html.

S. Kujala, V. Roto, K. Väänänen-Vainio-Mattila, E. Karpanos, and A. Sinnelä. UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, 23:473–483, July 2011. doi: 10.1016/j.intcom.2011.06.005. URL http://iwc.oxfordjournals.org/content/23/5/473.full.pdf+html.

A. Kumar. Machine Learning : Few rarely shared trade secrets, 2016. URL https://www.linkedin.com/pulse/machine-learning-few-rarely-shared-trade-secrets-ashish-kumar.

M. Kuniavsky. When the Cloud decides: Designing for Predictive Machine Learning for the IoT. In *O'Reilly Design Conference*, Jan. 2016. URL https://www.scribd.com/doc/296215323/When-the-cloud-decides-Designing-for-predictive-machine-learning-for-the-IoT-O-Reilly-Design-2016.

M. Kurt, S. Grabner-Kräuter, and S. Bidmon. *The Journal of Product and Brand Management*. doi: http://dx.doi.org/10.1108/10610420810875070.

E. L.-C. Law, V. Roto, M. Hasssenzahl, A. P. Vermeeren, and J. Kort. Understanding, scoping and defining user experience: A survey approach. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, Boston, MA, USA, 2009. CHI. URL `https://www.researchgate.net/publication/221518375_Understanding_scoping_and_defining_user_experience_A_survey_approach`.

J. Layton. How Amazon Works, 2005. URL `http://money.howstuffworks.com/amazon1.htm`.

Li Deng and Dong Yu. *Deep Learning*. NOW the essence of knowledge, 2014. doi: 10.1561/2000000039. URL `http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf`.

M. B. Lieberman and D. B. Montgomery. First-Mover Advantages. *Strategic Management Journal*, 9(Special Issue):41–58, 1988. URL `http://www.uni-oldenburg.de/fileadmin/user_upload/wire/fachgebiete/entrepreneur/download/Artikel_Internetoekonomie/Lieberman_First_Mover.pdf`.

G. Linden, B. Smith, and J. York. Amazon.com recommendations, item-to-item collaborative filtering. *IEEE Computer Society*, pages 76–80, Feb. 2003.

S. Lindmark. *Techno-economic analysis - an introduction*. Center for Intellectual Property studies, Chalmers University of Technology, Gothenburg, Sweden, 2006.

A. Linn. Microsoft releases CNTK, its open source deep learning toolkit, on GitHub, 2016. URL `http://blogs.microsoft.com/next/2016/01/25/microsoft-releases-cntk-its-open-source-deep-learning-toolkit-on-github/#sm.00006aswzu6i4eswtvu1zpyeg3e42`.

J. Lowenshohn. The science behind Netflix's first major redesign in four years, 2015. URL `http://www.theverge.com/2015/5/22/8642359/the-science-behind-the-new-netflix-design`.

S. Madappa. Ephemeral Volatile Caching in the cloud, 2012. URL `http://techblog.netflix.com/2012/01/ephemeral-volatile-caching-in-cloud.html`.

J. Mangalindan. Amazon's recommendation secret, 2012. URL `http://fortune.com/2012/07/30/amazons-recommendation-secret/`.

B. Marr. Machine Learning – Every Manager Should Read, 2016a. URL `http://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#7279dc29323f`.

B. Marr. Machine Learning: What It Is And The Milestones Everyone Should Know About?, 2016b. URL `https://www.linkedin.com/pulse/machine-learning-what-milestones-everyone-should-know-bernard-marr`.

F. J. Martin. The Past, Present, and Future of Machine Learning APIs, 2015. URL `http://www.slideshare.net/bigml/the-past-present-and-future-of-machine-learning-apis`.

W. Mason. Amazon Video is Bulding a VR Platform, 2016. URL `http://uploadvr.com/amazon-video-vr/`.

J. McCarthy and E. Feigenbaum. In Memorial. *AI Magazine*, 11, 1990.

R. Merrett. Where is machine learning heading in 2016?, 2015. URL `http://www.cio.com.au/article/590834/where-machine-learning-headed-2016/`.

