Artificial Intelligence: Are there any social obstacles?

An empirical study of social obstacles

ERIK LILIEQUIST
Artificial Intelligence - Are there any social obstacles?

An empirical study of social obstacles

by

Erik Liliequist

Master of Science Thesis TRITA-ITM-EX 2018:221
KTH Industrial Engineering and Management
Industrial Management
SE-100 44 STOCKHOLM
Artificiell Intelligens - Finns det några sociala hinder?

En empirisk studie av sociala hinder

av

Erik Liliequist

Examensarbete TRITA-ITM-EX 2018:221
KTH Industriell teknik och management
Industriell ekonomi och organisation
SE-100 44 STOCKHOLM
Artificial Intelligence - Are there any social obstacles?

Erik Liliequist

Abstract

Artificial Intelligence is currently one of the most talked about topics with regard to technical development. The possibilities are enormous and it might revolutionize how we live our lives. There are talk of robots and AI removing the need for human workers. At the same time there are also those who view this as deeply troublesome. Either from an individual perspective, asking the question what we should do once we do not need to work more? Or from an existential perspective, raising issues of what responsibilities we have as humans and what it means to be human?

This study does not aim to answer these grand questions, but rather shift the focus to the near future of three to five years. Yet, there is still a focus on the social aspects of the development of AI. What are the perceived greatest social issues and obstacles for a continued implementation of AI solutions in society?

To answer these question interviews have been conducted with representatives for the Swedish society, ranging from politicians, union and employers’ organizations to philosophers and AI researchers. Further a literature study has been made of similar studies, comparing and reflecting their findings with the views of the interviewees.

In short, the interviewees have a very positive view of AI in the near future, believing that a continued implementation would go relatively smoothly. Yet, they pointed to a few key obstacles that might need to be addressed. Mainly there is a risk of increased polarization of wages and power due to AI, although stressed that it depends on how we use the technology rather than the technology itself. Another obstacle was connected to individual uncertainty of the development of AI, causing a fear of what might happen. Further, several different ethical issues were raised. There was an agreement that we need to address these as soon as possible, but they did not view this as an obstacle.

Key-words: Artificial Intelligence, Social Obstacles, Ethical AI, Technical Polarization, Gender and AI, Technical Unemployment
Sammanfattning

Artificiell Intelligens är just nu ett av de mest omtalade ämnena inom teknisk utveckling. Möjligheterna är enorma och AI skulle kunna revolutionera hur vi lever våra liv. Det talas om att robotar och AI tar bort behovet av mänsklig arbetskraft. Samtidigt finns också de som ser stora problem med AI. Antingen från ett individuellt perspektiv, med frågor som vad vi ska göra om vi inte behöver jobba mer? Eller från ett existentiellt perspektiv, med frågor om vilket ansvar vi har som människor och vad det innebär att vara människa?

Den här studien syftar inte till att svara på dessa stora frågor, utan flyttar fokus till en nära framtid av tre till fem år. Dock är fokus fortfarande på de sociala aspekterna av utveckling av AI. Vilka uppfattas vara de största samhällsriskerna och hindren för en fortsatt implementation av AI lösningar i samhället?

För att svara på dessa frågor har representanter från det svenska samhället intervjuats, allt från politiker, fackliga och arbetsgivarrepresentanter till filosofer och AI forskare. Dessutom jämförs deras svar med en litteraturstudie av liknande studier.


Nyckelord: Artificiell Intelligens, Sociala hinder, Etisk AI, Teknologisk polarisering, Kön och AI, Teknologisk arbetslöshet
## Contents

1. **Introduction** ................................................................. 4
   1.1. Background ............................................................ 4
   1.2. Problem description ................................................ 5
   1.3. Purpose ................................................................. 5
   1.4. Delimitations .......................................................... 5
   1.5. Limitations ............................................................ 6
   1.6. Research question ................................................... 7

2. **Methodology** ................................................................. 8
   2.1. Methodological approach ............................................ 8
   2.2. Research process ..................................................... 9
       2.2.1. Literature review ............................................. 9
       2.2.2. Pre-study ..................................................... 9
       2.2.3. Empirical study .............................................. 10
   2.3. Reliability and validity ............................................. 11
       2.3.1. Researchers’ bias ........................................... 12

3. **Literature Review** .......................................................... 13
   3.1. Studies and predictions of AI ...................................... 13
   3.2. Methodological framework ......................................... 16

4. **Empirical Results** ......................................................... 18
   4.1. Pre-study ............................................................... 18
       4.1.1. Technical perspective ....................................... 18
       4.1.2. View of AI .................................................... 19
   4.2. Empirical study ....................................................... 20
       4.2.1. Definitions of AI ............................................ 21
       4.2.2. Opportunities ............................................... 22
       4.2.3. Risks .......................................................... 24
       4.2.4. Social obstacles ............................................ 25
       4.2.5. Specific obstacles .......................................... 26

5. **Analysis** ........................................................................ 30
   5.1. Definitions of AI ...................................................... 30
   5.2. Social obstacles ....................................................... 31
   5.3. Which social obstacles were considered major? .................. 33

6. **Discussion** ..................................................................... 36
   6.1. Gender and AI .......................................................... 36
   6.2. Nuances in public discussion ........................................ 38
   6.3. Compression to other technological changes ...................... 39

7. **Conclusions** ................................................................... 41
Preface

First of all, I would like to thank all of the interviewees being part of this study for all your insights and wisdom. I never believed it would be so easy to conduct these interviews and you have made the work with this project truly enjoyable. I left all of the interviews with lots of new perspectives and thoughts. Some of them ended up in this report, others I carry with me for the future. Once again, thanks!

Secondly, a big thanks to my supervisor Henrik Blomgren. Our discussions with regard to this project have been enlightening and you have pushed me to pursue paths I did not believe was possible.

Thirdly thanks to my seminar classmates for good feedback and interesting discussions. Looking forward to hear your final findings!

Lastly a huge thanks to Anette for all the help with reading the report at different stages, the great feedback and the interesting discussions. This report would not have been the same without you!

Erik Liliequist
Stockholm, May 2018
1. Introduction

This study revolves around social obstacles connected to the development of Artificial Intelligence and the perception of them. This section aims to introduce this study. To do so, a short summary of the background of AI is presented, motivating why the study is currently needed. Secondly the problem and purpose of the study are presented. Thirdly some of the limitations and delimitations of the study are discussed. Lastly, the research questions are presented and motivated.

1.1. Background

During the last few years there has been a huge improvement in the performance of different Artificial Intelligence (AI) systems. This improvement can be explained by several causes (Jordan and Mitchell, 2015). For instance, the increased amount of digital data generated through the use of Internet. This gives more training examples that can be used for training the AI systems. Further, improvements in hardware have allowed more complex computations to be made in a feasible time. Another cause is the improvements in learning algorithms. Yet, these improvements haven’t moved passed the academic world (Reynolds et al., 2017) and still haven’t affected individual everyday life nor society in general to any greater extent, perhaps with the exception of companies like Facebook and Google and their services (Andrews et al., 2016).

To some extent, the development and implementation of AI could be viewed as the next step of automation. As discussed by Schmitz et al. (2009) this could be viewed as “the Automation of Automation”, where AI is used to automate the supervising role that is needed today. This development have given rise to several different concepts, for instance the German term called “Industrie 4.0” (Drath and Horch, 2014) or the term used by GE “Industrial Internet” (Evans and Annunziata, 2012), which describes what they see as the next step of industry. Both share the view that everything need to be connected through the Internet-of-Things (IoT) and that in the end some of the decision making and adaptation should be made by intelligent machines. Apart from this development within the industry, there is also a possibility for a wider implementation in society by automating tasks and process that have been considered to complex to automate. Goldberg (2012) further argues that automation moves from classic manufacturing to “include applications in Healthcare, Security, Transportation, Agriculture, Construction, Energy, and many other areas.” An early study by Frey and Osborne (2013) gives an estimation that 47 percent of the employment in the US is at risk in an unspecified time frame, although they speculate that it could be “perhaps a decade or two” (p. 38). A similar Swedish study by Fölster (2014) suggest that 53 percent of employment in Sweden are at risk, within the same time frame. If these changes occur society will need to address some of the issues caused with regard to unemployment.

