NTNU – Trondheim
Norwegian University of
Science and Technology

MTM
Master of Technology Management

Intellectual Property and Machine Learning: An exploratory study

Lasse Øverlier

January 2017

MASTER THESIS

Department of Industrial Economics and Technology Management
Norwegian University of Science and Technology

Supervisor: Arild Aspelund

Mentor: Haakon Thue Lie

"*A breakthrough in machine learning will be worth ten Microsofts.*"

Bill Gates

Preface

This thesis completes a two year study program in Master of Technology Management (MTM) at NTNU. The study is completed while working at the Norwegian Defence Research Establishment (Forsvarets forskningsinstitutt, FFI), and the results of the thesis are in direct relevance to many of FFI's research areas.

The reader is not expected to have a deep knowledge of machine learning or the intellectual property rights of digital material. But the reader must have a good understanding of what makes innovation and intellectual property important for today's businesses, and be be proficient with new technology to understand how a company's freedom-to-operate can be affected by machine learning technologies.

Kjeller, 2017-01-31

Lasse Øverlier

"*The studies reported here have been concerned with the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning.*"

(Samuel, 1959)

"Some studies in machine learning using the game of checkers"

Acknowledgment

I would like to thank to my employer, FFI, for their support and willingness to let me do this study. And special thanks to Ronny Windvik for always having time to discuss my scenarios and ideas.

I also would like to thank my local supervisor Haakon Thue Lie for all the time he has spent discussing my high flying hypotheses and scenarios. It is great to discuss the potential impact of technology that does not yet exist, and have someone help me connect this to the relevant questions of the thesis.

And very special thanks go to my family who have endured my work hours.

L.Ø.

"*The only true wisdom is in knowing you know nothing.*"

Socrates

Executive Summary

Our research makes a contribution by exemplifying what controls the freedom-to-operate for a company operating in the area of machine learning. Through interviews we demonstrate the industry's alternating viewpoints to whether copyrighted data used as input to machine learning systems should be viewed differently than copying the data for storage or reproduction. In addition we show that unauthorized use of copyrighted data in machine learning systems is hard to detect with the burden of proof on the copyright owner.

We also demonstrate how existing products, both physical and software, becomes more vulnerable to reverse engineering with the rapid progress in machine learning, and how this challenges and reduces a company's freedom-to-operate and the way they appropriate innovation.

Another contribution is made from demonstrating how machine learning systems can create new valuable content from those patterns and structures found through parsing databases of texts, images, music or arts, and how this challenge the existing intellectual property regulations. We claim that with huge amounts of data used as input to a machine learning system, giving all intellectual property rights of output to the input data owners may be the wrong thing to do. We also think that intellectual property regulation should start discussing when a user of a machine learning system can be seen as having made a creative effort in the generation of the new content.

"*Intellectual property has the shelf life of a banana.*"

Bill Gates

Contents

"*You should be innovating so fast that you're invalidating your prior patents.*"

Elon Musk

Chapter 1

Introduction

The trigger for this thesis' research area started with a discussion over artificial intelligence (AI) and machine learning systems (MLS), and how new development in these areas challenge our current use and understanding of intellectual property rights (IPR). This is an approximate summary of this first discussion:

> "A human can look at copyrighted paintings and make a new painting where she/he clearly is influenced by the other painter(s). But can a machine learning system learn its tasks through finding structures and patterns in copyrighted material? Will the copyright owner of input data have intellectual property rights in the system output?"

We will get back to this and related questions later, but first we have to look into why this is an area of research and discussion.

Working inside the research industry, Norwegian Defence Research Establishment (FFI) has produced significant amounts of intellectual property since its start in 1946, and will continue this production as a research institute. Even if most of our research is open and freely available, we also need to control the use and access of some of our intellectual property (IP). Detection of abuse or lost IP is also important, even if FFI currently uses limited resources for this purpose.

In addition a lot of the research done at FFI is dependent upon using intellectual property from other businesses and research institutes. It has always been important for FFI to be consistent with the intellectual property rights of others and obtain the correct licenses for using others' intellectual property in accordance with international law. So there is a significant interest in looking at whether this will be different in the future of research, especially with the new emerging technologies and the huge amounts of data being created every day.

The main focus of our research is to identify and raise important intellectual property questions for businesses. When it comes to the IP questions regarding machine learning systems, none have been through the legal system, and most have not even touched by the legal system of any country. It will take some time until laws and regulations are able to address and set precedence on issues regarding the protection of intellectual property in the fields of machine learning and artificial intelligence.

Since the research questions raised in the thesis have a high degree of global relevance we are also referencing a lot of literature found in international databases, often legal research papers and articles from US and UK. We have made an effort throughout the study to increase the relevance towards Norwegian business practice, but the research will contain references to articles and regulations not directly relevant to Norwegian businesses. In spite of this we have chosen to include these to get a more complete view of the research problem.

1.1 Research Problem

There are challenges on multiple levels for a company attempting to establish or maintain presence in the field of machine learning. It could be a company developing new algorithms for machine learning. This company needs to know how they can protect their intellectual properties and establish or keep competitive advantage. It also needs to know whether the use of public data will influence how they can profit from the results and how the results can be protected. It is also essential to know if there are limitations to the (machine learning) use of any data, public or private, copyrighted or free.

In addition we have the dimension of the inventor, or creator, of the results. If a machine learning system finds patterns in the input data that can result in a profitable product or a service, the company needs to know who can claim ownership to the results. It is easy to see many potential claims rising from a relatively simple problem and that both the owner of the company/computer, the programmer, the owner of the ML algorithms, the user of the system, the owner of the input data, all may potentially claim some ownership in the results. This is another element in the discussion of a machine learning company's freedom-to-operate (FTO).

Freedom-to-operate can be described as the ability to perform business without infringing the intellectual property rights of others. Companies often perform an *FTO-search*, also called a *clearance search* or *infringement search*, before entering a new market to look for intellectual property claims that might exist. When working with machine learning and the system itself generates new valuable content, this seems to open up many new questions which we would like to address.

Research questions

This paper will look into the following research questions:

1. What parameters define freedom-to-operate for a company within the field of machine learning ?
    - What limitations can intellectual property rights on input data have on a company's research, development, products and services?
2. What intellectual property rights are suitable for protecting the intellectual assets of a machine learning company?
    - How will existing intellectual property protections be effective?
3. Are there differences in how a research institute, an established company with existing IP, and a new service/product developing company look at their FTO in this area of business?

1.2 Objectives

The new challenges exist both in the progress of inventions compared to the time it takes to protect the inventions through existing intellectual property protection mechanisms. In addition we have the fact that the machines "themselves" come up with the solution to problems. Machines cannot by today's laws be inventors and copyright holders. When all results and progress in the last decades' research in machine learning goes towards having machines become better at solving problems than humans, we are pushing the boundaries of today's regulations. This thesis will provide a discussion of these challenges for companies in this area.

1.3 Limitations

The most significant limitation of the study is the actual legal implications of our results and discussions. We conducted interview with a law firm, but acknowledge that without a legal background all legal topics in this paper must be taken as speculations or assumptions from our side. It is also limiting that we have used background material from both legal traditions, both common law and civil law, even if Norway is in the category of civil law.

Another limitation is that the interviews and discussions are made only with Norwegian companies and has few international viewpoints. This thesis is in the field of technology management, which includes management of intellectual property. The field of intellectual property law is very much harmonized (Lie, 2016). Business managers have to relate to foreign legal systems if they perform business abroad, so Norwegian firms must also understand e.g. US law. Our perspective is therefore from management science and not from jurisprudence.

1.4 Structure of the Report

This report starts by explaining the business challenges and opportunities that the recent introduction of machine learning has started in Chapter 2. In Chapter 3 we look at the theory of appropriability and IP-management in this new world of disruptive technologies, while Chapter 4 explains the methods used in our research. We start introducing the results from our interviews in Chapter 5, and continue with the rest of the results combined with a discussion of our findings in Chapter 6. We conclude in Chapter 7.

Chapter 2

Background

Machine learning (ML) was introduced as a term by A.L. Samuel in his paper "Some Studies in Machine Learning Using the Game of Checkers", and was split from artificial intelligence (AI) research in the early years of computer related research (Samuel, 1959). While AI research at that moment continued with the study of statistical pattern recognition, computational intelligence and later expert systems, machine learning has usually continued to be viewed as a part of this research area. In order to minimize confusion we will try leave the AI-term as a more general description of intelligent machines research, and use machine learning when most appropriate. But since most publications and press coverage describe challenges and evolution in AI, even when writing about machine learning, both terms will be used.

As shown in the title of Samuel's paper, machine learning research has, like AI research in general, always been closely connected with the automation of playing games. Most likely this is an easy-to-understand method of determining whether machines are better than humans at specific tasks. When systems capable of winning against the grand masters of chess (Hsu, 1999), and recently when defeating the best human player of the game Go (Yu, 2016), it has been taken as a sign of computers becoming better than humans in even the most advanced tasks. There are now special development communities for conducting research and developing game playing intelligence.

In 1993 Vernor Vinge came up with his statement describing *the singularity* as the moment in time where computers are more intelligent than humans:

> "*Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended.*" (Vinge, 1993)

But even if AI research brings us closer and closer to solving many of our current challenges, winning at games and driving our cars, few see *the singularity* happening very soon and the revolutionary breakthrough still seems far away (Chatfield, 2016; Achenbach, 2015). But when it comes to the term *creative singularity*, defining when computers overtake people as the primary source of innovation, this may be inevitable (Abbott, 2016).

Many computer systems may appear as intelligent. Like a chat-bot, trained on millions of movie scripts, able to hold a seemingly normal conversation with a human (Vinyals and Le, 2015), but they all have their limitations based on the training of the system (c.f. Chapter 2.1).

Professor Jon Bing, author and one of the pioneers of artificial intelligence and legal regulation of technology, believed that "it is not possible to copy the human awareness without copying a lot more than what we call intelligence"[1] (Rognli, 2007).

In order to understand the challenges and potential of intellectual property through the field of machine learning, we will first give an overview of the areas most relevant for our research.

2.1 Introduction to Machine Learning Systems

Machine learning has been making its way into every day life through better and more useful products in the last twenty years (McCorduck, 2004). But the limitations of computing power and efficiency of ML technology has held back research in this area. With the last two decades' exponential growth in computing power the impact from ML research has come in many areas. Products and services from robot housekeeping (autonomous vacuums, lawn mowers, smart lights, etc.) to self-driving cars with continuous eight camera visual analysis of the car's surroundings (Tesla, 2016).

Following all new technologies there are challenges on how to best protect your intellectual property when being a pioneer operating in this "unexplored area". This explosive

[1] Norwegian: "Jeg tror altså ikke det er mulig å kopiere den menneskelige bevissthet uten å kopiere veldig mye mer enn det vi kaller intelligens." (Rognli, 2007)

growth in machine learning technology has made many researchers claim that the existing protection mechanisms of intellectual property will not satisfy this new industry (Davies, 2011).

To understand the influence of machine learning technology we need to know how some of its basic principles work. A machine learning system can be viewed as a black box where you have an internal structure called the machine learning model. Into the black box you first enter training data and the box uses ML-algorithms to learn structures and patterns in this data and set the MLS' internal parameters (potentially billions of them) in the ML-model. After training you verify the trained model, and if the verification is satisfactory, you have created the machine learning system with the corresponding model that is to be used in the product or service.

We also would like to point out that there is a significant asymmetry when training and using machine learning systems. When *training* a complex machine learning system, you need huge amounts of both computing power and data. But when *using* the machine learning system only a small amount of computing power is necessary together with the input data for the product's task. This is why we are able to produce small devices capable of using advanced machine learning technology, like face recognition cameras, smart watches, intelligent apps with voice commands, intelligent drones with avoidance detection, etc.

The training of machine learning is often divided up into three main methods of learning; supervised , unsupervised and reinforcement. These methods describe if and how the machine learning system gets its feedback, the type of feedback, and how it learns.

Supervised learning The MLS is trained from a set of data that has already been labeled. E.g. a set of images and a classification of what is in each picture, like car, plane, dog, cat, ...

Unsupervised learning The MLS is trained from a set of data without any lables, and finds patterns and structures in the input data without input of right and/or wrong. Typical use is finding clusters of similarities in data set, or associations between data.

Reinforcement learning The MLS is trained from a continuous feedback (reinforcement signal) of information about whether its state is closer to achieving its optimum or not. Often used in automation tasks and robotics-related areas, and can be compared to the way a human learns to use a bicycle.

In the last years there have been three big breakthroughs which have had a great effect on the evolution of machine learning: Cheap parallel computing - GPUs[2], *Big Data*, and better algorithms (Kelley, 2014). GPUs and better algorithms have had significant effect, however it is the availability of huge amounts of data and associated technologies (*Big Data*), that has enabled MLS' recent growth. The fast evolution of the research areas affected by machine learning is also demonstrated by close-to-weekly revolutionary publications of research, like in (Esteva et al., 2017). This and the huge impact MLS has had on the last years technology evolution makes us consider machine learning to be a disruptive technology (c.f. Chapter 3.5).

The vast collection of data containing images, texts, videos, location, social network activities and more, has given the opportunity for computers to look for patterns and structures that could not be found in smaller data sets. *Deep learning* is one of the new methods of finding structures and patterns through multiple levels of abstractions (LeCun et al., 2015), and is currently used in most research involving machine learning. We will not go deeper into *deep learning*[3], but only accept that it is a term used when a machine learning system has a high number of so-called levels in its internal model. This enables a larger potential for finding other types of patterns and structures in the input data. We will stick to using the term machine learning in this paper.

2.1.1 The randomness of MLS

When training a machine learning system you have to give the internal model a set of initial values for the MLS to find an optimal setting for the training data. Using initial zero values is not possible (Sutskever et al., 2013). The system's parameters are therefore initialized with small random values, often thousands and millions of them upon the start of the training session.