C. Metz. Now Anyone Can Tap the AI Behind Amazon's Recommendations, 2015. URL `http://www.wired.com/2015/04/now-anyone-can-tap-ai-behind-amazons-recommendations/`.

Microsoft. Networking and Access Technologies, 2016. URL `https://technet.microsoft.com/sv-se/network/default`.

R. Miller. AWS Wants To Put Machine Learning in Reach of Any Developer, 2015. URL `http://techcrunch.com/2015/04/09/aws-wants-to-put-machine-learning-in-reach-of-any-developer/`.

M. Minsker. 10 Ways Amazon Has Changed Customer Service in 20 Years, 2015. URL `http://www.smartcustomerservice.com/Articles/News-Features/10-Ways-Amazon-Has-Changed-Customer-Service-in-20-Years-105131.aspx-cloud-arms-race-over/`.

B. Mitchell. What is computer networking, 2016. URL `http://compnetworking.about.com/cs/basicnetworking/f/whatsnetworking.htm`.

I. Mochari. The Simple Trick That Amazon, Apple, and Lego Know, 2016. URL `http://www.inc.com/ilan-mochari/strategy-execution-value-proposition-apple-amazon-lego.html`.

A. D. Moore. *Intellectual Property and Information Control: Philosophic Foundations and Fontemporary Issues*. Transaction Publishers, 2nd edition, 2009.

K. Murphy. *Machine Learning, A Probabilistic Perspective*. The MIT Press, Cambridge, Massachusetts, London, England, 2012.

National Conference of Commissioners on Uniform State Laws. Unifrom Trade Secrets Act with 1985 Amendments. In *Annual Conference Meeting in its ninety-fourth year*, Minneapolis, Minnesota, U.S., 1985.

Netflix. *Netflix Inc, Form 10-K(Annual Report)*. Edgar Online, 100 Winchester Circle, Los Gastos, CA 95032, United States of America, 2015.

Netflix. About Netflix, 2016a. URL `https://media.netflix.com/en/about-netflix`.

Netflix. Netflix Open Connect, 2016b. URL https://openconnect.netflix.com/en/.

Netflix. Quarterly earnings: Q4 15 letter to shareholders, 2016c. URL http://ir.netflix.com/results.cfm.

N. J. Nilsson. *Introduction to Machine Learning*. Stanford University, Departmant of Computer Science, 1998. URL http://ai.stanford.edu/~nilsson/MLBOOK.pdf.

E. Noren. Analysis of the Amazon Business Model, 2013a. URL http://www.digitalbusinessmodelguru.com/2013/07/analysis-of-amazon-business-model.html.

E. Noren. Analysis of the Amazon Business Model, 2013b. URL http://www.digitalbusinessmodelguru.com/2013/07/analysis-of-amazon-business-model.html.

J. Novet. Amazon Web Services brings in $2.4B in revenue in Q4 2015, up 69% over last year, 2016. URL http://venturebeat.com/2016/01/28/amazon-web-services-brings-in-2-4b-in-revenue-in-q4-2015-up-69-over-last-year/.

OECD. Glossary of Statistical terms - Market Power, 2002. URL https://stats.oecd.org/glossary/detail.asp?ID=3256.

R. Opdenakker. Advantages and Disadvantages of Four Interview Techniques in Qualitative Research. *Forum: Qualitative Social Research*, 7, Aug. 2006.

T. O'Reilly. What is web 2.0: Design patterns and business models for the next generation of software, 2005. URL http://mediaedu.typepad.com/info_society/files/web2.pdf.

D. Orenstein. Application Programming Interface, 2000. URL http://www.computerworld.com/article/2593623/app-development/application-programming-interface.html.

A. Osterwalder and P. Yves. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley Sons, Hoboken, NJ, United States of America, 2010.

OxfordDictionaries. Algorithm, 2016. URL http://www.oxforddictionaries.com/us/definition/american_english/algorithm.

K. Park. Data Communication and Computer Networks. In *Course CS536: Data Communication and Computer Networks*. Purdue Univeristy, 2016. URL https://www.cs.purdue.edu/homes/park/cs536-intro-1.pdf.