In late April 2018, after gathering the empirical results for this study, Vinnova published an overlooking study of AI in Sweden (Marklund et al., 2018). One field investigated in this study was that of social obstacles, but only through an literature review. This study therefore complement their views. Further, as the Vinnova study was ordered...
by the Swedish government in December 2017, this shows the need to explore this field.

1.2. Problem description

This study aims to investigate the social aspects of changes related to a wider implementation of AI in society. By identifying major social obstacles and issues, it gives the possibility to address the issues in an early stage of the transformation. This in turn might speed up the transition and decrease some of the skepticism towards AI. This study therefore mainly contributes to two separate fields of work. First, and foremost, it discusses attitudes towards technical transformations in general, and in specific which obstacles are likely in a transformation to an AI driven society. Apart from contributing to the academics, it also aims to help managers and innovators by identifying key social issues that they need to address. Secondly, it also could contribute to the field of AI by identifying issues in current systems, and therefore allows researchers to push technical solutions to address these issues.

Further, there is a need to clarify what is included in "changes related to a wider implementation of AI". As discussed above it includes all different kinds of automation that is made possible with the help of AI. This raises the need for a definition of what AI is. Unfortunately, there is still disagreement on this, and it does not exist any agreed upon definition of the term (Marr, 2018). For the purpose of this study a broad definition will be used, which have been inspired by Tegmark’s (2017) definition of intelligence:

Artificial intelligence are artificial systems which have the ability to achieve complex goals.

Hence the list of technologies included is very long and some examples are: search engines, automatically directed ads, autonomous vehicles, self-regulating industrial system, natural language processing, image recognition and AlphaGo. This broad definition also allows for a comparison between how different individuals and organizations define AI, which might also affect their attitude towards it.

1.3. Purpose

The purpose of this study is to:

- Identify social obstacles and issues that exists with regard to AI and when possible address these obstacles and issues.
- To understand the perceptions of Swedish politics, industry and trade with regard to a wide implementation of AI.

1.4. Delimitations

To delimit the scope of the of the study a focus will be on the issues and obstacles, rather than the possibilities. The main reason for this delimitation is the fact that the possibilities at this stage is far easier to identify and motivate, as it springs from the
technological advances. As it is easy to identify advantages and possibilities today, it is not uncommon to view AI as "magic" (Elish and Boyd, 2018). Thus, the industry and academics manage to find most possibilities themselves. Autonomous vehicles are a fitting example of this, as the possibility to have a completely autonomous vehicles fleet is obvious. From a technological point of view the autonomous vehicles even are safer than human drivers (Thompson, 2016). Yet, there are resistance towards autonomous vehicles due to factors such as the feeling of driving, who is legally responsible etc. Most of these factors have a clear social aspect, something that can’t be solely solved by technical solutions. Therefore, further investigating these social obstacles will provide more value. Similar reasoning hold true for most AI technologies.

Another delimitation of this study is to focus on representatives from unions, employers’ organizations and different government departments instead of individuals. This is motivated by a bigger gap in the academic work, as there have been several studies on individual views of AI (European Commission, 2017; Petronio, 2018). This delimitation is also motivated by the type of data that is possible to gather from these different sources. As surveys, such as European Commission (2017); Petronio (2018), gather a quantitative data there is a need to complement it with more qualitative data. To gather qualitative data that is not purely speculative, the respondents need to be chosen more carefully. Thus limiting the focus to representatives from unions, employers’ organizations and different government departments.

Lastly, this study do not focus on the effects of Artificial General Intelligence (AGI). This is motivated by the fact that it is unlikely that we will get AGI within the near future (Müller and Bostrom, 2016). As a result, the greater threats (such as AI controlling the world) and the obstacles connected to those types of threats are not discussed, although being important questions to explore.

1.5. Limitations

The limitations of this study are closely connected to the chosen method used to gather data. As a qualitative exploratory study the number of interviews conducted was rather low and the finding therefore lack generalizability. Although interviewed in the role of representative for different organizations, the interviewees all had similar background and can not be considered fully representative for the general public nor their organizations. To address these issue a quantitative study could be conducted, targeting those acting in the organizations or the general public. Secondly, the questions asked are of a highly speculative nature, thus both analysis and interviews should be viewed as that. Even if interviewees were chosen carefully to get as good speculations as possible, sudden events can change the conditions radically. This study should therefore be viewed as predictions given a limited timeframe.
1.6. Research question

This study aims to answer the following research question:

**RQ1:** What are perceived as major social obstacles for a continued wide implementation of AI solutions in the near future?

With the following sub-research questions:

- **Sub-RQ1:** What different definitions of AI exists among the stakeholders?
- **Sub-RQ2:** How do a stakeholder’s definition of AI influence which social obstacles that are identified as problematic?
- **Sub-RQ3:** Which different types of social obstacles exists?
- **Sub-RQ4:** Which social obstacles can also be viewed as an opportunity by another stakeholder?

The sub-RQ:s are motivated by the following reasoning. Before trying to understand which social obstacles different stakeholders raises as problematic, it is important to understand how each stakeholder define AI (sub-RQ1). Although above definition allows for a broad interpretation of what is considered to be AI, there is still a need to conduct some research in the field. This question could therefore be answered before finding any social obstacles, but it might need to be revised after investigating the obstacles and connecting different definitions to different obstacles (sub-RQ2).

To be able to answer RQ1 there is a need to understand which different types of social obstacles that exist. For instance, there could be issues on an individual level such as a lack of trust in AI or even a fear of it. Another possible type of issue is on a community level with a risk of increasing unemployment. To be able to identify the major obstacles, it is important to first understand which types of obstacles exist (sub-RQ3).

Further there is a high probability that different stakeholders don’t agree on what is an opportunity and an obstacle. As mentioned above the risk of increased unemployment could be viewed as a threat, but it could also be viewed as a possibility to reduce number of hours that is needed to work or as a possibility to bring previously outsourced tasks back to even lower costs. This could both vary between different sectors and within the sectors, as managers and employees might be affected differently by the changes. Investigating the issues identified from several perspectives therefore is need to further understand which are problematic (sub-RQ4).
2. Methodology

In this section the methodology is presented. First, a short presentation of the general methodological approach, connecting the chosen methodology to the purpose and research questions. Secondly, the process of this study is presented and motivated. Thirdly, the reliability and validity of the study is discussed, and the researchers’ bias is presented.

2.1. Methodological approach

This study is conducted within the interpretive paradigm (Greener, 2008), focus is how actors within organizations understand the changes that occur with regard to AI. A focus therefore is how the interview subjects interpret and reflect on the changes occurring, rather than the researcher. The interpretive paradigm further allows for unstructured interviews and a qualitative approach (Greener, 2008). Thus, this characterize the method used in this study. Further a slightly changed approach of grounded theory will be used, by conducting an iterative method working with interview recordings, transcripts and summaries to categorize findings. (Strauss and Corbin, 1998; Greener, 2008).

The study clearly falls within the cross-sectional approach, as there is a focus on a new technology emerging which affects several different actors (Greener, 2008). Thus several different actors, within different sectors, are used as sources. The data are collected within a short timeframe, therefore presenting a snapshot of their views. Although one aspect of the study is to extrapolate current themes into the near future, it is still based on a narrow current context.

It is also worth noticing this study’s claim to generalizability. Generalizability is defined as to which extent it is possible to extend the findings in a study to the population or society in general (Collis and Hussey, 2013). Although the goal is to identify obstacles that are common for several sectors, thus indicating a high generalizibility, it is important to remember that this study is a snapshot and might not be generalizable over time. Neither do this study make any greater claim to hold true for any time but the time being studied. Some of the obstacles identified might also be a higher issue in certain sectors, thus being less relevant across sectors.