During training the MLS change these internal values to find the optimum setting for its task, based on the training data and some verification input data. How these internal values change during training to find the optimal values is influenced by the training data itself and the order of the training data when being parsed. In addition most training of MLS now includes permutations of the input to get more variation, more easily understood with

[2]GPU - Graphical processing units capable of running thousands of mathematical operations in parallell.
[3]Pun intended.

images, like mirroring, rotating a few degrees in any direction, blurring, resizing, etc.

This is of course a highly simplified description. It is often difficult to properly train a machine learning system. But the main point is that minor variations in the training data, the MLS setup, or even the initialization, can give completely different values inside the machine learning system. Even in two machine learning systems that from the outside seem to perform identical work. In ML literature this is referred to as local minima, and there can be many of these in the same MLS (Goodfellow et al., 2016). This will therefore have an effect on the generation of content, and may have implications on intellectual property of output.

2.2 MLS and IPR

It is important to have a deeper look into which fields of machine learning that may create problems for intellectual property rights. We will do this through examples of usage areas that challenge today's rules and perceptions of legal use of freely available data. We will in the next chapter explain the methods of intellectual property protection, but will in this chapter expect a minimum of intellectual property knowledge. Like "copying without consent of the author is illegal", "if I create something new I have copyright to that", which both are correct with some minor exceptions to be discussed later.

We intentionally raise many questions throughout the chapter, but they are only asked to enlighten the problems for later discussion. And they are also used as a foundation for our interview series.

We use the term *producer* for the originator of the copyrighted work whether it is on paper, canvas, tape, concrete/marble or in any other analog or digital form. When a producer's works are digitized and put on-line, they become easy for machine learning systems to use and find structures and patterns. The producer has copyright on his work for a long time, which varies from country to country, and type of work. But in our research we only separate between intellectual property with copyright and intellectual property without.

This paper will not discuss the copying of others' copyrighted works for direct profit which is illegal. We will try to raise some questions about how to look at machine learning systems that are: influenced from, learning from, finding patterns from, creating new content from, and other new types of potentially using copyrighted material.

The main areas of our interest in MLS and IPR are summarized in these main parts:

- Restrictions on use of input data.
- Ownership of output from machine learning systems.
- Intellectual property protection of machine learning system.
- Intellectual property protection of output.

And we will look at these by looking at examples when MLS use and create different types of intellectual property: music, images, texts and art. Many of the challenges that affect one area, will also be similar for other areas. But it is important to visualize this in multiple areas to understand how machine learning actually will challenges the foundations of our current accepted methods of using and producing intellectual property.

2.3 Creating content

Intelligent machines with creativity have been the theme of many authors, and one of the earliest ones in the age of computers may have been Roald Dahl. In his short story "The Great Automatic Grammatizator" from 1953 he tells a story about a man using mathematical rules and grammar to construct a machine able to write profitable novels and mass produce them (Dahl, 1953).

With machine learning systems we are taking a big step towards non-human creativity. The important fact is that some machine learning systems make it possible to "reverse" the flow of computing. Meaning that you can produce content from your observed patterns and structures by specifying which features you want to emphasize and which to reduce. At the current moment this is a quite difficult task because it is hard to know which parameters represents what features. But with the exponential growth in the research area we expect this will soon become simpler.

This means that if you have learned patterns and structures from a database of texts, you can ask the MLS to create text using these patterns and structures by giving it a weighted list of feature values "explaining" how to generate the text (Sutskever et al., 2011). This could be new content of texts, software programs construction , article writings in $\LaTeX$, and poetry generation (Brooks, 2014; Burgess, 2016).

Even if both software programs and $\LaTeX$-code generated from these systems are very close to being syntactically correct and may be compiled with minor adaptations, the content does not make a lot of sense in the current state of research (Williams, 2016). But this is expected to change rapidly and we must therefore expect meaningful and creative texts coming out from these types of systems in the near future.

So if you train your machine learning system on all of Shakespeare's works you can create a system that writes like Shakespeare. If you do not add any additional external input it can generate texts with the same type of words, new sentences, new order of sentences, and same method of writing. But it is all based on the input from Shakespeare's works. Since there is no copyright on the input text, and new text has been created, a reasonable question to ask evolves. Who has copyright to the new text? Is it the input data owner? The programmer? The computer? The computer owner? The machine learning system creator? The user?

Referring back to Chapter 2.1.1, it will be hard even for two identically functioning machine learning systems to reverse the process and generate the exact same text. This is because even if the two machine learning systems work seemingly in the same way; the initialization, the training data, order of training data, and the possible permutation of the training data, may give the machine learning systems different internal parameters.

The research for content generation is also happening in the fields of art production (Sayej, 2016), music production, image generation and video generation (Joshi, 2016).

2.3.1 User of the system

What if the user of the system, the new producer, actually uses a lot of research and creativity to fine tune the reversing. Will this change anything? Many will argue that the new content might be a derivative work, or even still a copy, since all the patterns and structures are found from ONE producer's copyrighted material.

But what about when a system learns its structures and patters from ten producers? Do they all have one tenth rights to the new work? What if there are one million producers used in the MLS? In this case maybe no one can say “this is a copy of my work”, or perhaps hundreds will claim this of the same output work? Where and how can you find the proof of copyright infringements?

The user trying to describe and tune the content of the new work can put both a significant effort and creativity into it. When can we identify whether she/he is creative enough to qualify for having rights to the new creations?

In addition there is another dimension where the user can take into account which types of input in the training data that has better or worse acceptance from a specific audience and try to optimize for this. The new content produced from the MLS will now be optimized for best acceptance for the specific audience. Will this add to the creative effort for the user?

2.3.2 Music

One of the earliest examples of "assisted" production of music dates back to 1757 with Johann Philipp Kirnberger's "Der allezeit fertige Menuetten- und Polonaisencomponist" which was developed with the aid of dices (Nierhaus, 2009). Music composed with the aid from computers is as old as the computers themselves, and it is today almost unthinkable to create music without computers. Composers of today therefore have copyright even when using many computer based tools. But the basis for this is that the creative work has always been assumed to be from a human.

Iannis Xenakis built software and produced music from stochastic models already back in 1953 (Serra, 1993). Kostelanetz argued already in 1971 that the artistic input of the programmer of machine produced work, is enough for copyright (Kostelanetz, 1971). This has been the normal viewpoint on computer aided music production as long as the new creation is significantly different from other producers' works.

Even if the attempted production of a West End musical (Brown, 2015) might not be completely without human creation, others have been experimenting with automatic generation of folk music (Sturm, 2016), and of production of"new" Mozart compositions (Bloom, 2015). Generated music can be downloaded from these experiments.

Commercial music composed entirely by computers may not be that far away.

2.3.3 Images

Machine learnings system's most promising area of research is on visual analysis of images. With today's huge image databases consisting of millions of images, it has become easy to get plenty of data for training machine learning systems. The most common task for images in machine learning is object recognition where the revolution started with the mentioned *deep learning* and the so called AlexNet (Krizhevsky et al., 2012). Current research in object recognition is already better than humans at classifying objects in images (He et al., 2015).

But the challenge of creating new images using machine learning systems is only in its infancy. The first results were called "Inceptionism" (Mordvintsev et al., 2015) a term derived

Figure 2.1: Google's creations of funny animals from deep learning training data. (Mordvintsev et al., 2015)

from the movie Inception. These images are built by trying to force one specific category of the classification onto a new image and you have weird looking results as shown in Figure 2.1.

Figure 2.2: Drawing numbers from scratch (left) and drawing new images (right). (Gregor et al., 2015)

But learning how to sketch is also possible for machine learning systems. Gregor et al. demonstrates this in Figure 2.2 where the left field shows how the MLS iterates to create a digit. On the right they create images and display the closest match in the right column, next to the produced image (Gregor et al., 2015).

And this is only the very start of this research. When research image databases grows to Google or Facebook size, the machine learning systems gets better, and the potential for tuning the output is better understood, we will see new images with content that is more usable. And thereby also valuable.

2.3.4 Art

We have the same situation in art, where artists producing art in special ways may become more vulnerable for infringements and/or competition. The Internet project *The Next Rembrandt*[4] is demonstrating the potential use for machine learning in arts. This project uses machine learning systems to learn how Rembrandt painted from in depth analysis of his paintings, including high resolution imaging, three dimensional surface scanning, depth, paint stroke layers, etc. The main purpose is to create new paintings that look like they could have been painted by the master himself.

Using this type of technology will most likely only work with artists that have produced a significant amount of art.

2.4 MLS and reverse engineering

Reverse engineering can best be understood as "the task of finding out how something works". The legal framework for when reverse engineering is allowed has been fought in courts for decades. But we will only look at the technological potential that machine learning is adding to reverse engineering.

Typical reasons to use reverse engineering is to be able to understand, document, copy, improve, or reproduce the functionality of something. This something can be a piece of hardware, like Phoenix Technologies' reverse engineering of the PC BIOS in 1984 enabling the compatible PCs (Galler and Brown, 1995), or understanding how software works to build an alternative or a competitor, like DeCSS enabling all computers to play DVDs (O'Brien, 2003).

We will look at two different sides of how machine learning systems enables better methods of reverse engineering. First how machine learning systems are hard to keep secret, and then how machine learning can enhance science through reverse engineering combined with automation.

2.4.1 Protecting your MLS

Reverse engineering may also happen through creative use of the gadget, software, process, etc. that is being reverse engineered.

[4] *The Next Rembrandt* can be found at https://www.nextrembrandt.com/ (last accessed 27 January 2017).

Machine learning is often offered from cloud services as a place both to rent the huge computing power needed to train your machine learning models, and also as a place where you can launch your product as a service to others.

For your customers to actually be able to use your machine learning products you need to give them access in some way. If your product is a piece of hardware, they will have all access to both trying to open it up physically, and accessing it through input/output options available, like user and administrative interfaces.

When someone can see the output of your ML-product, they can send queries specifically designed to extract the inner workings of your machine learning system. Recent research in machine learning has proven that it is hard to keep your machine learning system secret if others can access it (Tramer et al., 2016). If you put your service out on a cloud system as a Machine-Learning-as-a-Service, detailed output may make this extraction even easier.

So when others can extract enough information to train their own machine learning model from yours, they have performed a type of reverse engineering purely based on use or access to the product or service. And protection from use is hard when you are trying to sell products and services.

It is important to mention that Tramer et al. do confirm that this only works for some types of machine learning systems, and they introduce countermeasures to slow down these attacks. But the principle is simple. If you let someone specify input and you e.g. perform a classification task, they will be able to tailor the input so their system ultimately will learn a significant amount of your machine learning model. How effective this is compared to training your own model may be discussed, but if you do not have access to the actual training data, this will give you information about the machine learning model without having the data.

2.4.2 The Robotic Scientist

So we have discovered that we can train our machine learning model based on minor input adjustments, and access to output. This is exactly what many types of companies need to make testing easier and faster.

Drug companies are already using machine learning and automation to do virtual drug screening for drug design, and it has been shown to be significantly better than existing commercial solutions (Unterthiner et al., 2014) .

This is further enhanced by automation in laboratories, like the design of a robotic scientist, “Adam”, to automate microbiological experiments on yeast. The system was simple, but capable of making hypotheses and runing the experiments. Here the creators argue for discoveries of novel scientific knowledge (King et al., 2009). The manual labor earlier done by graduate students and engineers are now being replaced by automated instruments, like robotic microscopes, liquid handling robots, imaging, etc. When a huge number of output from automated analysis is fed into machine learning systems it enables the discovery of patterns and structures not possible to observe by only a few results (Edwards, 2014). The precision in today's instrumentation setup also means that these types of processes can be parallelized and made even faster.

So making automatic analysis with feedback of the results is already possible and demonstrated. This means that when you e.g. have a production secret, but many factors are known, machine learning and automation may actually be able to simulate an equivalent of the production secret given enough time.

Let us take the hypothetical example of Coca Cola™. The company has a secret recipe of how to produce its Cola. Over a century of discussion over the recipe content, most ingredients are assumed to be known, but some are still only speculations. But even adding the “speculation ingredients”, it could now be possible to construct an automated lab where the process of mixing ingredients, heating, cooling, extraction, etc. is flexible and even the order of process elements can be automatically changed. Feeding all these conditions to a machine learning system, together with an analysis of how close the output resembles the real Coca Cola™, may actually over time produce a resulting recipe of ingredients and a process equal to, or an alternative to, the real secret recipe.

This was only meant as an example of where we might be heading. We believe that reverse engineering, combined with automation and machine learning will challenge us in ways never before deemed as possible.

Chapter 3

Theory

Our main research goal is to locate the possibilities and limitations that intellectual property rights can give a company in today's evolution of recent machine learning innovations.

How a company can secure the future value of their inventions is called appropriability. In (Ahuja et al., 2013) a company's appropriability is split into two forms, primary appropriability and generative appropriability. Primary is when company exploits inventions directly into becoming a product, service or licensable solution. Generative is how the company captures the highest possible share of future inventions coming out of (their) existing inventions. We will look at intellectual property rights in an appropriability context and how companies can appropriate value from their IPR in short and long term.

Generative appropriability is essential for companies using machine learning systems as everything the MLS creates can be used to gain new intellectual property. In order for a company to achieve this, it has to have rights to the creations of the machine learning system.

While some types of innovation are difficult to copy, like tacit and social complex knowledge (Schilling, 2012), other innovations are easy to copy or imitate and must be protected by other means, intellectual property protection. We will take a closer look into these types of protection mechanisms in this chapter.

3.1 Innovation and Intellectual Property Rights

In order to understand how intellectual property rights can be affected by the disruptiveness of machine learning systems, we will first present an overview of the different aspects of IPR.

IPR in a digital world consist of the following elements: copyrights, patents, trade secrets, design rights and trademarks. Trademarks and design rights relates to the general look and feel of objects, texts, logos and interfaces, and we will leave these outside of scope as these are of general interest and not specific to machine learning businesses.