A. Paterek. *Predicting movie ratings and recommender systems*. 2012. URL http://arek-paterek.com/book/.

PCMag. Defintion of: Application Program, 2016. URL `http://www.pcmag.com/encyclopedia/term/37919/application-program`.

M. W. Peng, D. Y. Wang, and Y. Jiang. An institution-based view of internal business strategy: a focus on emerging economies. *Journal of International Business Studies*, 39:920–936, Dec. 2008. doi: 10.1057/palgrave.jibs.8400377.

G. Peterson. InterViews: An Introduction to Qualitative Research Interviewing. *Journal of Phenomenological Psychology*, 28, 10 1997.

U. Petrusson. *Intellectual Property & Entrepreneurship Creating Wealth in an Intellectual Value Chain*. Center for Intellectual Property Studies Chalmers University of Technology, Gothenburg, Sweden, 2004.

U. Petrusson. *Forskning & Nytta*. Tre Böcker Förlag AB, Gothenburg, Sweden, 2015.

U. Petrusson and B. Heiden. Assets, Property, and Capital in a Globalized Intellectual Value Chain. In B. Berman, editor, *From Assets to Profits, Competing for IP Value & Return*. Wiley, 2008.

U. Petrusson, H. Rosen, and T. Thornblad. Global Technology Markets: The Role of Open Intellectual Property Platforms. *Review of Market Integration*, 2:333–393, Aug/Dec. 2010. doi: 10.1177/097492921000200307.

I. Pilaszy. Recommending new movies: even a few ratings are more valuable than metadata. In *RecSys '09 Proceedings of the third ACM conference on Recommender systems*, pages 93–100, New York, NY, USA, 2009. ACM. doi: 10.1145/1639714.1639731.

M. Pinola. Data Storage Technologies of the Future, 2015. URL `https://www.backblaze.com/blog/data-storage-technologies-of-the-future/`.

B. Popper. Netflix is testing ads for Netflix original content on Netflix, 2015. URL `http://www.theverge.com/2015/6/1/8700921/netflix-advertising-house-ads-test`.

PRNewswire. Online Streaming Entertainment Providers are rapidly building Brand Equity, accroding to the 2014 Harris Poll EquiTrend study, 2014. URL `http://www.prnewswire.com/news-releases/online-streaming-entertainment-providers-are-rapidly-building-brand-equity-according-to-the-2014-harris-poll-equitrend-study-249693911.html`.

Y. C. Prof. Hong and F. Clemens. Criticisms, variations and experiences with business model canvas. *European Journal of Agriculture and Forestry Research*, 1:26–37, Dec. 2013. URL `http://www.eajournals.org/wp-content/uploads/Criticisms-Variations-and-Experiences-with-Business-Model-Canvas.pdf`.

J. Prunty. Three Kinds of User Experinece Design Fuel Amazon, Google, Apple, 2014. URL `http://www.forbes.com/sites/sungardas/2014/03/21/three-kinds-of-user-experience-design-fuel-amazon-google-apple/#2b2cc1c5bdbe`.

G. Quinn. A Software Patent Setback: Alice v. CLS Bank, 2015. URL `http://www.ipwatchdog.com/2015/01/09/a-software-patent-setback-alice-v-cls-bank/id=53460/`.

E. Rahm and H. H. Do. Data Cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, 23:3–13, Dec. 2000. URL `http://lips.informatik.uni-leipzig.de/files/2000-45.pdf`.

R. Reinsberg. Business: Netflix, 2009. URL `https://www.technologyreview.com/s/416843/netflix/`.

P. W. Roberts and G. R. Dowling. Corporate Reputation and Sustained Financial Performance. *Strategic Management Journal*, 23:1077–1093, Sept. 2003. doi: 10.1002/smj.274.

J. Roettgers. Amazon Clocks $107 Billion In Revenue In 2015, 2016. URL `http://variety.com/2016/biz/news/amazon-clocks-107-billion-in-revenue-in-2015-1201691106/`.

M. Rose and T. Li. Application Program Interface (API). In *Essential Guide: SaaS applications integration: Strategies and best practices for success.* TechTarget, 2004. URL `http://searchexchange.techtarget.com/definition/application-program-interface`.