To complement the grounded theory, a social constructionist view is used (Andrews, 2012). It is argued that knowledge is constructed rather than created. Therefore, how the individuals interviewed view the world might influence what they view as relevant. As this study aims to capture perceptions of social obstacles, it is of importance to account for how these perceptions are constructed. Further, it is stressed that if the views of social constructionists is not based on a relative perspective, it is compatible with grounded theory. Along the lines of Hacking (1999) it is important to stress what is socially constructed in the case of AI. Is it the object, idea, or elevated word that is socially constructed? In the case of AI, all three could be relevant. AI as an object could be easily be viewed as socially constructed through the data it learns from and perspectives of the developers. The idea of AI can be socially constructed as assumptions on the development of the new technology is made. It can be constructed as the next step...
of digitalization, or the next step of the industrial revolution. Lastly, the elevator word intelligence could also in some sense be viewed as socially constructed. An anthropocentric view of what AI is could influence our construction of the term (Hacking, 1999). These constructions are not questioned to any greater extent in this study, but rather simply recognized.

2.2. Research process

This study consists of three major parts: a literature review, a pre-study and an empirical study. These parts is then complimented by a final analysis.

2.2.1. Literature review

During the entire study a literature review has been conducted. The purpose of this study have been two-folded. The first purpose was to identify different frameworks and relevant obstacles. This included fields such as previous technological changes (for instance environmental changes, the rise of Internet etc.). The second purpose was to identify other studies of opinions and attitudes toward AI and digitalization. This was done to provide a context for this study and to be able to further support or deny aspects found in this study. Thus, the later part could be viewed as another source for triangulation to increase reliability (Olsen, 2004).

For the literature study two main search engines have been used, Google Scholar and KTHB Primo. Both of these in-turn point to several different sources, such as journal, magazines and books. The literature have been found using both references from other authors, but also different combination of the following list of search terms:

Artificial Intelligence; Ethics; Economics; Predictions; Adoption; Internet; Climate Change; Technological Changes; Diffusion of Innovation; Social Obstacles; Threats; Possibilities; Risks; Gender; Superintelligence; Artificial General Intelligence; Industrial Revolution; Luddites; Requirements for Resistance;

2.2.2. Pre-study

The pre-study was conducted to gain insights into the field and a better understanding of which topics were relevant for the study (Collis and Hussey, 2013). This was done to be able to better understand which types of questions were relevant to ask during the empirical study. To some extent the goal was also to in part answer sub-RQ3, regarding which social obstacles exists. Further the purpose was to sharpen and confirm the research question for the study, ensuring the relevance of the research question.

The pre-study was conducted across several different sectors and positions, to catch as many different perspectives as possible. Shorter phone interviews (approximately 30 minutes) were conducted with five different actors, e-mail answers from one actor and one actor were interviewed in-person. The questions asked in all cases were the same. The phone interviews as well as the in-person interview were all of an semi-structured nature. The e-mail questions were open-ended and the response was thorough. Yet, this
response was more structured compared to those conducted orally. The different sectors of the respective interviewees are presented in Table 1.

The semi-structured form allowed the interviewees to talk freely about the topic, but still follow a general structure (Cohen and Crabtree, 2006). This fits the purpose of the pre-study well, as it is both to explore and confirm some topics. The questions used during the pre-study can be found in Appendix A.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Organization</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Employers’ organization</td>
<td>Phone</td>
</tr>
<tr>
<td>B</td>
<td>Employers’ organization</td>
<td>Phone</td>
</tr>
<tr>
<td>C</td>
<td>Employers’ organization</td>
<td>E-mail</td>
</tr>
<tr>
<td>D</td>
<td>Workers’ Union</td>
<td>Phone</td>
</tr>
<tr>
<td>E</td>
<td>Workers’ Union</td>
<td>Live Interview</td>
</tr>
<tr>
<td>F</td>
<td>Industrial company</td>
<td>Phone</td>
</tr>
<tr>
<td>G</td>
<td>Communication company</td>
<td>Phone</td>
</tr>
</tbody>
</table>

Table 1: Pre-study interviews

2.2.3. Empirical study

The empirical study is based on interviews with different actors, from the academics, politics and business sectors. The choice to use different sectors were motivated by the different background of the interviewees. As these might have different constructions of AI, their different views compliment each other. Further, it is worth stressing that political motives might influence how an interviewee view the change. By interviewing actors from different sides of the political scale reduces this risk. Another example is that of employers’ organizations and unions, where the two is likely to represent different views. The purpose of these interviews was to gather qualitative data and in-depth knowledge of their respective sectors with regard to social obstacles for implementation of AI. In specific the purpose was to find how the actors defined AI, which social obstacles they saw as most relevant, but also which possibilities they saw. Another purpose of the interviews was to compare the respective sectors, asking them to confirm or deny obstacles and possibilities as important. Thus confirming or rejecting the findings from the pre-study.

Semi-structured interviews were the method used for the empirical study. The interviews were conducted in two parts. First a part with open-ended questions which allowed the interviewees to respond freely. The interviewer could at this stage also ask follow-up questions, to explore certain areas deeper (Collis and Hussey, 2013). As a result new perspectives and ideas could be found. The second part of the interview contained questions that were less open-ended, and some possible to answer with “Yes” or “No”. These questions were constructed to further investigate the findings from the pre-study, thus the areas of Unemployment, Increased polarization of wages and power, Gender and Individual fears were in focus. Although these questions were not as open-
ended, the interviewees were encouraged to expand on their answers when possible. All of the interviews were recorded to be analyzed at a later stage, allowing the interviewer to focus on the interview. A summary of the interview was later sent by email to each interviewee to allow them to make sure they were interpreted correctly. The summary, rather than the recording, was then used as data for the report.

During the empirical study eleven persons were interviewed. In Table 2, the interviewees are presented. When contacted the interviewees were informed that the interview would take approximately one hour and that if possible preferred to be conducted in person. Most of the interviews were conducted like this, although some were limited on the request of the interviewee. In the case of the phone interviews, this did not have any greater effect and the interviews still went smoothly. The shorter interview on the other hand became slightly stressed, hence limiting the possibility to investigate some answers in-depth. The interview with Interviewee C was less structured than the rest, not following the questionnaire to any greater extent. This was motivated by the rather academic background of interviewee C, instead focusing on the philosophical questions that Interviewee C had studied. To find the questions asked during the interviews see Appendix B.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Organization</th>
<th>Form</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Governmental Official</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>B</td>
<td>Policy Institute</td>
<td>Live Interview</td>
<td>30 minutes</td>
</tr>
<tr>
<td>C</td>
<td>Philosopher</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>D</td>
<td>Policy Institute</td>
<td>Phone Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>E</td>
<td>Management Consultant</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>F</td>
<td>Journalist/Publisher</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>G</td>
<td>Employers’ organization</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>H</td>
<td>Workers’ Union</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>I</td>
<td>Policy Institute</td>
<td>Live Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>J</td>
<td>Professor Information Systems</td>
<td>Phone Interview</td>
<td>1 hour</td>
</tr>
<tr>
<td>K</td>
<td>Professor Mathematical Statistics</td>
<td>Phone Interview</td>
<td>1 hour</td>
</tr>
</tbody>
</table>

Table 2: Empirical study interviews

2.3. Reliability and validity

High reliability indicates that the study is done in a way that allows for consistency and that the design of the study is transparent (Greener, 2008). In other words, the study should be conducted so that the reader can trust the results. As this study is of a qualitative nature, it is important to think of biases from both participants and researchers. To address research bias, below is a clear statement of my views and opinions that might affect the research. Further, as suggested by Chenail (2011) a pre-study can be conducted to limit the bias of the researcher. Lastly, allowing the interviewees to confirm their responses decreases the risk of the research interpreting answers in a
biased way. To address participant biases an awareness when conducting and preparing interviews are needed, to not influence the respondents. Noting that respondents often want to confirm the views of the interviewer is important when formulating questions, to reduce the possibility of this (Furnham, 1986). To further increase the reliability triangulation can be used. Triangulation is the use of different methods to collect data which can be used to confirm data found through other methods (Greener, 2008).

Validity can be split into three categories: face validity, construct validity and internal validity (Greener, 2008). Face validity is gained through using a method that a first glance seems reasonable for the question asked. As a part of this study’s purpose is to complement the quantitative studies conducted through the use of surveys, an qualitative approach with semi-structured interviews has a high face validity. Construct validity is a measure of how well the method measures what the researcher think it measures (Greener, 2008). An inductive study needs an exploratory method, thus semi-structured interviews are a good option in producing open ended answers. The method chosen therefore gives the exploratory aspects needed for an inductive study. Lastly, internal validity raises the issue of causality. This is an important issue to have in mind when analyzing the data collected, not being biased to confirm causality when it is not clear.