Intellectual property (as in the Norwegian "åndsverk", French "droit d'auteur", German "Urheberrecht") can be defined as "the result of an intellectual creative effort made by a human"[1] (Wagle and Ødegaard, 1997). The etymology of these European terms indicate that they concern the right of the author and not the right to copy, as in copyright. The important terms in our definition are *creative effort* and *made by a human*. This means that not being human, not being creative and not making a real effort, will not produce intellectual property. We will get back to this in Chapter 6.

Businesses operate internationally so international IP-regulation does exist and is being handled by international agreements such as TRIPS (TRIPS, 1996) and Berne convention[2]. So even if common law (UK, US) and civil law (EU, Scandinavia, China, ...) are highly different, these international agreements have harmonized IPR for more than a century.

And there are other dimensions of intellectual property that we will not address. E.g. with music we have a plethora of people involved, rights for composer, lyrics writer, performer and more, called *related rights*. We will discuss the term copyright as if it covers both *related rights* and "åndsverk".

We will now make a deeper look into the relevance for copyrights, patents and trade secrets in this area.

3.2 Copyright in a digital world

The intuitive reasoning for copyright is to encourage authors and artists to create new works by protecting the producer's rights in what she/he has produced. We will look at copyright of digital works including digital versions of "analog" intellectual property. With "analog" IP we here mean physical works that are originally not produced in digital form. Typical examples are paintings, sculptures, printed books, etc. where the cost of copying usually is negligible from the cost of producing the intellectual property (IP). Like "Monolitten" a 17 meter tall sculpture in Oslo that took a team of sculptors 14 years to complete. With today's technology

[1] Norwegian: "Et åndsverk er et resultat av en intellektuell skapende innsats frembragt av et menneske." (Wagle and Ødegaard, 1997)

[2] Berne convention - http://www.wipo.int/treaties/en/text.jsp?file_id=283698

it can be scanned, digitized and mass produced by 3D-printers into all sizes in weeks.

Copyright laws exist for the copyright owner to be able to determine *if, when, where* and *how* the intellectual property can be published, copied, used, etc. It is important to notice that from the Berne Convention a copyright owner automatically has copyright to produced intellectual property upon creation, even without having to register the creative work.

A hundred years ago Borel used the analogy of using an infinite number of monkeys to type any known (and unknown) text if given some time (Borel, 1913). This is often visualized in later publications by referring to infinite monkeys typing Shakespeare's collected works. Some recent companies are trying to use this as a business model (QENTIS, 2015; All Prior Art, 2016) by generating all texts possible up to a specific length, and storing this in their databases. One of these companies is clearly doing this to gain copyright to all texts not yet written, but the other claims that their work will set people free by challenging the patent system into not being able to grant more patents.

3.2.1 Limitations to copyright

It is important to add that there are limitations to the rights granted. Use of copyrighted material for news reporting, teaching, research, criticism and comments, are covered by so-called *fair use* in the US. These are quite similar to other countries' exceptions. Fair use grants reproduction for the same purposes, but the problem arises because fair use is not a firm definition. Fair use is an evaluation of factors like (WSU, 2017):

Purpose of use Commercial vs. non-profit is perhaps the most important factor. Educational use is often positive, but many "educators" are way more commercial than university lecturers.

Amount of material used More than needed to achieve your (fair) purpose of use?

Effect on market Is the use harming the original market of the copyrighted material?

Nature of work What is the level of creativity involved in making the original material? Artwork vs. lecture notes vs. phone book listings?

How these factors are weighted is a complex problem specific to the individual use. The discussions of fair use when it comes to digital rights management have existed for a long time (Favale et al., 2016).

3.3 Patents

A patent granted to an invention is an authorization from the government in a country, to the inventor, to exclude others from producing, selling, importing and using the invention for a period of time. The patent is published and is fully available for the same period. To be granted a patent the invention must be new, useful, novel, non-obvious, adequately described or enabled, and claimed by the inventor in clear and definite terms (USPTO, 2016).

The patent system is generated to spur innovation by letting inventors get protection to exploit their inventions for a period of time, after which the invention can be built, used, sold by anyone. Patents are supposed to keep research and development investments high to appropriate value also in times of market uncertainty (Toole and Czarnitzki, 2006). Patents are an important part of a company's freedom-to-operate and to their appropriability.

The company AllPriorArt.com (All Prior Art, 2016) claims that they can strike against future patents by going through the US patent database and create descriptions of new inventions. This is not new, but they are trying to exhaustively create all possible descriptions of new inventions. By this they claim to be able to block new patent applications from being granted by being able to point to their own database where prior art exists (Rutkin, 2016; Noyes, 2016). This is quite similar to the strategy of infinite monkeys creating texts by claiming copyright to all texts the computer can produce. This will most likely not work (David Lizerbram & Associates, 2015).

Hattenbach and Glucoft addresses the IPR issues of computer-generated prior art and computer generated patents in (Hattenbach and Glucoft, 2015) by concluding that the future patent ecosystem must adapt to this. But they do not include how.

3.4 Trade secrets

A trade secret is information that is guarded and kept under protected access by the company. Typical information kept as a trade secret can be processes, formulas, software source code and designs. It is important to separate trade secrets which are company internal secrets, from classified information which is under governmental protection and regulation through national security acts, such as "Sikkerhetsloven".

A trade secret must include four concepts (Lie, 2016) :

1. It is business related technical or commercial information,

2. it is not known to the public,
3. it has commercial value because it is secret, and
4. it is the subject of reasonable efforts to be kept secret.

A trade secret is protected by law if the above requirements are held (Schilling, 2012).

Trade secrets are essential in today's markets for a business to develop and maintain a competitive advantage (EU Commission, 2013). If a trade secret is crucial for your company and it is stolen, either through illegal means or potentially through legal means, your company will by definition loose value. Therefore it is essential for a company to understand the threats against trade secrets. This is always hard in a fast evolving area of business as shown in Chapter 2.4.1 where new threat vectors may appear.

The important question is then whether the business can do a reasonable effort in keeping them secret (Stead and Cross, 2009). If the products or services are offered in such a way that they enable customers or others to extract or simulate the trade secret, it will no longer be a trade secret and the company may have lost its competitive advantage.

3.5 Disruptive innovation

When new technology emerges and starts to take over an existing market by creating a new market with a new value network, it is called a *disruptive technology* (Bower and Christensen, 1995). Disruptive *innovation* was later introduced because new technology usually is an outcome of an innovation (Christensen, 2016). A significant shift in technology will create great opportunities for new actors and huge challenges for existing companies in that market (Peng, 2013).

Our primary interest lies in the impact of machine learning on the research industry, established companies and new entrants. We look at machine learning as a technology where disruptive innovations already is taking place and will influence many others (Byrnes, 2016; Panetta, 2016). This is both because of the fast growing number of academic publications inside the field, and because machine learning is being combined with many other fields of research causing a significant impact on the evolution of those fields (Clark, 2016; Cave, 2016).

Disruptive technologies are often best visible after huge established companies vanishes,

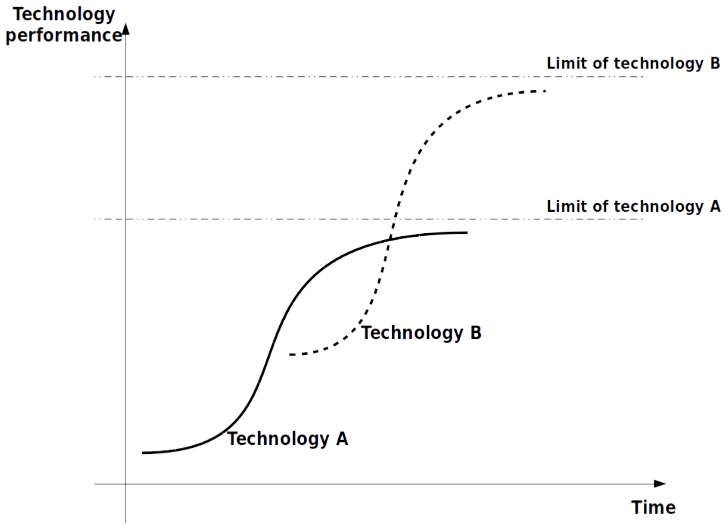

Figure 3.1: Techology S-Curve.

as with Kodak and digital photography, Blockbuster and on-line streaming of movies, Nokia and smart phones, etc.

The technology S-curve of R&D effort (Foster, 1986) demonstrates how technology performance (and value) evolves over time, shown in Figure 3.1. Significant[3] technologies starts out horizontally because in the beginning it is costly and time consuming to make the first iterations of new technologies. At some point the company is successful and the technology performance grows exponentially. And when you get closer to the limit of the technology, the growth stagnates and slowly approaches this limit asymptotically and improvements are more costly again. This has different meanings for the challenges and opportunities for established companies and for new entrants.

In Figure 3.1 we can think of “Technology A” as the long lived technology that has been the main factor behind the growth of an established company, and “Technology B” in its beginning as a technology that has shown potential for future products and services but is more costly to develop. Technology B is often the sole focus of new entrants trying to get into the market as Technology A might have lots of intellectual property tied to existing products and research. The start of Technology B happens in the “golden days” of the established company, and the company can easily defend to keep on investing in Technology A since progress there at that moment is significantly more lucrative than research in Technology B.

[3] *Significant* in the meaning that only technologies with an impact in the market is expected to follow the S-curve.

If an established company does not invest in Technology B from the beginning it may get huge problems when B's performance surpasses Technology A. Even when it has equal success in Technology B as the new entrant, the established company still carries an inertia from its old products and organizational model that must be converted to match the new market and its customers (Christensen, 2016, p.433).

The best advantage of an established company is that it *can* prepare for disruptive technologies. But established companies often do this the wrong way and focus on the technology itself, and not what the technology improves. In addition they often do not have a broad enough research and development (R&D) investment into what can become the new emerging technologies(Christensen, 2016). Thus they must put enough effort into appropriability of its R&D investments.

Another more successful move for an established company is to spin off an independent organization where more flexibility and better focus can be applied (Christensen, 2016, Chapter 10), or as an internal new venturing process (Hill and Jones, 2012, p.481).

One of the advantages of being a new entrant is that you can keep your focus on researching that specific potential disruption. Hill and Jones describes the scenario of having many ("a swarm of") potential followers to Technology A, but emphasizes that it is very hard to separate which will be the dominant factor in the market (Hill and Jones, 2012).

3.5.1 Machine learning and disruption

We look at the disruptiveness of machine learning both as what is happening inside the field of machine learning, but also machine learning as a catalyst for other technologies, products and services. Systems that earlier consisted of other types of algorithms with low accuracy, have been replaced with machine learning systems enabling revolutionary better products of cameras, language translators, voice recognition, video/image/text analysis and reporting, etc.

Machine learning creates the opportunities for several types of technological areas to experience the swarm of potential followers to the existing technology as described in (Hill and Jones, 2012, p.253).

3.5.2 IPR and disruption

The current IP system has lived through more than hundred years with only some adaptation. The reasons for adaptation could be instantiated to create a better balance between the dissemination of knowledge, and the needs for the inventor to appropriate value for the invention. An example is how the copyright system was adapted to appropriation of software code, and how the patent system has been gradually modified to not protect pure business methods (Borella et al., 2015).

Fraser claims that the potential for computer generated patents could result in a problem. That the cost of hampering competition is more useful than the value of patents as an incentive for innovation (Fraser, 2016). And maybe that the 20-year monopoly has become a too large reward compared to the cost and purpose of creating the patent.

3.6 Ethics and machine learning potential

Even if laws and regulations are what matters for the operations of a business, most companies have their own ethical rules. There are two types of ethical rules. Most businesses have so-called *normative ethical rules* meaning a code of conduct that is applied to everyday decisions and activities, while others use a *descriptive approach* and describes the actual behavior of the businesses and its employees (Clegg et al., 2012).

One important part about business ethics is to adapt to the cultural diversity of ethics in different countries. Some companies choose to adapt to the local ethical code even if it lowers the standards of their own ethical guidelines just to be able to do more business and make more money, like in the recent Yara corruption case (Jacobsen, 2015).

It is also important to realize that business ethics are not separate from personal ethics, and that ethics really is an individual responsibility even if it is heavily influenced by organizational culture and leadership, and internal decision making processes (Hill and Jones, 2012, p.403).

We can also separate between negative and positive ethics (Howard and Korver, 2008). *Negative ethics* consists of prohibitions like “You shall not copy others’ work.”, and are often easy to follow and guide from. *Positive ethics* are on the other hand obligations of the form “You shall...”, and are often not as clear and requires more effort to live by on a daily basis.

When it comes to machine learning systems, services, processes and products it is important to acknowledge that you need to have good ethical guidelines in place at the earliest stage especially when regulation is not in place (Cortez, 2014). Because, even if lack of explicit regulation today potentially opens up for discussion whether use of copyrighted material in an MLS is learning and not copying, it should be an ethical question for the companys guidelines. And even if it cannot be proven that your company has included copyrighted material in the training of your machine learning system, it does not make it less of an ethical question.

New disruptive technologies always push the edge of legal regulations and thereby also the ethical discussion.

“*The question of whether a computer can think is no more interesting than the question of whether a submarine can swim.*”

Edsger W. Dijkstra

Chapter 4

Method

We have based our exploratory research on an extensive search for relevant literature, papers and articles. Combined with this literature study we have conducted a series of interviews with businesses related to machine learning.

4.1 Target variance

We have tried to collect data from a variety of sources. Our broad collection is targeted to give views from most relevant angles, but was also necessary as there are a limited number of businesses that exists in the target group. We aimed for including both new/startup businesses, established industry, research institutions, business advisors and law firms. We would have preferred to include a higher number of companies from each of these groups, but were pleased with having from one to three interviews within each of these groups.

4.2 Inductive vs deductive reasoning

Deductive reasoning is made through the chain of:

$$theory \rightarrow hypothesis \rightarrow observation \rightarrow confirmation$$

In our field of research with this few observations available it would be hard to succeed by using deductive reasoning.