P. Rubens. Future of Data Storage: 8 Technologies Changing Storage, 2014. URL `http://www.infostor.com/disk-arrays/future-of-data-storage-8-technologies-changing-storage-1.html`.

R. P. Rumelt. Towards a Strategic Theory of the Firm. In N. J. Foss, editor, *Resources, Firms, and Strategies: A Reader in the Resource-based Perspective*, New York, NY, USA, 1997. Oxford University Press Inc.

SAS. Machine Learning, What it is and why it matters, 2016. URL `http://www.sas.com/en_us/insights/analytics/machine-learning.html`.

J. Shaeffler. *Using Data to Drive Profit: Top 10 Keys for using Data Analytics in the Media and Entertainment Industry.* NewBay Media Inc., 2015. URL `https://www.grapeshot.com/wp-content/uploads/2015/12/IBM-Whitepaper-Using-Data-to-Drive-Profit-3.pdf`.

P. Simon. *Too Big to Ignore: The Business Case for Big Data.* John Wile and Sons, Inc, Hoboken, New Jersey, USA, 2013.

SourceMedia. 6 Machine Learning Giants to Watch: Amazon to Saleforce.com, 2016. URL `http://www.information-management.com/gallery/6-machine-learning-giants-to-watch-amazon-to-salesforcecom-10027443-1.html`.

A. Spyker and R. Meshenberg. Evolution of Open Source at Netflix, 2015. URL `http://techblog.netflix.com/2015/10/evolution-of-open-source-at-netflix.html`.

T. Stenovec. The Secret Magic behind Netflix Customer Service, 2013. URL `http://www.huffingtonpost.com/2013/10/31/netflix-customer-service_n_4178662.html`.

A. Strange. Amazon Unveils Flying Delivery Drones on 60 Minutes, 2013. URL `http://mashable.com/2013/12/01/amazon-unveils-flying-robot-delivery-drones/#TQ9e2EV8SGqm`.

Stratecery. The Amazon Tax, 2016. URL `http://us5.campaign-archive2.com/?u=a04b7dc3f187a3b6f52e8de49&id=0dbe603c42&e=bab35dd08e`.

R. Stuart and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 1301 Sansome St, San Francisco, CA 94111, USA, 2013.

L. Sullivan. Apple, Amazon, Facebook Gain Significantly In Brand Value, 2015. URL `http://www.mediapost.com/publications/article/259667/apple-amazon-facebook-gain-significantly-in-bran.html`.

D. Teece. Business Models, Business Strategy and Innovation. *Long Range Planning*, 43:172–194, April/June 2010. doi: 10.1016/j.lrp.2009.07.003.

E. Teijlingen. Semi-structured interviews, 2014. URL `https://intranetsp.bournemouth.ac.uk/documentsrep/PGR%20Workshop%20-%20Interviews%20Dec%202014.pdf`.

J. Tirole. Cognition and Incomplete Contracts. *Amrican Economic Review*, 99(1): 265–294, Mar. 2009. URL `http://econ.lse.ac.uk/staff/lfelli/teach/Tirole%20AER%202009.pdf`.

U.C. Berkeley. Distributed and Parallel Computing. In *Course CS61A*. U.C. Berkeley, 2011. URL `http://wla.berkeley.edu/~cs61a/fa11/lectures/communication.html`.

L. Ulanoff. Amazon Knows What You Want Before You Buy IT, 2014. URL `http://mashable.com/2014/01/21/amazon-anticipatory-shipping-patent/#mnw7NA1N.ZqN`.

U.S. Copyright Office. Report On Legal Protection For Databases, 1997. URL `http://www.copyright.gov/reports/db4.pdf`.

UsabilityFirst. Screen Real Estate, 2014. URL `http://www.usabilityfirst.com/glossary/screen-real-estate/`.

USPTO. General Information Concerning Patents, 2014. URL `http://www.uspto.gov/patents-getting-started/general-information-concerning-patents#heading-2`.

USPTO. Types of Patents, 2016. URL http://www.uspto.gov/web/offices/ac/ido/oeip/taf/patdesc.htm.

USPTO: Office of Policy and External Affairs. Trade Secrets Protection in the U.S., 2016. URL http://www.nist.gov/mep/upload/marinaslides.pdf.