2.3.1. Researchers’ bias

In this section I address some of my personal views and opinions that might cause me to be biased when conducting this study. By acknowledging these I let you, the reader, read the findings knowing what might have caused me to reason in certain ways. This being said, I have tried to be as impartial as possible in my analysis.

First of all I am a firm believer in technology, and especially in the possibilities of AI. This causes me to be to positive toward the technology, and perhaps neglecting or underestimating some of the risks and obstacles. Further this could have the effect that I overestimate the possible outcomes of AI, thinking it is better than it actually is. Secondly, I believe in humankind. Although some obstacles raised are really problematic, I do think we can solve them. This also might cause me to underestimate some obstacles and risks. It is also worth noting that what I consider to be an obstacle or risk might be considered something positive by others. Or in other words, what I view as obstacles is what could stop or delay the development of AI, although it might be of benefit to humanity.

Another aspect worth mentioning at this stage is that of Artificial General Intelligence (AGI). AGI could be seen as the opposite of the narrow AI, such as chess-playing or automobile driving. AGI rather is an AI that have self-control and can adapt to any sort of problem solving (Goertzel and Pennachin, 2007). The estimations of when AGI is here varies a lot. According to an interview with Hiroshi Yamakawa the average estimate of when an AGI on human level could be possible is 2030[^1]. Although exciting and thrilling, I still believe this estimate is on the lower side and therefore AGI is not relevant for this study.

3. Literature Review

The literature review consists of two parts. In the first section, studies, reports and predictions with regard to the future of AI are presented. At this point a short summary of the studies are presented, to be discussed in the analysis section of this study. The second section presents studies made using different methodological approaches, which in turn can be used in this study.

3.1. Studies and predictions of AI

The Future of Employment: How susceptible are jobs to computerisation? (Frey and Osborne, 2013)

By using a statistical approach Frey and Osborne have calculated how susceptible jobs are for computerisation. To do this they have used the O*NET information on different occupations. The occupations are described using different variables. Experts were then asked to classify 70 out of 700 occupations as automatable or not. Using regression the rest of the occupations are evaluated.

Using this method Frey and Osborne find that “47 percent of total US employment is in the high risk category, meaning that associated occupations are potentially automatable over some unspecified number of years, perhaps a decade or two” (p. 38). Further they also provide probabilities of how susceptible different occupations are to computerisation. Some of the occupation at high risk of computerisation are transportaion, logistics, office and administrative support workers, labor in production and service occupations.

Vartannat jobb automatiseras inom 20 år - utmaningar för Sverige (Fölster, 2014)

Using a similar methodology as Frey and Osborne (2013) Fölster shifts focus to the Swedish labor market. Fölster concludes that Sweden is even more susceptible for computerisation than the US, with 53 percent of occupations at risk. Further, also inline with Frey and Osborne, he finds that occupations that require a high amount of finger dexterity, artistry, social competence, persuasion and consideration/care are at lower probability of computerisation.

To address these changes Fölster points to four key conclusions. First he stresses the importance of a sharp innovation policy. Secondly, innovation also need to occur with respect to how we educate ourselves. A higher focus need to be on ongoing education during work life. One way to solve this could be through digital education. Third, there is a need to realize that workers might not be the future tax base. Fourth, instead of focusing on taxes, there is a need to focus on economic growth to support different government programs for social security.

Automation, skills use and training (Nedelkoska and Quintini, 2018)

This OECD study from 2018 is made to contrast the Frey and Osborne (2013) study, using slightly different statistics for the predictions. The main critique of the Frey and

---

2https://www.onetonline.org/
Osborne study is that they have too aggregated job descriptions, thus losing nuances in which jobs might be susceptible to automatization. Instead of using the O*NET database, this study uses Survey of Adult Skills for statistics with regard to competences.

With these different statistics they find that 14 percent of workers face a high risk (probability of over 70 percent) of losing or fundamentally changing their job. Additionally, they find that 32 percent of jobs are at between 50 and 70 percent risk of being automated.


This study published by McKinsey and Company in October 2017 aims to discuss how the labor market will change by 2030 in the European “digital front-runners” (Belgium, Denmark, Estonia, Finland, Ireland, Luxembourg, Netherlands, Norway and Sweden). This studies prediction of the number of tasks automatable is in agreement with both Föllster (2014) and Frey and Osborne (2013), predicting 44 percent of current work hours being at risk. Yet, they argue that less than 20 percent of jobs will be made completely obsolete. Further, they argue that more new jobs will be created, thus creating a net surplus of jobs. Hence, they find “[…] that automation and AI bring significant benefits, including new jobs and increased productivity.”

To be able to fully use the technology they also suggests a few key areas that needs to be focused by the digital front-runners. First there is a need to keep supporting the infrastructure needed for AI, and lower barriers for innovation. Secondly, support local AI ecosystems. Third, focus on education for future tasks. Fourth, help workers transition to new tasks or jobs. And lastly, help shape the global policies regarding AI.

Artificial Intelligence and Economic Growth (Aghion et al., 2017)

This report focus on the economical aspects of AI and how future economic growth can be affected by AI. To do this several frameworks and theories from economics are used to develop a model of how an increased automatization will impact economic growth. In conclusion they find that “our model […] generates a rich set of possible outcomes” (p. 46). Several of these outcomes allows for a balanced growth with a constant capital share well below 100 percent, with a nearly fully automated economy. Thus indicating that there are possible scenarios were a full use of AI is possible, still keeping part of the tasks available for labor incomes. This is motivated by the fact that some tasks are hard to automate and still will require a human hand.

Further, the authors argues that AI in part discourages innovation in two separate ways. First, AI encourage imitation, therefore reducing the possible gains from innovation and limiting the returns to the inventor. Secondly, AI supports a rapid creative destruction allowing a shorter life-span of new inventions, thus reducing the possible economic returns.

http://www.oecd.org/skills/piaac/
Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics (Brynjolfsson et al., 2017)

This report starts in the paradox of the AI hype and the lack of support for this in economic statistics. To explain this mismatch they suggest four different explanations: false hopes, mismeasurement, redistribution, and implementation lags. Although all explanations have separate cases and validity, the authors argue that the implementation lags are the core explanation of the mismatch. Further, they argue that there is a need of complementary innovations to fully achieve the possibilities of AI.

Another insight is raised with regard to whom are pessimistic and optimistic about AI. In general, they argue, that technologists and venture capitalists are optimistic about the future, while economists, sociologists, statisticians and government officials are pessimistic. Interaction is much more present within these groups compared to between them, and when communication occurs between the groups they are often talking past each other. Lastly, they also argue that there is a need to update how we measure productivity to fully capture the productivity gains from these new technologies.

Special Eurobarometer 460: Attitudes towards the impact of digitalisation and automation on daily life (European Commission, 2017)

The Special Eurobarometer 460 is a survey among all 28 EU countries. The survey was asked to individuals in the respective countries, ensuring a statistical significance for each country. One of the key findings from this survey is that Europeans in general are positive towards recent digital advancements, 75 percent believe it has a positive impact on the economy, 67 percent on the quality of life and 64 percent on society. Yet, they also find this to differ among different groups. Men, younger respondents, well-educated, daily Internet users and those with less financial stressors are more positive towards digitalization and robotization. Further, they find that the opinion of AI strongly depend on the knowledge of AI, with a greater knowledge increasing the positive view. Although in general positive, 72 percent believe that robots and AI steal people’s jobs and "almost nine in ten respondents agree robots and AI are technologies that require careful management (88%)” (p. 6).

Optimism and Anxiety - Views on the Impact of Artificial Intelligence and Higher Education’s Response (Petronio, 2018)

This survey by Gallup and Northeastern University was done to understand the US public opinion of AI. Similar to the Special Eurobarometer 460 (European Commission, 2017), they found that 76 percent of Americans are positive towards the impact of AI on society. In general, white-collar workers and highly educated (bachelor’s degree or higher) individuals are more optimistic regarding AI. Although optimistic, the Americans still were aware of the risks and 73 percent believe in a net lose of jobs. In contrast, only 23 percent believed their job was at risk. The authors concludes that this "... suggests some may be underestimating the likely impact of this technology” (p. 30). Further, the American public is unsure about what type of skills are needed to succeed in an AI society, split even between soft and hard skills.