Inductive reasoning can be simplified through the chain of:

> *few observations → seeking patterns → forming hypothesis → likely conclusion/theory*

Inductive reasoning is therefore a more realistic approach for our research (Johnson-Laird, 1999; Bradford et al., 2015). By collecting interviews from researchers, product developers, business advisors inside these groups, we have tried to collect an overview of how these new technologies challenge and can influence appropriation and thereby also decision making and management.

One methodology of executing inductive research is described in (Gioia et al., 2013), but we found it too extensive to be used in our research. We are basing our research on grounded theory where we seek to discover emerging patterns in the collected data (Glaser and Strauss, 2009), and have chosen a more simplified method based on the chain of inductive reasoning given above.

4.3 Qualitative vs quantitative

For inductive research the small amount of observations or an intensive study are the best reasons for choosing a qualitative approach (Stern, 1980). If there are a large number of observations or an extensive study the most suitable would be a quantitative approach. With our qualitative approach we try to collect enough data to locate significant patterns and draw conclusions which can help build theories from the data.

There are only a limited amount of businesses operating in our selected area of machine learning services, products and research. In addition the new significance of intellectual property inside this field, makes the collection of data highly time consuming and located at only a few sources. The research is therefore completed as a qualitative study.

4.4 Literature study

A literature study should always be in place for background and discussion. In the case of fast evolving technologies like machine learning the number of publications and articles are growing exponentially which has made the literature study highly challenging. Many of the publications referenced in this paper are published during the period of performing the research. 37 out of 81 references are newer than 2015. By having this very dynamic environment we have also had to base the literature study on a large amount of on-line articles and magazine publications.

The wide literature study made it possible to have knowledge stretching into many related topic areas during the interviews, which made discussions easier. But the literature study also had an extended length in time and had to continue both during and after the interviews because of the dynamics in this area of research.

4.5 Interview series

The field study of the research is performed through an interview series with participants from different business groups: research institutions, established industry, startup businesses, company incubators/advisors.

Since we are exploring a new field of research there are a limited amount of businesses and people with high competence and experience. This made us seek in-depth interviews in place of big focus groups or surveys (Boyce and Neale, 2006). This also meant that we could not perform a huge amount of "fast interviews", or a large survey for quantitative results. Because of this we ended up making a set of qualitative in-depth interviews.

A total of seven businesses were interviewed within a period of approximately one and a half month during fall of 2016. All companies were interviewed under the promise of anonymity to allow as free and open dialogue as possible. Some views may push todays regulations and laws and we did not want those views connected to any business even if they were the views of individual employees of that company. Because of this the recorded interviews were transcribed immediately after the interview and anonymized to only belong to a company in a group as shown in the list below:

(A) One business advisor/management company.

(B) One law firm with digital IPR as their specialty.

(C) & (D) Two newly started businesses working with machine learning as main component in their research, product and/or service development.

(E), (F) & (G) Three established industry companies with various degree of research spending, but all having large amounts of existing intellectual property.

By not using survey style interviews and adapting to a more loose interview style we avoid the strict questioning and keep room for interpretation, new ideas, new viewpoints and an open discussion. In addition the anonymity hides any controversial views and statements from being attached to any one person or company.

But there are challenges of holding this kind of more loose style of interviews as discussed in (Boyce and Neale, 2006).

- Loose interviews are time consuming as you have to plan well in advance to be able to reach all the appropriate areas. In addition it takes a lot of time to record and write down everything that is said and then put everything into system and correlate with the the other interviewees.[1]

- The interview can be prone to bias if e.g. seeking to confirm pre-existing theories or not asking all the questions (Kahneman, 2013). This also covers the challenge of preparation and having an interviewer with experience.

- In addition the results are not generalizable because of small sample set.

We have made an effort to avoid these challenges in our loose style interviews by going through an *interview guide* at the end of each interview where the interviewees are able to extend, comment or change their statements. The interview guide is included in Appendix A.

It is important to keep our focus on addressing our research questions when analyzing the interviews. Therefore we have kept the main structure of the research questions throughout both the interview results and the discussions. This enables us to create some structure in the huge amounts of input gathered from the loose interview style.

Grounded theory (Glaser and Strauss, 2009) guided us towards tagging collected data (quotations) with codes and sort them into categories based on areas related to our research questions. The resulting structure after categorization and sorting can be seen in the structure of Chapter 5. This collection and categorization of input data was then discussed, put in relation to our literature study, and we built our conclusion.

[1] Our seven interviews of 45-60 minutes in length, came to a total of more than 60 full pages of digital text, excluding text from our introductory presentation.

Chapter 5

Interviews

The basis of the interview sessions were to introduce the companies, which we from now on refer to as A, B, ... H, into the potential of machine learning systems and especially the possibility for a machine learning system to create valuable content. Even when we had made a decision to have a loose interview type, we needed to introduce our setting and areas of discussion.

We completed the interviews in the fall of 2016 while publications and articles were published weekly in the machine learning area. The order of the interviews were not A–H. The law firm interview suddenly became possible at the end and we have pushed most of the statements from this last interview out of this chapter and into the discussion. Not interviewing the law firm early in the research made it easier to ask and discuss all kinds of views on use, creation, protection, etc. of content with all the companies.

The interviews were structured to follow this proposed agenda:

- Introduction of interviewer, background and reason for research.
- Introduction of interviewees, the company and their background.
- Introduction to machine learning including reason for choosing them, and introduction to freedom-to-operate and intellectual property protection.
- Discussion around the company's use of machine learning, if any.
- Introduction to new creative use of machine learning - especially the case where systems are generating new valuable content.

- Discussion on learning vs. copying. Copyrighted vs. free vs. expected free input data, etc.
- Intellectual property questions in this new scenario.
- Discussion about intellectual property and machine learning in this new setting.
- Go through check list and see whether all points are covered during discussion.
- Optional questions and discussions.
- Expressing thanks for participating.

And for most of the interviews this was the progression of the interview. With companies A and B we skipped the discussion of their company's potential use of MLS. In these two interviews it was more important how they looked at the raised questions and how they would give advice to their customers if they were in similar situations. They took the role of advising both owners of copyrighted data, and users of others' copyrighted or free data. In the interview with the law firm we also added some additional points to the interview guide as shown in Appendix A.

It must be emphasized here that none of the persons participating were legal experts with the exception of the lawyer from company B which is one of Norway's foremost experts in the field of intellectual property rights. The other companies all made this clear, but answered to the best of their knowledge and with what to them seemed reasonable in the current setting.

Just a few of the interviewees were highly knowledgeable when it came to machine learning technology and its in depth functionality and use. So some misunderstandings were inevitable, but the statements shown here were the most relevant ones. So even if not all statements makes 100% sense we are able to understand what they meant.

Disclaimer:

All interviews were conducted in Norwegian. Therefore all quotes from the interviews have been translated by the interviewer. In some statements the questions had to be integrated into the quote of the company to make sense of what the statement referred to.

In addition the most discussed subject was images and the creation of images, as these are relatively easy for everyone to relate to. So if the type of intellectual property is unspecified it will in most cases be related to images.

We present here the main results from the interview series to show the width of the discussions that arose. There was a lot more material that could be included, so just a small portion of the expressed opinions from all companies are used here. But we have through our coding and categorization of statements, been able to keep focus on the most relevant statements from each category. We have of course been forced to leave out many good statements. Not because they were in conflict with what others said, but most often because they had similar or overlapping views.

We start by using a statement of when something becomes an intellectual property. The process does not have to take time, but it does have to be creative. Or as one company so eloquently said:

> B: The amount of work is not significant. You can spend ten thousand hours and not create any intellectual property, and you can spend one minute and create the world's most beautiful poem - that IS intellectual property. It is not the amount of time spent, it is the creativity.

5.1 Limitations on freedom-to-operate

What early became clear through the literature study was that questions about the use and rights to input data were essential in controlling their freedom-to-operate. This was emphasized throughout the interviews as all companies had strong views on this.

5.1.1 Use of open/free data

Use of free data, or data licensed free for use, may still have limitations on them. Or they may put some demands on the resulting system, service or product.

> A: We use data licensed under Creative Commons where the only thing the data owners ask for is attribution, even if used commercially. That is a fair claim.

In addition there were concerns with the integrity of free data, whether the data is genuine and how can you be sure of that:

> E: We cannot use open data to train MLS used on critical infrastructure. What if something happens? You need other sources of data and not open or "Google-based" data.

Putting together many information pieces from open and/or free sources may still give problems with privacy issues and classified material. Some knew that combinations of information and collections of information may be regulated by other rules than IPR. Even if collected on free/open sources.

> A: There was this case involving Norwegian researchers walking around and taking pictures. As they claimed they had only used open and accessible information, but when put together it gave a description of the Norwegian navigational system that the Norwegian Defense claimed was classified after the Security Act.

> A: The way you put (small parts of personal) information together can make you end up with privacy regulated data.

5.1.2 Use of unmarked data

If data is of unknown origin but expected to be free, like image collections, there were larger concerns, but they could all be addressed to be risk management for the company.

> A: This is risk management. We would recommend them to isolate the risk so they easily could eliminate it/them.

> E: I would be terrified in using others' IP as input to our commercial machine learning systems, especially since laws and regulations lack behind. Freedom-to-operate would be threatened for me.

G: By using others' data you may end up with a model, or several models, that are alike but trained on different data. Then you can risk that the best model really cannot be used because of the (restrictions on the) training data.

5.1.3 Use of commercial data

All companies would purchase commercial data before using them. But some expressed concerns if the source of data inserted watermarks or other visible IP protection mechanisms.

G: If someone I purchase images from injects watermarks into them I would not be a happy customer. Because then I have not received pure images.

And if you should be so unlucky that one of your used images had another license than what you initially thought - like discovering a commercial image in someone else's image collection, you need a plan.

A: By removing (copyrighted) images upon request we most likely have not broken EU law. We have been told something is wrong and immediately taken action to correct this.

A: But if I build my system so I am not able to remove it, I have a problem. You need traceability, so when you are forced to remove images from the learning process, you are able to do this.

5.1.4 Use of internal data

The established companies were concerned with protecting their own existing data, something the new entrants were not so concerned with.

F: We have received all rights to images put into our systems. In some areas we do not, but these are kept separate.

G: We are highly interested in protecting our own production of data.

G: Our gut feeling tells us that if we have the rights to use the images, we own the information we are able to extract from them. This is like Facebook and Google.

As one of the new entrants put it. Protection of their technology is their first priority and more important than data.

C: For us data protection is not so important as algorithms and systems and where the borders of technology (protection) are.

5.1.5 Internal IP management

For all the companies it is important to have control over your intellectual property and over the intellectual property that you use in your systems.

C: We purchase all data and have clear license agreements with what we can do with the data.

There are no limitations for using the data in MLS.

Some have fair use limitations and we cannot reproduce everything, but perhaps a summary.

Most companies were in agreement when discussing how it would be hard to detect the abuse of intellectual property in large machine learning systems.

B: I believe it will be extremely hard to detect that your intellectual property is used in an MLS.

C: It is very hard to determine whether someone has used your IPR in a huge machine learning system.

But the companies are taking precautionary steps to try to be ahead of the competition and potential abuse.

> F: We have received all rights to images put into our systems. In some areas we do not, but these are kept separate.

The same company expressed some concern about users putting copyrighted information into their database.

> F: We do not know if there are copyright on all images in our database. We do make users "sign" that they have all rights to images they upload, but some do not (have all rights).

5.1.6 Creating new content

The idea of using machine learning to create new valuable intellectual property was quite new to all companies participating in the interviews. Therefore it was highly interesting to hear their first thoughts on the subject.

Most of the companies are here expressing their gut feeling and have not had a lot of time to think through the potential challenges. But they had many opinions on the appropriability for their company. Both when they were the creators and when their IP had been used to create new valuable content.

The aggregation of data is one of the important foundations of research, and for machine learning it is essential as structures and patterns are found only in large enough data collections.

> A: There is a long history of use of open data where knowledge is aggregated, so why should open data not be used in machine learning systems? The principle is there, but maybe with machine learning you make another step, not just using the data, but creating new content is potentially a lot more valuable.

> G: I look at machine learning as a tool to build the product (output).

Some saw the creation of new content not as a problem of today, but more of a future challenge that currently does not affect them.

F: Creating new material is highly interesting. We are not affected from this threat at the current moment.

E: These (creating new content) are very interesting problems. We have not thought this far ahead, but it is good to be challenged.

But there were many viewpoints on what generation of new content really means for intellectual property.

C: I would look at machine learning more like humans learning. A person reading a lot of books is allowed to write in the same style. Intuitively I would expect that a machine can generate new content. That it makes the computer the owner of the content.

D: If you create something new from the MLS, it becomes a legal question whether it is “close enough to” something existing.

E: It matters whether the system creates the same image every time, if there are predictable methods to produce the image, or if there are random elements involved making every image unique.

One company also related this to music production.

G: In a music production it is in varying degree the person behind the computer that makes the result the way it is. A lot is made automatically.

And there were also some challenging open questions asked about creating new content.

A: Maybe new content cannot have copyright?

Maybe new images will be seen as photographies?

Or maybe just using images is breaking business ethics?

Some of the companies had strong views on the situation where the producer of new content was Google or other large intellectual property owners.

C: Intuitively if Google computers creates an image, Google is the artist because they used a tool, the computers they have bought and the software they have written. This is quite similar to a painter with a brush and a canvas. Google owns the new image.

E: When producing new images there will be a change as soon as e.g. Google charges customers for new images. Then users will look at how to get a part of the revenue.

C: If they have used a thousand images as input/inspiration and generated something new, I would say they (Google) are the owner. Similar to a human inspired by a thousand images of Odd Nerdrum, paints his own inspired by them, he is still the intellectual property owner.

E: It seems natural that those who put resources into a system also wants some rights to the output. So it is natural that Google and others will claim some rights.

And one company was particularly interested in the meaning of the future of intellectual properties when it came to the creation of new content, raising many interesting viewpoints.