P. van Eijk. Business Model Canvas for Cloud Providers, 2014. URL http://www.clubcloudcomputing.com/2014/08/business-model-canvas-cloud-providers/.

S. Vanghan-Nichols. Amazon EC2 cloud is made up of almost half-a-million Linux servers, 2012. URL http://www.zdnet.com/article/amazon-ec2-cloud-is-made-up-of-almost-half-a-million-linux-servers/.

M. Wade and J. Hulland. The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *MIS Quarterly*, 26:107–142, Mar. 2004. doi: 10.2307/25148626.

J. Wagner. Top 10 machine Learning APIs: AT&T Speec, IBM Watson, Google Prediction, 2015. URL http://www.programmableweb.com/news/top-10-machine-learning-apis-att-speech-ibm-watson-google-prediction/analysis/2015/08/03?page=2.

H. Wang, J. He, and J. Mahoney. Firm Specific Knowledge Resources and Competitive Advantage: The Roles of Economic and Relationship Based Employee Governance Mechanisms. *Strategic Management Journal*, 30:1265 to 1285, Dec. 2009. doi: 10.1002/smj.787.

T. Warren. The Amazon Fire Phone was always going to fail, 2014. URL http://www.wired.com/2015/01/amazon-fire-phone-always-going-fail/.

B. Wernerfelt. The resource based view of the firm. *Strategic Management Journal*, 5:171–180, 1984.

M. Winter and S. Siddharth. Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods*, 44:1–23, Mar. 2012. doi: 10.3758/s13428-011-0124-6. URL http://research.microsoft.com/en-us/um/people/nath/infographics-study/articles/mturkmethods.pdf.

C. Wodehouse. A Guide to Database Technology, 2016. URL https://www.upwork.com/hiring/data/a-guide-to-database-technology/.

World Intellectual Property Organization (WIPO). Trademarks: Valuable assets in a changing world, 2009. URL http://www.wipo.int/wipo_magazine/en/2009/04/article_0002.html.

World Intellectual Property Organization (WIPO). What is Intellectual Property?, 2013. URL http://www.wipo.int/edocs/pubdocs/en/intproperty/450/wipo_pub_450.pdf.

World Intellectual Property Organization (WIPO). Frequently Asked Questions: Patents, 2016. URL `http://www.wipo.int/patents/en/faq_patents.html#basics`.

A.-M. Yaser. The Learning Problem. In *Learning From Data*. Caltech's Provost Office, E&AS Division and IST, Apr. 2012. URL `http://work.caltech.edu/slides/slides01.pdf`.

S. Yegulalp. How IBM, Google, Microsoft, and Amazon do machine learning in the cloud, 2016. URL `http://www.infoworld.com/article/3039052/cloud-computing/how-ibm-google-microsoft-and-amazon-do-machine-learning-in-the-cloud.html`.

J. Zenor. Video privacy protection act. In K. Harvey, editor, *The Encyclopedia of Social Media and Politics*, Los Angeles, US, 2014. SAGE Publications.

K. Zetter. Apple's FBI Battle Is Complicated. Here's What's Really Going On, 2016. URL `http://www.wired.com/2016/02/apples-fbi-battle-is-complicated-heres-whats-really-going-on/`.

Appendix A

Appendices

APPENDIX includes a list over the interviewed for this research as well as the template for the interviews.

A.1 Appendix A: Interview Template

The interviews have been designed according to the research questions. Since a semi-structured interview method was used, the template was used as a starting point and focal point of the interviewed depended on the expertise of the interviewed. For example, a Machine Leaning researcher was only asked questions from the "MLS Characteristics" section of the template. Similarly, IP experts were mainly asked questions related to control mechanisms. Below the template for the interviews is presented.

Introduction
Short description of the purpose of the study as well as the theoretical foundation.

Short introduction of the authors and their academic background.

Clarification of how the data gathered will be used and ensure that consent is given for publishing the information:

"Before we get started we just want to clarify how the information shared during this interview will be used. The data will be used for our master thesis, which will be published at Chalmers University of Technology. For that reason, please let us know if there is anything you don't want us to include. Otherwise we'll assume that we can publish the information."