To be able to adapt to these changes, the authors concluded that universities and
collages need to take a key role as they are best suited to educate the workforce. There is also a need for the universities and collages to collaborate with employers, as most respondents believe they will get their education through their employer.

3.2. Methodological framework

To understand social obstacles with regard to AI, it is important to understand the individual perceptions of AI. A recent study focused on how the New York Times had reported about AI during the last 30 years (Fast and Horvitz 2017). To do this they developed a framework to capture ‘‘[…] a diverse set of hopes and concerns about the future of AI’’ (p. 964). This framework was divided into three major groups of measures. Two types of General Measures are suggested. First is a measure of engagement, how much is written about AI over all. Second is a measure of attitude towards AI, optimistic vs pessimistic.

Hopes of Artificial Intelligence is the second group of measurements, these focus on the positive attitudes towards different aspects of AI. These measures tries to capture what extent different aspects of AI are regarded as positive. The third group is Concerns of Artificial Intelligence, which focus on the negative attitudes towards AI. All of the measures suggested in these categories are presented in Table 3.

Although this study focused on an analysis of news articles, the measurements are still valid for this study. The measures suggested in Table 3 can be further divided into sector specific measurements, such as Education, Transportation, Healthcare, Entertainment and Military applications, and more general issues that is relevant regardless of sector. These general issues are directly applicable to this study, while the sector specific is of more use to give examples and understand where AI can be used. It is further possible to use these sector specific data to confirm or deny attitudes found during the interviews.

An article by Dietterich and Horvitz (2015) raises five classes of risks as highly relevant in the near future, and areas where further research is needed. First, is the risk of bugs which includes all the types of errors that can be caused by the developer. Second, cybersecurity is an ever increasing concern, and AI is as vulnerable as other software solutions. For example, manipulation of training data, changes of preferences and changing

<table>
<thead>
<tr>
<th>Hopes</th>
<th>Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact on work (positive)</td>
<td>Impact on work (negative)</td>
</tr>
<tr>
<td>Singularity (positive)</td>
<td>Singularity (negative)</td>
</tr>
<tr>
<td>Merging of human and AI (positive)</td>
<td>Merging of human and AI (negative)</td>
</tr>
<tr>
<td>Education</td>
<td>Loss of control</td>
</tr>
<tr>
<td>Transportation</td>
<td>Military applications</td>
</tr>
<tr>
<td>Healthcare</td>
<td>Absence of Appropriate Ethics</td>
</tr>
<tr>
<td>Decision making</td>
<td>Lack of progress</td>
</tr>
<tr>
<td>Entertainment</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Measurements to capture hopes and concerns with regard to AI.
the goals. Third, the risk of “Sorcerer’s Apprentice”, where the AI simply follows the instructions, without regards for other rules or what we would consider common sense. Fourth, shared autonomy is a class of risks that relates to the issue of setting clear goals and states. There is a need of communication between human users and the AI system, sometimes at a fast pace. In these settings it is important that the current states are very clear, to not cause accidents like China Airlines flight 006. Fifth, is the risk of socioeconomic impacts, urging an increased understanding of how AI influence distribution of jobs and wages.

Several of these risks are of a technical nature, yet they have impact on society and individual views of AI. Therefore, these sets of risks should be accounted for when investigating social obstacles, although not in focus.

\[^4\]https://en.wikipedia.org/wiki/China_Airlines_Flight_006
4. Empirical Results

In this section the empirical results from the pre-study and the empirical study are presented. Although the pre-study was conducted to help the empirical study, some of the findings still are valid for later analysis and discussion. The results are summarized in this section, for full accesses to the data contact the author.

4.1. Pre-study

During the interviews questions were asked with regard to what they thought of and how they viewed AI. The first section was to establish their description of AI and how they defined it. The second section was instead to discuss their attitudes and opinions about AI. The results are presented using the same division.

4.1.1. Technical perspective

From the interviews it is apparent that there is no clear definition of AI, as is clear from the associations presented in Table 4. When asked how they view AI today, examples of the technology ranged from pattern recognition to increased automation through robotizing. At the same time, some of the interviewees did not think robotizing as equal to AI, or that robotizing might not benefit to any greater extent from AI. The general theme that most of the interviewees agreed upon was the need for large amounts of data. Thus, they reasoned that the information driven parts of the organizations are where AI could have the highest influence. Several of the interviewees focused on decision making, and the possibility to make smarter decisions. Several of the interviewees also noted that this was a change since previous automation, as white-collar jobs are targeted instead of blue-collar jobs.

<table>
<thead>
<tr>
<th>Associations to AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-learning systems</td>
</tr>
<tr>
<td>Not human programmers</td>
</tr>
<tr>
<td>Re-shoring jobs through AI</td>
</tr>
<tr>
<td>Increased robotizing</td>
</tr>
<tr>
<td>Sharpen human senses</td>
</tr>
<tr>
<td>Pattern recognition</td>
</tr>
<tr>
<td>Robots is not equal to AI</td>
</tr>
<tr>
<td>Mainly in organizations with high amounts of data</td>
</tr>
</tbody>
</table>

Table 4: Summary of associations to AI

An interesting fact to notice during the interview was that of autonomous vehicles. Non of the interviewees thought of this as an example of AI when asked to say what they associate to AI. When mentioned by the interviewer almost all of them agreed upon it being a good example.
Another interesting perspective raised by one of the interviewees was that of AI as a way to sharpen the human senses. The interviewee connected AI to technology such as Virtual Reality (VR) and Augmented Reality (AR), which the interviewee saw as other ways of sharpening the senses. While VR and AR focuses on our five input senses, AI instead aims to sharpen our minds.

Viewing five years ahead, the key prediction was that the technology will keep developing without being disruptive. Several of the interviewees saw AI as a tool, or a compliment, in the near future rather than something that will cause unemployment. A big driver in this development they believe to be the possibility to save resources, which in turn can lead higher quality or lower prices. Their believes in AI as a tool to improve decision making is present in the near future as well, with several mentioning data mining and data analysis as areas where most development will occur.

Although this rather positive view of the technology in five years, the interviewees still managed to identify several challenges (see Table 5). The challenges could be divided into three categories: Knowledge of AI, Relationship between machine and human and Costs and technical challenges.

The most striking challenges from the interviews were those in the category Knowledge of AI. Several of the interviewees believe that there is a need for a high general knowledge in society. Further they see that different actors think of AI slightly differently. This in turn causes bad communication and a lack of focus towards common goals. Further they stressed that the general public (read the customer or user) also need to be educated about AI. Some of the interviewees said a lack of knowledge causes fear, which might influence the relationship between AI and individuals. The employers’ organizations also stressed the importance of gaining expertise in sectors other than IT. They argued that to be able to understand, identify possibilities and implement AI other organizations need the expertise as well.

The Relationship between machine and human were not discussed in any greater detail, but all of the interviewees mentioned it in some way. It still was an rather abstract challenge, but they noticed it as a challenge. Most of the discussions were with regard to how individuals should relate to AI. Cost and Technical challenges is slightly out of the scope of this study, and therefore the interviewer did not investigate to deep in the field. The interviewees that raised the issue did it as a challenge to solve technical problems while maintaining a low cost to be able to sell the system. They argued that if it is to expensive, very few will be interested in implementing the systems.

4.1.2. View of AI

Overall the interviewees had a positive view of AI and the future of AI. Yet, most of them was quite cautious and uncertain. Most of these skeptical feelings were connected to how humans should relate to AI. Issues such as "What should we do with all time when we get even more support in our workday?" were raised. Most of these are of a rather philosophical nature, and several of the interviewees found that there probably is no universal answer to these questions.