E: It makes no sense to have intellectual properties rights in this new picture, because you can always make a new image that is quite similar.

If the variation in images gives everybody a different image. Then I cannot see anyone having any intellectual properties in that image.

> We also risk that this is a passing problem. When too many have access to too much, this goes away. We will have thousands of systems that can make any image, so the rights will have to be quite small.
>
> We might see this as a temporary problem. You can end up with a situation where you generate an image you pay for once. Meaning you pay for the service and not for the image. Because keeping up with copyright in this scenario makes no sense.

5.1.7 Reverse engineering

Most companies were fully aware of the legality to reverse engineer products. Some did see this as a potential threat, and there were varying views on the types of products and the level of details that should be allowed to reverse engineer.

> E: Combined reverse engineering from video and multiple sources is worse for us. We do want to protect ourselves from this, but currently we have no plan to counter this.

> D: Reverse engineering using brute force into a black box I would view as copying and a break of copyright. Maybe machine learning models shoud be looked upon as some kind of design works?

Some also had the understanding that user agreements and terms-of-use contracts could restrict the threat of reverse engineering.

> C: If you have found out how the MLS works, but not copied the code. They are not protected by copyright. Then you might have broken terms-of-use or a contract you have signed to get access. In most of these contracts it states that reverse engineering is not allowed.

And some questioned the legality of agreements that forbid reverse engineering.

A: Through observation and use - have you done reverse engineering or not? When you get a license or buy something you sign that you will not perform reverse engineering. Is that according to local law?

D: The question of whether it is legal to reverse engineer an MLS by normal access is very interesting. If the MLS is more than a simple algorithm it should have some protection because it involves a piece of design work.

A: Could you have a conversation with the chief security officer where she/he explains that you can use our MLS, but you cannot do "this and that" and especially not try to reverse engineer it?

Business opportunity was also a potential result according to one company. Maybe automatic reverse engineering has a potential for future engineering?

E: What if we instead of employing smart engineers can have machine learning systems trying permutations 24/7 and eventually succeed with creations?

And others could easier see the amount of work needed to copy a machine learning model through usage, and therefore viewed this as a minor problem at the current time.

G: It is not "free reverse engineering", as there will be a huge number of queries. Maybe just as easy to train from scratch?

And questions were asked whether reverse engineering could be stopped through local or contract law or have the machine learning model viewed as a trade secret.

A: Maybe you can protect the internal parameters of a machine learning system by defining them as a trade secret. It would be easy if it was a microprocessor that it took one year to find out how it works. A new drug where you cannot observe how you made it is also OK. But many other things are not OK to define as a trade secret.

5.2 IPR suitable for machine learning

The other research question was to find out what type of intellectual property rights that are best to use for a business involved in machine learning. We have tried to split the views into the three fields of IPR we identified earlier as most relevant for machine learning businesses: patents, copyright and trade secrets.

There were expressed opinions about using design patents as a way of protecting your machine learning system service or product.

> A: I think design is important, because when you need to relate to these kind of systems the user interface is important, and I think secrecy is important.

> B: When it comes to design it is just the layout that matters. Screen shot. All it takes is a claim for being new.

Several companies expressed concern with starting to change the definitions, regulations and agreements of IPR since there already are a lot of existing international agreements already.

> E: Changes in existing IPR systems will create problems with tens of thousands of existing patents. Changing this is extremely difficult as existing regulation is accepted by 150-200 countries.

But even when having identified these international challenges we took a deeper look into the existing protection mechanisms of patents, copyright and trade secrets and kept our focus on the machine learning challenges.

5.2.1 Patents

There were many expressed meanings on what a patent is and how it can, or cannot, be granted.

A: To get a patent you need to have a technical effect, and knowledge is not a technical effect.

C: Patenting ML technology is hard because it is usually software.

A: Most patent applications will not go through because they are not novel or because of prior art. And a lot will be stopped for not being patentable, stopped by the ALICE-case in the US, and in the EU by being a process.

Many applications will be minor problem solving products and very hard to patent.

G: Maybe you can patent the numbers in the MLS and the training procedure, which order you have on your input data and preprocessing.

A: Practically we must relate to the patent system because it exists and is not something you can ignore.

And there were different views on the usefulness of patents in the future of IP protection and whether they could protect themselves with patents.

A: It is an interesting question whether patents promotes innovation or not, inside the area of machine learning. It may be that the problem solves itself because it either has no big effect, or because it cannot be patented within the regulations we have today.

C: We have attempted to patent technology through a patent firm with a negative reply. Not possible in Europe, and unlikely in the US.

F: We have no patents but we do have trademarks.

Patenting machine learning was also interesting for all companies. Especially if it really is possible, and how this could be done.

G: If you patent the entire MLS, all the way down to the parameters. Then it is a question of how much must be different before you have actual change? You can change some and still have the same result, even if this is a new MLS model.

C: From our understanding software is not patentable, but a system as a whole, hardware and software and its functionality, may be patented.

One company drew the parallel between a machine learning system and its millions of internal parameters, to the music/sound standard mp3. Some parts of the mp3 standard is patented and the patents expire in 2017.

G: I think that the machine learning algorithm is quite similar to the mp3 protocol/description. While the content is the sound itself.

And there were many concerned views on the patent from Google often referred to as the "Drop patent" or the "Dropout patent" (Hinton et al., 2014). This patent covers a technique that is used in almost all the most promising deep learning systems of today. The patent was granted to Google during the writing of this thesis, and a great deal of uncertainty lies in how Google choose to enforce its rights.

C: The drop patent from Google is interesting. What does Google want to use it for? Will they sue others? What about the risk for other companies using patented technology? What if they are sued? Are management responsible? What might Google do in the future?

The Google drop patent is really a software patent. To get this they must have described it as a system. It is still unclear to me where that borderline is and what it takes to be on the correct side.

5.2.2 Copyright

Every company knew that only humans could own copyright, but some of them were also open for discussing this when it came to machine learning. They saw that the aid from computers in creating intellectual property continues to grow extremely fast, and that this will be a question for the future which cannot be avoided much longer.

G: I think the argument of no copyright when created by a computer is strange. Most of what we produce today is the result of computer assistance.

A: There is a risk of not knowing how the law applies to this area. Maybe copyright will collapse when machine learning systems are able to create valuable content?

As there already have been new creations of music from machine learning systems the companies did not know about, these are reasonable questions and statements (Bloom, 2015; Sturm, 2016).

Some companies used analogies that implied that the use of machine learning is similar to storing a compressed version of the original image.

D: I prefer to look at machine learning as a compression algorithm. So if we were to compress an image to jpeg, it is still the original image we are dealing with. In ML you compress the essence of the image. Making it a smooth transition from when you compress the essence to when you compress the actual image.

If you change the original image enough it will become a new derivative work and a new creation. For me the work of ML is creative work.

G: If you have learned something from the image, you have compressed and stored the image in compressed form. Then you in some sense still have the image, only compressed.

I think you (in ML) can say you have stored a compressed version of the image you have used.

Examples of creative use of others' intellectual property that has been accepted was also pointed out.

A: In the TV-series Expanse you have something from all sci-fi. Space stations, fighting space ships, Blade Runner investigator, mystery. *Everything* is there. It is obvious that they have asked the public what they wanted, and made it. But in *every* scene I think: "I have seen this before", "this is Alien",... The show is completely shameless. It is good. This is how fiction works. This is how poets work.

D: Someone trained an MLS with Picasso images and made a rendering of the movie 2001 made up with Picasso paintings on all objects (Joshi, 2016). I would say that definitely is derivative work of both 2001 and Picasso, and has its own intellectual property.

5.2.3 Trade secrets

It became clear through all the interviews that secrecy was an important protection mechanism for the companies.

C: Until now we have based out IPR on keeping our technology secret.

D: Our advantage is that we do not take input and directly give output, but give a general output (of accumulated information). So secrecy is the model we base our business on.

G: For protection we believe in secrecy.

The big companies (Google, Facebook,...) claim to go the other way and publish as much as possible, and that this is the only way to be ahead in the market.

The vulnerability of full disclosure was something at least one company was concerned with.

E: The consequence of someone having complete knowledge of algorithm increases the possibility of abuse. It is easier to do mischief if you know how something works in detail.

Some expressed concern with the protection of their data and results when they purchase a machine learning service, e.g. at a cloud provider, and use for development of your product or service.

> G: You can already buy a framework where you can train your data, and you own the data. There is the question about the MLS after you have trained it and how to protect it. Maybe secrecy is the best solution here?

> A: What is a "reasonable effort" to protect the trade secret from disclosure? Is a reasonable effort that you sign an NDA before you get access to the system?

5.3 Company ethics - do no evil

In addition we have also included company ethics in this chapter. Even if ethics is no protection mechanism it does have a significant impact on a company's vulnerability and potential for appropriation. Most countries have laws or court rulings on "fair business practice". This means a basis for legal actions against a company using data in a way the business community considers unfair or unethical.

Some ethical questions came up in the discussion of using data of unknown or unwanted origin, or collected by/of some entities.

> G: Some might have ethical considerations when purchasing products. What type of data is used in this MLS? Or the other way - where will I allow my data to be used?

> A: The customer might demand to not use a specific set of input data, due to potential political issues like "no Russian data", to "I know child labor has been used in those data".

> D: We are interested in finding legal guidelines and rulings on input data and referrals, in addition to whether it is OK to summarize and link to others' data.

And when using data you might have to have internal routines to handle potential unintended problems of intellectual property.

> A: Use of others' intellectual property is a potential problem that must be isolated.
>
> As a customer, depending on my area of business, I would have demanded some form of traceability of which input data had been used.

Some already had internal (unwritten) guidelines for what kind of data to use, and how to use it.

> G: If it feels wrong to use the image in a presentation externally, I would not use it in a product I am selling.
>
> If there is a license where everything is allowed I would not have any problems using it, but if the license states "only to be used for presentations", it would make me think twice.

Some asked questions on how others could collect their large databases without breaking any laws and whether it is ethical. Especially if it involves collecting from private citizens, even with their "consent".

> E: Tesla gathers data from their entire car fleet and use this to train the cars to perform their tasks better. Which rights have Tesla and which rights do the users have as passive "guinea pigs"?

And if this collective gathering of participants data really could be one company's or one group's property.

> E: The algorithm that Tesla in the end develops, do they really have all rights to this? They can try to secure themselves upon purchase, but... (How about reselling and optionally new owner?)

> G: There are huge challenges in getting enough (training) data. Constructing it yourself is one way of doing this. Many cooperate and share data with specific partners.

5.4 FTO and business area

All companies had plans for protecting their FTO and some had experienced loss of IPR.

> C: But we do want to protect ourselves with patents if possible. We just hired a guy with a master's degree in law on software patents.

> A: For our freedom-to-operate I would make sure that I had "kernel knowledge" that I knew was safe. And I would build other sections where I had traceability of used input data.

> F: We write that all content is protected by Norwegian Copyright Act and that systematic and regular extraction methods are not allowed without explicit consent. But it still gets done.
>
> At some occasions it has been troublesome that our database gets scraped of all data.

> C: We will have a contract with all customers where terms-of-use will forbid reverse engineering, redistribution of data, and things like that.

> E: The data, the big data. That is what is worth something. This is what Google and the others have realized.

Some companies also made a point of how challenging it can become to protect your own intellectual property. And that it will be hard to detect the use of your intellectual property in another company's machine learning system.

> E: You can protect algorithms and data structures. I do not believe that it will help you. You will not be able to prove that it (image) has been copied or used.

> C: We will not make any technical protection or detection actively.
>
> Giving away your data reduces your FTO. If the data is really not within your business area you have in reality not lost anything. You loose a potential opportunity.

Some also expressed hope for the use of patents in order to protect a company's FTO also in the field of machine learning.

> A: But there will no doubt be some real invention gems coming that are patentable, and the inventors will be sitting on those patents for 20 years.

Chapter 6

Results and Discussion

From the literature study and the interviews we identified many levels of potential IPR challenges for companies in the field of machine learning, and there were a limited amount of background resources available. Because of this we needed to use the interviews as well to locate more unanswered questions and identify potential challenges to IPR in machine learning.

This chapter will show and discuss our main findings from the literature study and from the interview rounds. We have moved some statements from the interviews into this chapter to better visualize and understand the results.

6.1 Limitations on freedom-to-operate

We will here look at the major types of intellectual property rights that usually are found on input data together with some of the related statements made by the companies.

6.1.1 Use of data

The intellectual property rights of input data are critical for a company's freedom-to-operate. The companies had some different views on this. The new entrants were not so restrictive in their views and discussed the regulations of making summaries and headlines of content pulled from Internet. While the established companies were more strict and would not use data they did not have explicit consent to use.

One analogy we used during the interviews addressed that if a human can read texts and view images from the Internet, can a machine learning system do the same and update its internal parameters? The image is not stored, but the internal model of patterns and structures are adjusted internally.

The law firm claims that every use of copyrighted material must be explicitly allowed by the copyright owner:

> B: But you have to ask the originator “may I use your IP this way?” It is you who reproduces it and makes it available to the users, and then you must have consent.

Faced with the challenge of having to relate to different types of intellectual property, there was advice of separating unconfirmed data from confirmed data, and they also had a plan for how to address this:

> A: We would have asked for a positive acknowledgement for using data where no license is present, or where the data is only presented as free/open data. If only one or two images are having problems (with licensing) there can be extensive problems just by this. If I build my system so I am not able to remove it, I have a problem. You need traceability, so when you are forced to remove two images from learning process, you are able to do this.

Interesting points were made about the use of open/free data in machine learning systems for something important, like products that might actually hurt someone or cause damage. At least you need to have control over the training, so it is not influenced in a matter that makes the system vulnerable and less secure. Not letting your system be manipulated like the retraining of Google Translate to give out false translations in order to embarrass your neighboring country (Sharkov, 2016).

Using millions of images

Maybe the most interesting view on the use of machine learning on image collections surfaced when one company made the analogy of the machine learning internals to be a mosaic of images.