MLS Characteristics

1. Can you give us a short description of your role?
2. What characterizes [your] Machine Learning System?
3. What different technology fields does the system constitute of?
4. What characterizes each technology field?
5. How do the different characteristics and technology fields create value for the end-user?

Potential sources of sustained competitive advantage

1. Can you give us a short description of your company and what do you consider to be your core business?
2. Can you give us a short description of your role?
3. Can you tell us about [your] products/services that are based on Machine Learning technology and that has been successful?
4. In your opinion, what [make]/[made the] product(s) or service(s) successful?
5. What makes [ML companies]/[you] competitive over time?
6. What would you consider to be the most valuable assets for being successful over time?
7. What would you consider to be the most valuable capabilities for being successful over time?
8. How would you describe [your]/[market leaders'] position on the market?
9. How does market position influence your long-term competitiveness?
10. What would you consider to be the key ingredients to sustained competitiveness in this industry?
11. In hindsight, is there anything you would have done differently?(if applicable)
12. Looking forward, what do you think will be most important for companies to remain competitive over time?

Control Mechanisms
Short introduction of terminology that will be used, eg. the definition employed in this study for technical control mechanism.

1. How do [you]/[market leaders] use Intellectual Property Rights? Why/why not?

2. What type of legal tools are most commonly employed for inventions related to Machine Learning Systems?
3. Looking at property-based control mechanism, and specifically IPRs, how are IPRs used today for machine learning inventions? And what are the main implications with using IPRs to control businesses based on Machine Learning technology?
4. Are there any determining court cases that have had an impact on the deployment of legally enforceable control mechanisms in Machine Learning based business today?
5. To what extent is secrecy used to prevent imitation and to control businesses based on Machine Learning?
6. To what extent is technical used to prevent imitation and to control businesses based on Machine Learning?
7. To what extent is contractual control used to prevent imitation and to control businesses based on Machine Learning?
8. To what extent is market power/reputation used to prevent imitation and to control businesses based on Machine Learning?
9. Are there any other control mechanisms that you believe we have overlooked?
10. In hindsight, is there anything you would have done differently? (if applicable)

A.2 Appendix B: List of interviewed Persons

Name	Company	Description	Interview Date
Danny Bobrow	Parc	Expert within Artificial Intelligence with over 100 published papers, books and patents	2016-03-09
Oliver Downs	Amplero Inc.	Chief Scientist and CTO	2016-03-05
Hoda Eldardiry	Parc	Research Scientist, Machine Learning, Knowledge Discovery and Data Mining	2016-02-22
Sam Funnel	Stratified Medical	IP Manager with 25 years of experience working with IPRs in the ICT industry	2016-04-19
Industry Expert	Expert	Over 40 years of experience in Natural Language \& Machine Learning Technologies (AI)	2016-03-29
Eran Kahana	Mason LLP \& Stanford Law	IP Attorney and Research Fellow at Stanford Law School lecturing on legal aspects of using Artificial Intelligence	2016-04-04
Mike Kuniavsky	Parc	Principal Scientist, Ethnographer \& UX Designer	2016-03-22
Damon Matteo	Fulcrum Strategy	CEO, with 25 years of experience within strategic creation, funding, management and commercialization of high-value innovations and IP assets across industries in an international context. Named one of the "Fifty Most Influential People in Intellectual Property".	2016-04-05
Marzieh Nabi	Parc	Research Scientist, Data Mining, Machine Learning and System Theory (control and optimization)	2016-03-08
Bob Price	Parc	Research Fellow, Machine Learning, Interference, Tracking, Model-based Control, Modeling online purchasing behavior and creating predictive models of web-searching behavior and constructing recommendation sets.	2016-03-18
David Rose	Ditto Labs Inc.	CEO	2016-03-02
Michael Sollami	Ditto Labs Inc.	Chief Scientist	2016-03-02
Mathew Shreve	Parc	R&T Computer Science Engineer, develop and apply computer vision and Machine Learning algorithms	2016-03-09
Shivon Zills	Bloomberg Beta	Partner \& Founding Member of the investment company Bloomberg Beta	2016-03-10