With regards to risks of implementing AI three topics where raised, Neglecting the
Challenges

**Knowledge of AI**
- A lack of expertise
- A lack of expertise in sectors other than IT
- Shifting knowledge of AI in society
- Poor understanding of AI systems in society

**Relationship between machine and human**
- Trust to AI and machines
- Patience in training AI

**Costs and technical challenges**
- Cost of developing AI
- Too much unstructured data
- Technical advances

Table 5: The greatest challenges for a continued development of AI

*humans, Poorly trained AI and Bad intents. Neglecting the humans* is a risk closely connected to the challenge category of *Relationship between machine and human*. The extreme fear is that of AGI, when machines have the possibility to control the world. Even if this extreme is not likely, they still identify minor identity issues connected to for instance losing jobs. Both *Poorly trained AI* and *Bad intents* give the rise of bad or evil AI, the difference being if it was intended to do bad or not. Some of the interviewees also connected this to the fact that we lose control of how the AI makes decisions, thus making it hard for us to determine if the decision was fair. Another interviewee raised the issue that we train the AI to learn our bad traits, such as discrimination and racism.

### 4.2. Empirical study

The results from the empirical study are presented using the same structure used during the interviews. First the interviewees definition of AI, and which technologies they include in the term. Secondly what they identify as the greatest opportunities, risks and obstacles following a wide implementation of AI. Lastly, the interviewees thoughts of certain obstacles are presented. Each section contains two parts, first a general summary of the opinions during the interviews, and secondly examples from individual interviews.

*During this section interviewee statements will be presented. Those in italic and within quotation marks are quotes from the interviews translated by the author. The statements presented in plain text are rephrasing done by the author, but agreed by the interviewee.*
4.2.1. Definitions of AI

When asked to define what AI was, it became apparent that there still is no common definition of the term. Table 6 is a summary of what the different interviewees wanted to include in the definition. Although the question asked was to give a definition, non were able to form a concise reply but instead started to reason about characteristics and technologies. Therefore, the summary contains very broad terms, with several interviewees connecting to several different definitions.

<table>
<thead>
<tr>
<th>&quot;Definition&quot; of AI</th>
<th>Interviewees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithms and systems that uses huge amounts of data</td>
<td>A, D, G, H, J</td>
</tr>
<tr>
<td>Learning through previous experiences to use in new situations</td>
<td>A, B, D, H, J</td>
</tr>
<tr>
<td>A system that can make autonomous decisions</td>
<td>D, F, G, K</td>
</tr>
<tr>
<td>Systems that copies human intelligence</td>
<td></td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>E</td>
</tr>
<tr>
<td>AI is software that is good on specific tasks</td>
<td>I</td>
</tr>
</tbody>
</table>

Table 6: Interviewees "definitions" of AI

Table 6 is a summary of what was actually said during the interviews, but it is possible to further summarize these if instead looking on what was being conveyed. First, as a distinction between different types of AI, all of the interviews were talking about specific AI, not AGI. As apparent from Table 6 there was focus on how it was done, rather than what was achieved (Pattern Recognition, Learning from previous experience, Huge amounts of data). Yet, as argued clearly by interviewee E and supported by most of the interviewees, there is a need to make a distinction between systems such as Robotic Process Automation (RPA) and Internet-of-Things (IoT) as they were on its own not considered to be intelligent. Thus, as clearly stated by interviewees F and K, there is a need to mimic some sort of human behaviour.

In contrast to the view that AI needs to mimic human intelligence, interviewee A stated that

there is no need for a system to be highly intelligent to be considered to be AI.

Interviewee B reasoned in a similar way, stating that

there is no need for a system to be conscious to be intelligent.

Interviewee I summarized the relation between human and artificial intelligence in a way that echoed through the interviews:

What separates AI from human intelligence is that AI is much more specified and lacks the possibility to make decisions in the way a humans can.

As a part of the discussion about definitions, interviewee E raised the issue of how it is defined on a societal level. Interviewee E argued that
from societal perspective three major distinctions can be made: viewing AI as a trend or hype, the next step of digitalization and a new phase in our development as a society.

Depending on which of these perspectives you view AI from, different actions will be considered necessary to address the changes.

Lastly, interviewee F raised the issue that the definition of AI is ever-changing. Technologies in the past that were considered to be AI might not be considered to be that today. For instance, the chess-computer Deep Blue that played, and won, against Kasparov in 1997 was considered to be a prime example of AI. Yet, viewing the techniques used by Deep Blue it is more or less brute force searches of the best possible move, which is not considered to be to intelligent (Campbell et al., 2002). Further, tests such as the Turing Test might be made obsolete, as they do not capture the core aspects of AI.

All the interviewees were also asked to associate AI to services used in everyday life. Their associations are presented in Table 7. Although there are several different associations, most of the interviews discussed the technical platforms, such as Facebook and Google, and different types of personalization through recommendation algorithms. One important aspect to account for both when evaluating which services and technologies that are associated with AI today is our willingness to pay, as expressed by interviewee H. The willingness to pay will also influence which services and technologies will emerge in the future.

<table>
<thead>
<tr>
<th>Technologies and services</th>
<th>Interviewees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search engines</td>
<td>A, B, E, I, J, K</td>
</tr>
<tr>
<td>Technical platforms (Facebook, Google, etc.)</td>
<td>D, E, F, I, K</td>
</tr>
<tr>
<td>Personalized feed</td>
<td>B, D, E, J</td>
</tr>
<tr>
<td>Image recognition/analysis</td>
<td>A, G, K</td>
</tr>
<tr>
<td>Autonomous cars</td>
<td>B, K</td>
</tr>
<tr>
<td>Administrative processes</td>
<td>H, J</td>
</tr>
<tr>
<td>Digital Personal Assistants (Siri, Alexa)</td>
<td>F</td>
</tr>
<tr>
<td>Chat bots</td>
<td>F</td>
</tr>
<tr>
<td>Watson</td>
<td>F</td>
</tr>
<tr>
<td>Natural Language Processing (NLP)</td>
<td>G</td>
</tr>
<tr>
<td>Medicine and medical diagnosis</td>
<td>H</td>
</tr>
<tr>
<td>Sensor technology (through IoT)</td>
<td>H</td>
</tr>
</tbody>
</table>

Table 7: Interviewees associations of AI

4.2.2. Opportunities

All interviewees were asked to identify the greatest possibilities of a wide implementation of AI in the near future, three to five years. Numerous different opportunities were identified. Four of these were major fields identified in several interviews, see Table 8.
Throughout the interviews, a clear theme was that of opportunities in the public sector. Although increased efficiency in administration and services are applicable in both private and public sector, the examples mainly were related to public sector. All of those supporting increased efficiency in administration and services did so out of the possibility to solve some of the issues related to an aging population. As the population grows older, there is a need to solve the issue of how to finance public services. AI can help reducing the costs of public services by increasing the efficiency. Interviewee B further argued that an increased productivity (efficiency) is not the same as an improved quality. One effect from an increased efficiency in production of goods and services is the possibility to increase focus on quality.

This becomes more apparent within healthcare and teaching, where the main contribution is an increased quality. Within healthcare, the possibility to have an AI system as a support in diagnosis, or even as a back-up doctor, provides huge potentials for both preventive and curative care, argues Interviewee C and E. Interviewee C further argues that the combination of AI and human intelligence is where the greatest gains are in the near future. Within teaching, the opportunities is to personalize education and allow the teachers to focus more on teaching.

Last, and most concrete, of the four major themes is that of autonomous vehicles. Although being mentioned as an opportunity in four of the interviews, it was discussed in next to all. Several of the interviewees were reluctant to include autonomous vehicles in to what was considered AI. Yet, most seemed to agree after thinking about it for a while. A part from further showing the complexity of the term, it also indicates the issue of AI closer to automation not being viewed as a part of AI. At the same time, interviewee A gave an example of how AI can improve production through using an AI to operate a drill it can make more well-informed decisions on how far it is possible to push the tool without increasing the risk of breakage.

Interviewee F argued along the same line by saying that technology infused with AI can be easier for the user to use, thus reducing the barrier for new technology to diffuse as less education is needed.