> G: Somehow you can say you are building a mosaic of extremely many images. The more you add - the smaller the individual mosaic pieces.

But we argue that there is only a limited amount of storage in a machine learning system. The example of using e.g. one gigabyte equal to eight billion bits for internal storage of data, we also have to ask ourselves a question: What happens when you in this system use the huge databases and have sent more than eight billion images through the system? Meaning that there is less than one bit of storage per input image? Can you then (really) claim that the intellectual property owners of the input data still have claims to the output?

This can also be addressed as an IP-management problem of enormous dimensions. If the owner of the input data also owns the right to the output from the machine learning system, there is also the question of how do you solve the situation with multiple owners of the input data? While this may be solvable for small systems, new machine learning systems using billions of input IP with millions of content owners make this an impossible solution in the long run. Not even Bing's suggested markup language (Bing, 2004) could handle this complex situation.

> B: When you have a result out of learning from millions of images, you have the actual problem. You enter so large data sets that one person will not be able to say "That is my picture." Or potentially hundreds will claim significant similarity.

This is a likely problem of the future of machine learning unless laws and regulations handle these challenges in another way.

Proof of IP abuse

If your competitor launches a machine learning service and you suspect it has used your intellectual property to train the system. Can you detect this? Can you prove it? Detection of wrongful use of data is almost impossible as all the companies claim. Especially since the burden of proof is on the copyright owner.

> B: If someone claims that you have used his IP, he has the burden of proof. He must prove that it is more than 50% likely that you have used his IP without consent.

Some might use this as a free card for using anyone's IP in their machine learning systems, but with the potential infringement compensation on millions of images, this will likely make all serious companies attempt to have full control over their use of external IP.

To prove abuse of your IP, you most likely need to have complete access to the machine learning system's internal parameters and perform lots of manual testing. This becomes quite similar to expert witnesses called in to evaluate abuse of code in other's software systems (Hollar, 2004).

There might be ways in the future to assist in locating abuse of your IP. Watermarking of input was suggested, but others claimed that this would reduce the value of the IP when used and would not purchase watermarked data.

6.1.2 Internal IP management

All of the companies using images for machine learning in product/service development or in their research, purchased these images or developed them internally. One company assumed more rights than the others, like automatic creation of summaries as fair use of data, and another acknowledged that users may put copyrighted images into their system even if they agreed not to.

This will make a difference for using the images in a machine learning system. If you do not have complete control over your input, you are opening up for potential law suits.

> B: The main challenge is to make sure that you have control of the IPR on the training data. Because there will be infringements there. Someone will claim "Hallo! This is mine!". Most likely an author, composer or photographer.

And the law firm was very clear on who were responsible for the data in the machine learning system.

B: The company with the most responsibility is the one entering the data sets (training of system). This is where you add/copy others' intellectual property.

6.1.3 Creating new content

The creation of new content from machine learning systems generated a lot of discussion and interest. Most were interested in who really have intellectual property rights to the output. There are many contributors to the new material in a product or service that is able to influence the creation of new content.

Programmer of machine learning system If you have implemented the system that enters the data and identified which features the system shall use when creating the new content, could you claim IPR to the new content?

Computer itself A computer cannot own intellectual property in Norway today. Could we see this change soon?

System owner So even if a computer itself cannot own IP, maybe the system owner that has bought the computers where the MLS is running and all the data is stored should own some IP in the new content?

Owner of training data If you are the owner of the input data that have contributed to the structures and patterns in the system, could you claim IPR to the new content?

System/service user If the user specifies what it wants to have created, and the system creates something that resembles this, is this effort enough to own some IP in the new content? It would not have been created in this way without the user's specification.

So with all these seemingly open question we include statements from the law firm:

B: Even if I am an artist and create B from A, adding something new, change and dependence on A, adaptations, is covered by copyright (opphavsrett) and you need consent.

> B: But if you can recognize an (input) image from the created image, then it will be an adaptation and you need consent from the IPR owner.

And clarifying that using external material is in need of consent, and that the programmer or her/his employer will have copyright to the code but no IP to the new content.

> B: From multiple input pictures/IPR then you get an actual problem. Basic question is: Who has created the image? The programmer made the software. The software code is his copyright. She/he did not build the new image. (The image is the result of the images used in training.) So the programmer will not have any rights to the new image.

> B: Before you entered the (training) data set you did not know this would be the result. Those that made the software will therefore not get any rights to the created images.

The computer cannot own IP. Owning a camera does not automatically give you rights to the pictures taken. It is the photographer that creates the pictures and can claim IP. So the system owner does not own any IP in the new content:

> B: What you showed was highly interesting because it shows that it seems like the computer actually creates new images. But the computer does not get any rights.

Since the software in the machine learning system is static, and all that affects the content generation is the training of the system and the user's interaction, these are the two possible sources of creativity.

But can the user's effort be deemed as creative? If the users only select from predefined lists of options, the answer is most likely no. This is quite similar to generating new content by not being creative (enough):

> B: When you press a button and create random noise, have you then made a creative effort? EU and Norwegian law say no.

So if you are a non-creative user we end up having to own the right to use the images in your machine learning system to control the intellectual property created by the system.

An interesting view point on what kind of uniqueness that must be present in order to demonstrate creativity is the principle of double creation.

B: For music generation there is the principle of double creations. If I and my brother asked for music of our preference from a system, we would likely end up with quite similar, maybe identical results. So creativity must be over a threshold to be an intellectual property.

Another interesting development is that the first movers in the market, able to create profitable valuable content and keep intellectual property might get a significant head start and may dominate the entire market with their productions (Peng, 2013, p.163). It is interesting to see that this kind of development was predicted already in 1953 in Roald Dahl's "The Great Automatic Grammatizator", where they in the end reveal that half of the world's new English texts are created by the machine (Dahl, 1953).

But we also have to look at the possibility of user creativity. What if we find users and types of use that can be defined as creative?

User creativity

As described in Chapter 5, many companies looked at computers as a tool used to create new content. Why not in generation of new content from machine learning systems? In music production computer tools have long been accepted, and someone takes credit even if music production can be made quite easy (Nierhaus, 2009; Brown, 2015).

B: With music the composer gets all rights. She/he must write down the composition and deliver it and claim "this is my song". If a computer auto-generates a song anyone will be able to claim rights. But will anyone listen to it?

But this only breaks down to the user taking credit for using a system in a specific way. So maybe a user of the machine learning system also can be seen as creative?

> B: In order for the user (that is creating new works from MLS) is able to create something new, she/he must use creativity and make a lot of choices so that it is defined as an original work.

We introduced the randomness of a trained system in Chapter 2.1.1. As a consequence of this it may be necessary for the user to have the complete understanding in order to be able to fine tune and correctly set the thousands of parameters that may be necessary to get exactly "this image". Though it also will be possibilities to only set a few and still get an image, we think this has the same implications as the user of a music composing tool. And thereby the user should have the same rights.

There is still the line defining when something is creative enough, but as quoted above, will anyone listen to it? And will it then actually be worth anything? The same principle should apply to image generation as well.

If the new content is not good enough or usable, why should you even want to have intellectual property to it? Especially since new images of equal quality can be generated with ease.

Computer generated patents

Companies are already trying to block patent applications on all new inventions by publishing "all possible future ideas" from a machine learning system trained on all prior patent applications and inventions (All Prior Art, 2016). This is not creative enough to be deemed as prior art, and also not a manageable index of publishing.

A more useful case is that patent lawyers are using machine learning systems to assist in finding the most probable outcome of any patent dispute (Harbert, Tam, 2013). This already has quite an impact on a huge market.

Computers with creativity will be an issue of the future (Kohlhepp, 2008), but the discussion of creativity will inevitably come. The claims of computer generated patent applications are discussed in articles and publications already (Hattenbach and Glucoft, 2015), and remain an open question to current regulation.

The problem of computer generated patents is a lot bigger. We can easily see today's patent system get swamped by patent applications trying to either maintain a company's freedom-to-operate, or try to block a competitor's progress. Almost like patent trolls are doing today (Plumer, 2011). These computer generated patent applications are most likely profitable if they are granted. And even if they are computer generated and only some tiny percentage of them are granted, just having claimed them might even be worth it. But the big question then becomes: Should we scale up a patent system already struggling to respond fast enough? And in addition the flood of new applications will make each application take even longer.

Computer generated inventions

A computer cannot have intellectual property by today's laws. But the idea of computers making inventions on their own is not far off (Fraser, 2016), as is neither the idea of creative computers including the question of users vs. owners vs. developers (Abbott, 2016).

Recalling Adam, the Robot Scientist (King, 2011), and the claim for novelty in results, this will only become a bigger topic of discussion in the future.

Abbott writes a hypothetical but quite realistic and up-to-date story about the development of new vaccines being made by computers (Abbott, 2015). This is a new generation of Adam, where the company is also aiding patent applications from the software, called Hal. This system can design drug trials, run clinical simulations, search for new use of existing drugs, and assist writing patent applications. In the paper the software developers do not qualify for IP to the new drugs, as they only developed a tool. The paper claims that the employees that guided Hal to test that specific drug, may meet inventorship criteria if Hal worked under the supervision of them. Meaning that Hal can not claim intellectual property to the results, but that the users of the system might.

This could have implications on how we can look at results from machine learning systems in general.

6.1.4 Reverse engineering

We have entered a time where many new types of engineering can be automated and results analyzed in far more detail. Machine learning adds a completely new level to the possibilities of reverse engineering. Combined with current robotics technology this can automate the reverse engineering of new types of products. Not only reverse engineering of software, but reverse engineering of electronic black boxes, production processes, chemical reactions, etc. (Nosengo, 2016)

B: In the entire world there are rules for reverse engineering. For computer programs it is allowed to look at the compiled code and try to guess what the source code is. In Norway, you are allowed to reverse engineer.

There will be more reverse engineering with more powerful tools. It will likely be OK to extract the machine learning model by external query of the machine learning system.

Will the huge advantages in using computers and machine learning systems in the process of reverse engineering challenge today's regulations?

As explained earlier it is possible to reverse engineer an existing machine learning system, or at least build your own version closely resembling it purely by having access to use the other system for some time.

Let us visualize this by using a machine learning system to learn how to play chess, and play it well. We do not know how to make an intelligent chess playing program, but we will train our machine learning system by playing against state of the art chess software. So we purchase one license of e.g. the Play Magnus chess app, and automate the process of playing over and over against this application and learn from this. After a while we will have created our own chess playing software that could be quite sophisticated, but all it has learned is coming from playing against the Play Magnus application. With today's regulation we have most likely not broken any, but it would probably not be popular with the owner of the application used for training.

Then we can ask the following questions. Is this really any different than having a machine learning system learn how to build images from looking at (or browsing, or reading, ...) existing images from Internet? Can we say that one type of use is allowed and another is not?

Protection against reverse engineering

Several of the companies already felt threatened by the possibility of having their products and services reverse engineered.

Some suggested to use copyright to protect themselves from reverse engineering.

C: To protect themselves from reverse engineering they need to use copyright or patents.

And some wanted to fight this with user agreements and terms-of-use, but this is especially hard in Norway.

> B: A user agreement cannot take away the right to reverse engineering. By Norwegian law you are allowed to do this and it cannot be waived by an agreement.

Some more creative ways to change the rules of reverse engineering came when we looked at the potential difference between a computer aided reverse engineering and a human-performed reverse engineering. One company proposed to only allow manual reverse engineering as a potential way to counter the potential in automation of reverse engineering.

> E: Can you change the law and say that reverse engineering is allowed, but not with computer aid? It is allowed done as a human, but undefined how you are allowed to perform the reverse engineering?

But with today's regulation explicitly allowing reverse engineering, we claim that many will have problems protecting their products and services from reverse engineering. Especially if the products or services can be accessed and used by an automated process.

6.2 IPR suitable for machine learning

We will start our discussion of suitable IPR by quoting the law firm:

> B: I do believe that with copyright, patents and other IPR they (law makers) do not need to make many changes. The rules of today are quite technology neutral.

And they also pointed out why copyright was chosen to protect software source code.

> B: Copyright was the chosen method to protect software (source code) as it enables changes and adding more code. Patents is more "shutting others out of the market".
>
> But I can make a description of how the software works and get a patent on that.

Building on these views we will have a closer look at how patents, copyright and trade secrets are used in the field of machine learning.

6.2.1 Patents

Patents have always been an advanced topic of discussion. There are so many dimensions and areas to discuss as shown in the interview chapter. We will have to focus on the areas related to protecting machine learning systems and this seems to be quite hard.

B: I do not believe that patents will be important for machine learning. You can only use another method to create this. It might be as simple as that.

C: Machine learning does not make patenting easy as there are many ways to reach your goal.

So trying to get a patent for your machine learning system may not be a good way of securing your company's appropriability. First of all because it seems easy to avoid existing patents, but also because they are hard and often resource demanding to get approved.

There is another important dimension involved - time. It takes several years to get a patent application approved and in the meantime anything can have happened in this highly dynamic market. Maybe most patent applications are not even useful anymore now that "all" problems can be solved with the aid from machine learning and artificial intelligence? If it becomes near to impossible to create novel patent applications and most innovation come in increments, then we might have a bigger challenge in approving these new patent applications.

We hope that an evaluation of the validity length of patents, amount of work involved in "approving", or having "approved light" or a simpler type of registration, could be topic for future discussions. Especially whether or not the patent system of today really is working as intended, now when we enter a time of computer generated patent applications and inventions.

6.2.2 Copyright

B: Copyright is only a result of an intellectual effort. That is why we (in Norway) have the condition for original work ("verkshøyde").

The challenges of copyright varies from country to country, but we will keep to the Norwegian challenges. And we also got confirmation of the need for creativity:

> B: Writing two sentences of text – you do not get copyright from that. It must be an effort worth protecting.
>
> If you translate from Norwegian to English which has five times more words, you can make a huge amount of choices. And if you choose one way you create something new.
>
> If a computer translates it does not *create* anything.