Table 8 presents the rest of the possibilities identified during the interviews. Some of these might overlap each other and those presented in Table 8 but the distinctions were made based on the discussions in the interviews.
### Opportunities (interviewee)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency in production (A)</td>
<td>AI (robots) physically stronger than</td>
</tr>
<tr>
<td>Increased productivity (B)</td>
<td>humans (A)</td>
</tr>
<tr>
<td>Reducing human biases (B)</td>
<td>Increased security (B)</td>
</tr>
<tr>
<td>Increased diffusion of technology (F)</td>
<td>Using AI to structure unstructured</td>
</tr>
<tr>
<td>Image recognition/analysis (G)</td>
<td>Smart energy grids (H)</td>
</tr>
</tbody>
</table>

Table 9: Possibilities suggested by individual interviewees

#### 4.2.3. Risks

In Table 10 a summary of the different societal risks of a wide implementation of AI are presented. Several of the top identified risks are overlapping to some extent and are connected. In short, two themes of risks were identified among most interviewees, ethical and moral issues and transparency issues. Most of the ethical and moral issues were related to how AI should cope with biases introduced in the data or how developers should remain unbiased. Next to all interviewees argued that we should not let the AI be neutral and simply reproduce the current structures and biases. This in turn connects to the issue of whose ethics should be implemented. Interviewee D stated this as

"When you acknowledge that something is not simply a platform anymore, it becomes a very tough challenge of deciding whose ethics, whose regulation and whose normativity should be applied."

<table>
<thead>
<tr>
<th>Risks</th>
<th>Interviewees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neglecting moral and ethical issues</td>
<td>A, E, G, H, J</td>
</tr>
<tr>
<td>Theft or misuse of data</td>
<td>C, D, E, I</td>
</tr>
<tr>
<td>Transparency of AI services</td>
<td>D, E, G, I</td>
</tr>
<tr>
<td>Diffuse or incomprehensive reasoning behind answers or actions</td>
<td>A, B, I</td>
</tr>
<tr>
<td>Who should control the AI?</td>
<td>G, H, I</td>
</tr>
<tr>
<td>Unanticipated outcomes from actions</td>
<td>A, F</td>
</tr>
<tr>
<td>Biased AI</td>
<td>B, E</td>
</tr>
<tr>
<td>Decreasing number of persons needed to work</td>
<td>A, I</td>
</tr>
<tr>
<td>Intelligent weapon systems</td>
<td>B, K</td>
</tr>
<tr>
<td>Decreasing number of tax payers</td>
<td>A</td>
</tr>
<tr>
<td>Negative impact on the climate issue</td>
<td>F</td>
</tr>
<tr>
<td>Reducing our ability to communicate and socialize with other humans</td>
<td>F</td>
</tr>
</tbody>
</table>

Table 10: Risks of a wide implementation of AI

Transparency issues are somewhat connected to the ethical and moral issues, but also
include several other aspects. As argued by interviewees A, B and I, one of the major transparency issues is that of understanding the reasoning made by the AI. They argued the relevance of this on different levels. First of all, AI have the possibility to process vast amounts of data which by sheer numbers is not possible for humans. Therefore, a human observer will not be able to identify the reasoning behind decisions in a feasible time. Secondly, the current development within Deep Learning, where deep neural networks are used, the decision making can be to abstract for a human to fully comprehend. Thirdly is the issue of the developers sharing all their source code and data. Even if it would be feasible to validate the reasoning, it is not possible due to the developers not sharing the material. Interviewee B even suggested that the lack of transparency can lead to AI making decisions that we can not predict

"Systems can process such vast amounts of data that we can not understand what they do with it and how they learn from it, which might cause problems that we can not even understand."

A part from these major themes there is also worth mentioning the risk of cyberattacks. Interviewees C, D, E and I mentioned this as a real threat. The increased amount of data used by AI systems, often of a personal nature, increases the reward for cyberattacks, thus creating a higher risk of cyberattacks. Stated explicitly by interviewee A, and mentioned by several others, is the need to have a global perspective on the development, especially the risks. Interviewee A even compared it to the threat of climate change, arguing that both of theses changes are abstract, remote and disputed. Therefore there is a need to make agreements like the climate agreements with regard to AI and its development.

4.2.4. Social obstacles

Table 11 summarizes the obstacles identified in the interviews. One theme throughout most of the interviews was that of uncertainty and fear about the development of AI. The category social uncertainty about the disruptive nature of AI is one example of this. This revolves around the unpredictability of what outcomes might come and how it will affect us as humans. As argued by interviewee C as humans, we are creatures of habit, reluctant to make fast changes. Sudden changes to our habits puts us under a great amount of stress, thus making our reactions unpredictable.

In turn, this connects to the risks of unemployment and reduced tax bases, by creating a fear of these changes. As argued by interviewee F, the fear might be a bigger issue than the actual risk.

One of the risks of AI is that of unemployment, but this risk might be enlarged. Yet, the fear of unemployment might be a big obstacle, as this fear raises voices to stop the development.
Further, there is a risk that some individuals and policymakers wants to stop the development or are skeptical towards the new technology. This could result in actions similar as those of the Luddites during the industrial revolution, with actions taken to prevent development.

<table>
<thead>
<tr>
<th>Social Obstacles</th>
<th>Interviewees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social uncertainty about the disruptive nature of AI</td>
<td>A, B, C, F, H, I, J, K</td>
</tr>
<tr>
<td>Legal restrictions</td>
<td>A, E, H, I, K</td>
</tr>
<tr>
<td>Technical skepticism</td>
<td>B, C, F, H</td>
</tr>
<tr>
<td>Stopping the development</td>
<td>A, F</td>
</tr>
<tr>
<td>Lack of competence outside of technological companies</td>
<td>E, G</td>
</tr>
<tr>
<td>Increasing need for transparency</td>
<td>D</td>
</tr>
<tr>
<td>Addressing biases and structures in data</td>
<td>D</td>
</tr>
<tr>
<td>Attitudes of how we view and structure data</td>
<td>E</td>
</tr>
</tbody>
</table>

Table 11: Social obstacles for a wide implementation of AI

A second theme was that of our attitudes towards sharing data and the lack of coherent standards across industries. In a basic form, this revolves around laws and legal restrictions. Although out of the scope of this study, several of the interviewees argued that there is a lack of agreement on what the laws should regulate, even before starting to formulate the laws. This issue therefore is of a social nature, revolving around which ethics and morals should be agreed upon. As AI is built upon data, there is the question of how we, as individuals, should share our data and be ensured it is used in a safe and legal way. As many opportunities were identified in the public sector, interviewee E raised several important obstacles within Swedish public sector. In short, there are currently regulation of how different public sectors can share information with each other. As most of the Swedish municipalities are rather small, there is a lack of data to be able to fully use the possibilities presented by AI in each municipality on its own. One way to solve this would be to share data between municipalities, but that is not currently allowed. Further, even if this would be allowed, there is a huge problem with a lack of standards.

Lastly, a theme running through most sections of the interviews, was that of transparency of solutions. There is a need to overcome the issue of how to explain the decisions and actions of AI. Another aspect of transparency is how to handle different sort of biases. These types of obstacles are hard as they are close to ethics, and there is a need to solve these issues before implementing AI solutions widely. Or as expressed by interviewee G and J

"We should invest one dollar into ethics for each dollar invested in AI."

4.2.5. Specific obstacles

Four specific obstacles were investigated deeper during the interviews, Unemployment, Increased polarization of wages and power, Gender and Individual fears.
**Unemployment**

During the interviews, none of the interviewees viewed unemployment as an obstacle, or at least not to any greater extent. In short, this view was motivated by a belief in that new jobs would be created, explicitly stated by interviewees A, D, E, H and I. Another fact reducing the fear of unemployment in Sweden is the fact that Sweden currently is facing a lack of labor. Interviewee F argued that this lack of labor in part could be solved by increasing the productivity through AI. Further, interviewees D and I pointed to the fact that the discussion regarding unemployment is very polarized. Either there is a belief that all jobs will be lost, or non. Both argued that more nuances is needed in the discussion.

Another aspect of unemployment was raised by interviewee A who stated:

Employment will be affected differently during different phases of diffusion of AI. Within IT, first there might be increased employment due to cheaper production costs and increased demand. Once the supply catches up with the demand, and assuming a continued increased efficiency through the new technology, the employment will decrease. Lastly, new types of employments will be created, thus stabilizing slightly above the original level of employment. An important perspective is that these gains and loses can occur in different parts of our global economy.