What is challenging in machine learning is the difference between learning from viewing and reading data into a machine learning system.

Learning from viewing What all humans do when looking at art, reading books, listening to music, etc. This is OK and you can use that knowledge later when you try to be creative, as long as we leave forgery out of the question. You can write like Shakespeare did, you can paint like Rembrandt, but it might give you less credit than being completely original. It is still legal to try to use your knowledge to make your creations better.

Reading data into an MLS This is viewed upon as something different. Even if the machine learning system cannot reproduce your image, and has only enhanced its "perception" (internal parameters) of patterns and structures, the data must first be read into the system. And this is the problematic part according to company B.

> B: But even if (the data is) only read once, it makes a copy. Copyright protects copying. For the data sets to be used there must take place a copying. That makes copyright important.

And from the other companies we have the analogy of storing a mosaic of all the images in the machine learning system. We do not agree with this analogy as:

1. there is a limited amount of internal storage, no matter how many images you use in the training process, and
2. you cannot reproduce the original image unless you construct your system in that way and explicitly store all necessary parameters, which will be copying.

We are not trying to dispute the problem of having to copy the data to be able to read, but this is a technical definition. You have still not published the data or a new image in any way, and the copy is deleted upon parsing. And we still think that letting a computer read and parse input data with or without intellectual property is something that must be discussed more thoroughly. Especially since the system will be unable to reproduce the input and only use input data to adjust its internal parameters. Some of this was discussed in Chapter 6.1.3.

One of the companies visualized the value of having copyright with an example from of non-human production of art:

> B: In Sweden a few years back an unknown artist sold paintings for 100.000 kronor. They were created in a totally new way. Then it was revealed that a monkey made the paintings and the value went to zero.

Another copyright story is about the selfies[1] taken by a monkey. The photographer from whom the monkey borrowed the camera claimed copyright, but the images were ruled to belong in the public domain, free for all to use (Domonoske, 2017).

> B: The monkey took the camera and used it on himself. The photographer owned the camera, but you do not get any rights from owning a camera. You must take the picture. Copyright is not to protect the trivial. Like taking a picture by pressing a remote. You do not get copyright from that. It must be a creative effort.

How can copyright in the age of machine learning be solved? Maybe it will be like we discussed under "new content" in Chapter 6.1.3. The law firm did have views on what could happen from their point of view:

> B: What I believe will happen is that Google or other huge companies will make user agreements for using their code and their data sets.

It is highly likely that many will purchase their licenses from a company that contributes with both the machine learning system, the machine learning model (internal parameters), computing resources, databases of training data, and software binding all this together. But

[1] Selfie: A self portrait taken with camera in hand.

this will not work for all companies, so there must be regulations addressing all types of use for this to work for the entire industry.

And abuse of output is likely to happen, as it always have:

> B: And if the user creates something fantastic like a new image, she/he will save it and claim copyright, even if it is not her/his image.

6.2.3 Trade Secrets

Even if all companies were using secrecy as their main method of protection the lawyers added the concept of good business practices, which is protected by law in Norway.

> B: Trade secrets are extremely important. We also have laws for similarities that are not protected from design. This is breaking the laws of good business practices.

But the most relevant question about trade secrets is - does it work for machine learning technologies? The major problem is described earlier through the possibility of reverse engineering the system. Meaning that some machine learning systems may not be protected by trade secret legislation as there is an opening for extracting information through use of the systems, protected by the right to reverse engineering.

A trade secret must be given a “reasonable effort” in protection. When new technologies like automation combined with reverse engineering can be used, this moves the borders of what should be considered a “reasonable effort”. And many companies may not be aware of these potential security challenges.

These issues are a threat to the companies’ appropriability by preventing effective protection of intellectual property.

6.2.4 Company ethics - do no evil

Some companies may have internal ethical guidelines that indicates that all use of unknown images are risky and should be avoided. Or they have guidelines of how the source of origin for the images can be acceptable. This is risk management through ethical guidelines.

But we also entered the discussion around copyright through the creation of new images. Should some types of machine learning systems be banned or avoided simply because they *can* be used to break copyright?

> D: I think what it all comes to now, is how close output is to input. Even if the MLS can reproduce the work, it does not mean that the MLS must be legally enforced to avoid some types of use.

We think not, and will look at this as censorship of technology.

None of the companies expressed opinions in censoring some type of systems or functionality, but we suspect that this is not the last time this type of censorship is discussed.

Another ethical discussion took place in the collection of data from users or the public, and what must be in place for these data to be allowed to be used in a machine learning system. E.g. location data of smart phone users are already being collected in the billions every day, and assist people driving cars to avoid congestions, choose the optimal paths, etc. People might understand collection of location data, but what if an insurance company collects location data on you continuously? Will the insurance company use the fact that you moved faster than the speed limit on multiple occasions to raise the price of your car insurance? How can the end users actually understand how their data can be used? These types of ethical questions should be understood by the users, prior to the start of collection.

In addition some companies should have ethical guidelines not only on how they use their data, but how they share them with others. But this is outside the scope of this paper.

6.3 FTO and business area

All companies want to protect their intellectual property. There was no difference in this when comparing the business areas. But for the new entrants it was more important to protect their systems and algorithms, and for the established companies the focus of protection were primarily on their data.

The companies were also clear about the importance of protecting the data, and that it is hard to detect whether someone has used your data in another machine learning system. And they were all open for discussing if and how a computer can create copyrighted material.

G: I think the argument of no copyright when created by a computer is strange. Most of what we produce today is the result of computer assistance.

We found that the new entrants were more willing to push the borders of what data you can use in your machine learning system and how intellectual property of the result should be granted. E.g. by defending Google's rights to own output of new content generation.

6.4 Other views

We also collected some other views and questions that were not specific for our research, but should be part of our results to see the width of the challenges discussed during the interviews.

6.4.1 Human rules and computer rules

Most companies were interested in the discussion of humans learning vs. computers learning. Like looking into if there is a real difference between a human learning process and a machine learning process, and how this can be viewed in the future. Can science and business live with a conclusion of having different rules for humans and intelligent computer systems?

G: There is a difference in what you have published for viewing by humans and thereby not been copied and stored, compared to what a machine can do.

6.4.2 Update of software without humans making the code

If you get software itself to update its own code. Making this update solely based on the machine learning system and what the MLS has learned from its sensors, with no humans involved; who will own the new source code? What about the new data in the machine learning system?

E: Can software update itself and learn from (recorded) experiences without humans actively in the loop?

If code evaluates itself based on input data and changes itself to better interpret the input data and perform its tasks, who is to separate this from humans updating code? And what happens to the new source code copyright?

6.4.3 Manipulation of machine learning system

How can you protect your machine learning system from being manipulated?

E: Influencing a machine learning system is possible. Ukrainian opposition trained Google Translate from Russian to English about Putin and discovered how to make the translation to their version. (Sharkov, 2016)

Automation and manipulating systems are as old as the systems themselves and the protection mechanisms trying to prevent manipulation.

E: I think it is quite impossible to protect the machine learning algorithm.

6.4.4 No action until blame is to be given

Maybe updating the regulations will get more priority when something bad happens and someone gets hurt? Or what if you automatically extracted news, or new content consists of illegal material, or perhaps points to illegal material?

E: Who is to blame if something goes wrong will be the essential question. If you claim ownership to the algorithm you must have some kind of responsibility.

C: I do not believe this (generating art) will be a problem legislators will even look at until it affects someone for real. It is like autonomous cars. Who is responsible is someone is killed? Driver? Car maker? Artificial Intelligence makers? No one?

Chapter 7

Conclusion

Our research has made a contribution towards understanding the challenges that new machine learning technologies can put on our current understanding and use of intellectual property rights. It is essential to address these potential problems early on so that companies can be aware of them and take precautions. In addition it is important for law makers to see if the existing regulation will be effective also with the new technology evolution taking place.

7.1 Use of data

We have discussed the never ending topic of difference between humans learning and computers learning. If humans can learn from viewing images, texts and videos, and be accredited with intellectual property of the new content; why cannot a machine learning system be interpreted to “learn” from input and “own” the output in the same way?

The main reason for this is that today’s regulation defines the act of reading data into a computer as an act of copying, and thereby a breach of copyright. Our research indicates that there are alternating viewpoints to whether copyrighted data as input to machine learning systems should be viewed differently than copying the data for storage/reproduction.

We also point out that even if today’s regulation does **not** allow a company to build their new machine learning products/services by using data with copyright, it is very hard to detect such abuse after the product/service is developed. And this burden of proof will lie on the copyright owner suspecting abuse.

With the huge databases of so-called free data that exist today each with millions of entries, there are bound to be mistakes and some copyrighted data added by accident. By building your products with uncertainty in the origin of data, you expose your company to potential future law suits. With today's regulation a company should be able to remove the influence of specific input data, if this becomes possible.

7.2 Reverse engineering and upcoming challenges

Our contribution has been to explain and visualize the impact that reverse engineering can have on existing products and services, when combined with the newest machine learning technology and with today's potential of automation. We claim that a significant amount of existing products and services will get problems simply because they let their customers use the products or services. This access, physically or logically (through computer control), combined with machine learning systems and automatic adjustments of the input, will open up for reverse engineering of products in ways not possible until now. This creates a great deal of uncertainty for existing companies and their options for protecting their products and services, and will reduce a company's freedom-to-operate and appropriability.

In addition we found that machine learning companies are especially vulnerable to the same type of challenges, and that terms-of-use or other types of user agreements and contracts cannot take away the right to reverse engineer (at least not in Norway).

From this follows the questions of how the technology will challenge today's intellectual property regulation and methods of protection. One company proposed the interesting idea of making a distinction between human reverse engineering and computer aided reverse engineering.

7.3 Creating new content

Another contribution is our discussion regarding the creation of new content from machine learning systems. Even if the quality of the production demonstrates that the technology is in its infancy, we have discussed the question of who will have intellectual property rights in the new content.

In our research we demonstrate how machine learning systems can create new valuable content from its internal data of patterns and structures, and how this is different from copying content. One important argument is that a machine learning system only has a limited internal size for storage of patterns and structures. As an example we looked at a case where you put more than eight billion images into a machine learning system with one gigabyte of internal storage, meaning that less than a bit of information potentially could represent the input of each image. Even when you add more images and adjust the internal patterns and structures parameters, it cannot be said to store a copy the image.

Neither the programmer, the computer itself, nor the system owner will have any claims in the new content. But from today's regulation the owner of the input data is essential and will be able to claim ownership, making licensing input data the most important part for your freedom-to-operate. From the last years' development we predict that the user of the machine learning system through her/his analysis of the internal patterns and structures, and by extracting the new content may have claims depending of the definition of user creativity and novelty of new content.

7.4 Research questions

By looking at the sections above "Use of data", "Creating new content", and "Reverse engineering", we have answered our research questions throughout our discussions. To briefly summarize:

FTO in MLS We discuss the ownership of input data and internal models, where control of input data is essential by todays regulations. We address the conflict of creative users and input data owners when it comes to IP in new content. And we find that most will base their protection on trade secrets, but that this might be challenged by the new opportunities through reverse engineering, automation and machine learning.

Suitable IPR Trade secrets are important but we explain why they may be vulnerable to reverse engineering together with automation and machine learning. We emphasize that you need to gain copyright to created content. And we believe patenting machine learning technology is hard to accomplish.

FTO and business areas All business areas face the same challenges and they are aware of this with patents, input data and trade secrets. We found that established companies may have more focus on keeping their data from being used without permission, while new entrants may have more focus on retrieving and using data without breaking any laws.

7.5 Recommendations for Further Work

We would like to see the research area continued into several directions.

- The actual legal implications and challenges to both the Norwegian and other legal systems when having computer systems creating content like images, texts, art, computer software, improving themselves, etc. And the legal ramifications this has on our current regulations of intellectual property protection mechanisms.

- Research of freedom-to-operate for companies producing new content, and the direct challenges to intellectual property of that new content. Including research specific for defining the creative user vs. creative software, and if/how/when non-humans can have claims to the output.

- More research into the future of intellectual property while looking at reverse engineering in conjunction with machine learning systems and automation. What happens when a robotic scientist creates something new every day, and it was the robotic scientist that built the hypothesis, constructed the experiment, put together the lab, ran the experiment, collected the data, analyzed the results, concluded that there were commercial value in the result, and completed by submitting a self generated patent application?

Appendix A

Interview guide

Through the introduction of the topics in the interviews there were many discussions and exchanges of view points on the subjects. Since these discussions were unpredictable we had generated a list of check points that could be addressed if not discussed earlier in the interview. Some items in the list were not applicable for all businesses, and the last points (10–16) were added specifically for the interview with the law firm.

The initial guide was produced and used in Norwegian as all interviews were conducted with Norwegian companies, but we attach a translated version here.

1. Are you using open/free/external data in your IP systems? Do you check licenses? How?
2. How do you view the use of other's IP as input to an MLS? Why?
3. What controls your company's FTO today? In the future?
4. How can IPR be improved for MLS?
5. How can patent protection work for MLS? Future change? Wny (not)?
6. How do you look at trade secrets, copyright, design protection?
7. Can you protect your company's IP from being used in MLS? Can you detect abuse or recognize unauthorized use?
8. Do you want to protect yout IP from MLS use? (Why?)
9. Does your company have a strategy for protection of new content from MLS? (How?)

10. Will IPR-owner have claims in a machine learning system trained on their data? Having IPR on all data used to train the system? What if someone has IPR on only one of one million input images?

11. Challenges when unable to detect use of specific image(s) in MLS. Must product developer prove non-use?

12. Will computer/system owners have claims? Regulated by agreement?

13. Will programmers (creator of learning model) have claims? Is model just a tool? Is code just a tool? Is output derivative work of code? Where is programmer's expression in output?

14. Will end users (asking for specific image) have claims? System is only a tool for end user? "it creates whatever I ask for..."

15. What kind of IPR can be applied on new machine learning results/products? (does existing IPR mechanisms deal with this new technology?) Will it be similar to patents on software? Can patents be effective here? Copyright? Design? Secrecy?

16. Protection against reverse engineering of machine learning model? Made reasonable effort to protect against reverse engineering?

Bibliography

Abbott, R. (2015). Hal the Innovator: Computational Invention and Its Patentabilty Implications. *Big Data Is Not a Monolith, MIT Press (Hamid Ekbia, et al. eds.)(Forthcoming)*.

Abbott, R. (2016). I Think, Therefore I Invent: Creative Computers and the Future of Patent Law. *Boston College Law Review*, 57(4).

Achenbach, J. (2015). Some scientists fear superintelligent machines could pose a threat to humanity. http://www.washingtonpost.com/sf/national/2015/12/27/aianxiety/. Last accessed 4 January 2017.

Ahuja, G., Lampert, C. M., and Novelli, E. (2013). The Second Face of Appropriability: Generative Appropriability and Its Determinants. *Academy of Management Review*, 38(2):248–269.

All Prior Art (2016). About All Prior Art. http://allpriorart.com/about/. Last accessed 4 January 2017.

Bing, J. (2004). Copymarks: a suggestion for simple management of copyrighted material. *International Review of Law, Computers & Technology*, 18(3):347–374.

Bloom, J. (2015). Asking RNNs+LTSMs: What Would Mozart Write? http://www.wise.io/tech/asking-rnn-and-ltsm-what-would-mozart-write. Last accessed 27 January 2017.

Borel, E. (1913). La mecanique statique et l'irreversibilite. *Journal de Physique Theorique et Appliquee*, 3(1):189–196.

Borella, M., Sheppard, C., Kesan, J., Remus, M., and Pedraza-Fariña, L. (2015). Business Method Patents Following Alice v. CLS Bank.

Bower, J. L. and Christensen, C. M. (1995). Disruptive Technologies: Catching the Wave. *Harvard Business Review*, January - February.

Boyce, C. and Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input*. Pathfinder International Watertown, MA.

Bradford, A., 23, L. S. C. |. M., and ET, . p. (2015). Deductive Reasoning vs. Inductive Reasoning. http://www.livescience.com/21569-deduction-vs-induction.html. Last accessed 8 January 2017.

Brooks, R. (2014). Computer Algorithm Generates Poetry As Good As Shakespeare's. http://www.psfk.com/2014/01/shakespeare-machine-learning-poetry-app.html. Last accessed 13 January 2017.

Brown, M. (2015). World's first computer-generated musical to debut in London. *The Guardian*.

Burgess, M. (2016). Google's AI has written some amazingly mournful poetry. http://www.wired.co.uk/article/google-artificial-intelligence-poetry. Last accessed 13 January 2017.

Byrnes, N. (2016). AI's Very Disruptive Time. https://www.technologyreview.com/s/600994/ais-very-disruptive-time/. Last accessed 10 January 2017.

Cave, K. (2016). IDG Connect – The next wave of disruption: Graph-based machine learning. http://www.idgconnect.com/abstract/18124/the-wave-disruption-graph-machine-learning. Last accessed 10 January 2017.

Chatfield, T. (2016). How much should we fear the rise of artificial intelligence? *The Guardian*.

Christensen, C. M. (2016). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press, reprint edition edition.

Clark, T. (2016). Forbes SAPVoice: Machine Learning Examples That Will Amaze You. http://www.forbes.com/sites/sap/2016/10/31/machine-learning-examples-that-will-amaze-you/. Last accessed 10 January 2017.

Clegg, S. R., Kornberger, M., and Pitsis, T. S. (2012). *Managing and Organizations: An Introduction to Theory and Practice*. SAGE Publications Ltd, Thousand Oaks, CA, 3 edition edition.

Cortez, N. (2014). Regulating Disruptive Innovation. *Berkeley Technology Law Journal*, 29(1).

Dahl, R. (1953). *Someone Like You*. Knopf, first edition edition edition.

David Lizerbram & Associates (2015). Can a Computer Own a Copyright? http://lizerbramlaw.com/2015/03/10/copyright-question-can-computer-create-copyrightable-works/. Last accessed 13 December 2016.

Davies, C. R. (2011). An evolutionary step in intellectual property rights – Artificial intelligence and intellectual property. *Computer Law & Security Review*, 27(6):601–619.

Domonoske, C. (2017). Monkey Can't Own Copyright To His Selfie, Federal Judge Says. Last accessed 30 January 2017.

Edwards, C. (2014). From lab rats to robots. *Engineering Technology*, 9(11):76–79.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, advance online publication.

EU Commission (2013). Study on Trade Secrets and Confidential Business Information in the Internal Market. http://ec.europa.eu/internal_market/iprenforcement/docs/trade-secrets/130711_final-study_en.pdf. Last accessed 14 January 2017.

Favale, M., McDonald, N., Faily, S., and Gatzidis, C. (2016). Human Aspects of Digital Rights Management: the Perspective of Content Developers. *SCRIPTed*, 13(3):289–304.

Foster, R. N. (1986). *Innovation: The Attacker's Advantage*. Summit Books, New York.

Fraser, E. (2016). Computers as Inventors – Legal and Policy Implications of Artificial Intelligence on Patent Law. *SCRIPTed*, 13(3):305–333.

Galler, B. A. and Brown, J. E. (1995). *Software and Intellectual Property Protection: Copyright and Patent Issues for Computer and Legal Professionals*. Quorum Books,U.S., Westport, Conn.

Gioia, D. A., Corley, K. G., and Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, 16(1):15–31.

Glaser, B. G. and Strauss, A. L. (2009). *The discovery of grounded theory: strategies for qualitative research*. Aldine, New Brunswick, 4. paperback printing edition. OCLC: 553535517.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., and Wierstra, D. (2015). DRAW: A recurrent neural network for image generation. *arXiv preprint arXiv:1502.04623*.

Harbert, Tam (2013). Supercharging Patent Lawyers With AI. http://spectrum.ieee.org/geek-life/profiles/supercharging-patent-lawyers-with-ai. Last accessed 5 October 2016.

Hattenbach, B. and Glucoft, J. (2015). Patents in an Era of Infinite Monkeys and Artificial Intelligence. *Stan. Tech. L. Rev.*, 19:32.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv:1502.01852 [cs]*. arXiv: 1502.01852.

Hill, C. W. L. and Jones, G. R. (2012). *Strategic Management: An Integrated Approach*. South-Western College Pub, Mason, OH, 10 edition edition.

Hinton, G. E., Krizhevsky, A., Sutskever, I., and Srivastva, N. (2014). System and method for addressing overfitting in a neural network. International Classification G06N3/08, G06N3/04; Cooperative Classification G06N3/084, G06N3/082, G06N3/0472, G06K9/4628, G06N3/0454.

Hollar, L. A. (2004). Requesting and Examining Computer Source Code. *Expert Evidence Report*, 4(9):238.

Howard, R. A. and Korver, C. D. (2008). *Ethics for the Real World: Creating a Personal Code to Guide Decisions in Work and Life*. Harvard Business Review Press, Boston, Mass, 1 edition edition.

Hsu, F.-H. (1999). IBM's Deep Blue chess grandmaster chips. *IEEE Micro*, 19(2):70–81.

Jacobsen, S. (2015). Former Yara executives sentenced to prison in corruption case. *Reuters*.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual review of psychology*, 50(1):109–135.

Joshi, B. (2016). 2001: A Picasso Odyssey. https://vimeo.com/169187915. Last accessed 6 December 2016.

Kahneman, D. (2013). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, 1st edition edition.

Kelley, K. (2014). The Three Breakthroughs That Have Finally Unleashed AI on the World. https://www.wired.com/2014/10/future-of-artificial-intelligence/. Last accessed 4 January 2017.

King, R. D. (2011). Rise of the robo scientists. *Scientific American*, 304(1):72–77.

King, R. D., Rowland, J., Aubrey, W., Liakata, M., Markham, M., Soldatova, L. N., Whelan, K. E., Clare, A., Young, M., Sparkes, A., Oliver, S. G., and Pir, P. (2009). The Robot Scientist Adam. *Computer*, 42(8):46–54.

Kohlhepp, P. M. (2008). When the Invention Is an Inventor: Revitalizing Patentable Subject Matter to Exclude Unpredictable Processes. *Minn. L. Rev.*, 93:779.

Kostelanetz, R. (1971). Machine art. *The Centennial Review*, 15(3):229–249.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lie, H. T. (2016). Introduction to Current Trade Secret Management Research. SSRN Scholarly Paper ID 2869296, Social Science Research Network, Rochester, NY.

McCorduck, P. (2004). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. A K Peters/CRC Press, Natick, Mass, 2 edition edition.

Mordvintsev, A., Olah, C., and Tyka, M. (2015). Inceptionism: Going Deeper into Neural Networks. https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html. Last accessed 27 January 2017.

Nierhaus, G. (2009). *Algorithmic Composition: Paradigms of Automated Music Generation.* Springer Science & Business Media. Google-Books-ID: jaowAtnXsDQC.

Nosengo, N. (2016). Can artificial intelligence create the next wonder material? *Nature News*, 533(7601):22.

Noyes, K. (2016). How do you stop patent trolls? This algorithm just might do the trick. http://www.pcworld.com/article/3060106/how-do-you-stop-patent-trolls-this-algorithm-just-might-do-the-trick.html. Last accessed 4 January 2017.

O'Brien, T. L. (2003). Norwegian Hacker, 19, Is Acquitted in DVD Piracy Case. *The New York Times*.

Panetta, C. (2016). Gartner's Top 10 Strategic Technology Trends for 2017. http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/. Last accessed 10 January 2017.

Peng, M. W. (2013). *Global Strategy*. South-Western College Pub, Mason, Ohio, 3 edition edition.

Plumer, B. (2011). Innovation's down, but patent trolls are thriving. https://www.washingtonpost.com/blogs/ezra-klein/post/innovations-down-but-at-least-patent-trolls-are-thriving/2011/09/21/gIQABGdKlK_blog.html. Last accessed 4 January 2017.

QENTIS (2015). QENTIS. http://web.archive.org/web/20150915155038/http://www.qentis.com/. Last accessed 13 October 2016.

Rognli, E. (2007). Ingen Wintermute i sikte. Et intervju med Jon Bing. *Pegasus*, pages 24–25.

Rutkin, A. (2016). Computer generates all possible ideas to beat patent trolls. https://www.newscientist.com/article/2084755-computer-generates-all-possible-ideas-to-beat-patent-trolls/. Last accessed 4 January 2017.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.*, 3(3):210–229.

Sayej, N. (2016). Vincent van Bot: the robots turning their hand to art. *The Guardian*.

Schilling, M. A. (2012). *Strategic Management of Technological Innovation.* McGraw-Hill Education, New York, NY, 4th edition edition.

Serra, M.-H. (1993). Stochastic Composition and Stochastic Timbre: GENDY3 by Iannis Xenakis. *Perspectives of New Music*, 31(1):236.

Sharkov, D. (2016). Google Translate Glitch Turns Russia Into Mordor. http://europe.newsweek.com/ukraine-google-translate-russia-mordor-411723. Last accessed 25 January 2017.

Stead, D. R. and Cross, A. R. (2009). The management and security of trade secrets: an exploratory study. *International Journal of Intellectual Property Management*, 3(3):256–277.

Stern, P. N. (1980). Grounded Theory Methodology: Its Uses and Processes. *Image*, 12(1):20–23.

Sturm, B. L. (2016). "Lisl's Stis": Recurrent Neural Networks for Folk Music Generation. https://highnoongmt.wordpress.com/2015/05/22/lisls-stis-recurrent-neural-networks-for-folk-music-generation/. Last accessed 27 January 2017.

Sutskever, I., Martens, J., Dahl, G. E., and Hinton, G. E. (2013). On the importance of initialization and momentum in deep learning. *ICML (3)*, 28:1139–1147.

Sutskever, I., Martens, J., and Hinton, G. E. (2011). Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pages 1017–1024.

Tesla (2016). All Tesla Cars Being Produced Now Have Full Self-Driving Hardware. https://www.tesla.com/blog/all-tesla-cars-being-produced-now-have-full-self-driving-hardware. Last accessed 8 January 2017.

Toole, A. A. and Czarnitzki, D. (2006). Patent Protection, Market Uncertainty, and R&D Investment. *Berkeley Center for Law and Technology.*

Tramer, F., Zhang, F., Juels, A., Reiter, M. K., and Ristenpart, T. (2016). Stealing Machine Learning Models via Prediction APIs. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 601–618, Austin, TX. USENIX Association.

TRIPS (1996). Agreement on Trade-Related Aspects of Intellectual. Property Rights. *WTO.*

Unterthiner, T., Mayr, A., Klambauer, G., Steijaert, M., Wegner, J. K., Ceulemans, H., and Hochreiter, S. (2014). Deep learning as an opportunity in virtual screening. *Advances in Neural Information Processing Systems*, 27.

USPTO (2016). Patent FAQs. https://www.uspto.gov/help/patent-help. Last accessed 14 January 2017.

Vinge, V. (1993). TECHNOLOGICAL SINGULARITY. http://www.frc.ri.cmu.edu/~hpm/book98/com.ch1/vinge.singularity.html. Last accessed 2 January 2017.

Vinyals, O. and Le, Q. (2015). A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Wagle, A. M. and Ødegaard, M. (1997). *Opphavsrett i en digital verden*. Cappelen, Oslo.

Williams, R. (2016). Google's artificial intelligence writes miserable poetry. http://www.telegraph.co.uk/technology/2016/05/16/googles-artificial-intelligence-writes-miserable-poetry/. Last accessed 4 January 2017.

WSU (2017). Fair Use Doctrine | University Communications | Washington State University. https://ucomm.wsu.edu/fair-use/. Last accessed 18 January 2017.

Yu, H. (2016). AlphaGo's Success Shows the Human Advantage Is Eroding Fast. *The New York Times*.