Both interviewee C and G expressed similar reasoning. Although being of a speculative nature, these example presents two important perspectives. First is that there is no guarantee that the adoption of AI will be smooth, in fact as stated above the opposite is more likely. Interviewee G further argued that as a result the adoption will be slower, allowing society to adopt along the way. Thus, the impact on employment will be spread over a longer timeframe. Secondly it is of importance to remember that society now acts in a global market. As this study is done in a Swedish setting, there might be global causes and effects with regard to our adoption of AI. For instance, Sweden making an active choice of not developing AI will probably have limited effect as some actors on the global market will develop it which will impact our market and society. Or the other way around, Swedish companies reshoring tasks from abroad, which will reduce the number of jobs abroad.

**Increased polarization of wages and power**

With regard to polarization of wages and power, there were no clear trends from the interviews. Most interviewees found it hard to reason about these questions, and that it was of a very speculative nature. The general view was rather well summarized by interviewee A and D:

Hard to reason about how AI will affect the polarization of wages and power, but it will probably not help.
Interviewee B, F and H pointed to the fact that the technology itself does not cause polarization, but how we as a society use it does. A part from this some thoughts of a speculative nature was presented. Interviewee E, G and H viewed AI as a continuation of digitalization, which in turn already have created some increase in polarization. Although reasoning about which jobs are at highest risk of being made obsolete is hard, interviewee B, F and I all speculated that mid-wage jobs are at risk. This in turn might cause an increased polarization, but depending on which actions is taken it does not have to cause polarization.

Gender

Table 12 summarize the interviewees view on gender and AI, and which outcomes they viewed as possible. Although the question were asked from a gender perspective, some of the discussions during the interviews were made including several other grounds for discrimination. Most interviewees also stressed the speculative nature of their reasoning, and that they found it hard to reason about these topics.

<table>
<thead>
<tr>
<th>Possible outcomes of Gender and AI</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different market segregation causes different effects of AI</td>
<td>A, B, E, G, H, I, J, K</td>
</tr>
<tr>
<td>Biased data will generate biased AI</td>
<td>D, E, F, G, I, J</td>
</tr>
<tr>
<td>Male dominated development environment causes biases</td>
<td>C, E, F, H, I</td>
</tr>
<tr>
<td>Increased need for abilities that AI lack</td>
<td>A, B</td>
</tr>
</tbody>
</table>

Table 12: Four general possible outcomes of AI with respect to gender

When asked the question of gender and AI, most associated with different biases that are introduced to AI systems. This could either be in the form of biased data or a biased developing environment. Some discussions was of conscious biasing (where the developer makes the system biased in some way on intention), but most found it more interesting to discuss the unconscious bias that occur. With regard to data, this goes back to the issue of unclean data and to what extent we should clean it. For instance, training on historical data will with high probability include biases of different kinds, should the developers then have a responsibility to clean it. Secondly, with regard to the developing environment, most agreed that there is an issue that most developers are male and with a similar background. To create AI that is suited for use in society, there is a need to include other perspectives than a technical, argued interviewee J.

When presented with the case, or raised the issue themselves, of which jobs was at highest risk of automation through AI and the statistics of current gender segregation, more than half of the interviewees agreed that women and men will be affected differently. Interviewee I presented the case in the following way

Men are both the big winners and losers as a result of an increased use of AI. For instance, male dominated jobs such as truck drivers probably will be automated relatively shortly, while other male dominated jobs such as programmers will gain an increased status. With regard to female dominated
jobs, most predictions are harder to make as the new technology will rather act as a complement than replacement.

Lastly, interviewee A and B believed that since some jobs will be lost, these workers have to search for something new to do. As AI currently is not performing as well in tasks which involves lots of creativity, empathy and care, these types of characteristics will be more sought after. This could in-turn change how we view female and male norms.

**Individual fears**
Interviewee A, C, F, G and I all explicitly discussed issues of an existential nature with regard to individual fears. For instance, what should we do if we do not have to work anymore (interviewee C)? What do I have to do to make sure there is food on my table (interviewee J)? Interviewee J raised an interesting example of how we as a society believe in and trust the individual to a great extent. When changes such as this occur, this leads to a lot of the focus is on how I as an individual should cope with the changes, while it might be better to ask the question of how society should cope with it. This focus on the individual creates a lot of unnecessary fear and anxiety.

A part from these existential issues, there is also a fear with regard to responsibility. As individuals we feel a need to know who to blame in case something goes wrong, or at least get an explanation why. These types of questions are harder when an AI is involved. For instance the example of "**Who is responsible in case of an autonomous vehicle crash?**" came up in next to every interview. In the end, interviewees G, H and I argue that this leads to a fear of lack-of-control when an AI makes more and bigger decisions.
5. Analysis

This section aims to answer the research questions stated in the introduction. The answers are based on the empirical study and the literature review. First Sub-RQ1 and Sub-RQ2 regarding definitions of AI are answered. Secondly, social obstacles are identified by answering Sub-RQ3 and Sub-RQ4. Lastly, the urgency of these obstacles are discussed to answer RQ1.

5.1. Definitions of AI

Sub-RQ1 “What different definitions of AI exists among the stakeholders?” and Sub-RQ2 “How do a stakeholder’s definition of AI influence which social obstacles that are identified as problematic?” both address how the stakeholders define AI. As apparent from the interviews, as well as from the literature, there is no consistent definition. What is apparent though is how the academic definitions are slightly different than those used in society. From an academic point of view it is more formal and strict, which is no surprise. Yet, there is a difference in that in the academic it is more common to talk about abstract goals and intelligence rather than concrete examples. For instance, my own definition, inspired by Tegmark, “Artificial intelligence is artificial systems which have the ability to achieve complex goals” states achieve complex goals as the measure. The recent Vinnova [Marklund et al., 2018] study defines it as:

“Förmågan hos en maskin att efterlikna intelligent mänskligt beteende. Det vill säga den förmåga hos maskiner som möjliggör för dessa att fungera på meningsfulla sätt i relation till de specifika uppgifter och situationer de avses utföra och agera inom. Artificiell intelligens är också det vetenskaps- och teknikområde som syftar till att studera, förstå och utveckla maskiner med intelligent beteende.”

”A machines ability to mimic intelligent human behaviour. In other words the ability that enables machines to act purposely in accordance with specific tasks and situations that it was made to act within. Artificial intelligence is also the scientific and technological field that aims to study, understand and develop machines with intelligent behaviour. (freely translated by author)”

Aghion et al. [2017] suggests that AI can be defined as “the capability of a machine to imitate intelligent human behavior” or “an agent’s ability to achieve goals in a wide range of environments.” in their summary of definitions of AI.

The advantage, and disadvantage, of these definitions is that they can include next to anything. By contrast the definitions suggested during the interviews were more tangible. The categories of Learning through previous experiences to use in new situations and Algorithms and systems that uses huge amounts of data both describe the technology rather than what it should achieve. These categories of course contain weaknesses, as you need to define what learning and huge amounts of data are. Yet, these definitions appear to be easier to distinguish technologies than those of “complex goals” or “intelligence”.

30
Further the suggested category of a system that can make autonomous decisions can be considered as wide as the academic in which technologies it includes, but in most cases it presents a distinct line between what is AI and not. A weakness of this definition, as expressed by interviewee K, is that it lacks a dimension of complexity. For instance, a thermostat regulating the temperature is making autonomous decisions, yet most would agree that this is not considered to be AI.

Moving on to Sub-RQ2 there were no clear patterns in how the interviewees defined AI and what they viewed as social obstacles. Yet, three minor observations can be made:

- Four of five who defined AI in part as Learning from previous experiences identified the disruptive nature of AI as an obstacle. In total eight of eleven identified this as an obstacle.
- Four of five who defined AI in part as Huge amounts of data identified neglection of ethical and moral issues as a risk. In total six of eleven identified neglection of ethical and moral issues as a risk.
- Three of four who defined AI in part as Autonomous decisions identified autonomous vehicles as an opportunity. In total five of eleven identified autonomous vehicles as an opportunity.

These observations should not be viewed as any trends or major conclusion, and no statistical significance can be assured. As interviewee E argued, there might be difference depending on the definition of AI from a societal or economic perspective, if it is viewed as a trend, the next step of digitalization or a new industrial revolution. This type of definition was not gathered during the interviews but extrapolating from their responses most would probably have a quite similar economic or societal definition. Further, all interviewees had some higher education and to some extent worked with questions relating to digitalization and AI. This, and the fact that no clear differences were identified in their technical definitions, might explain why no connection between definition of AI and views of AI were found.

In conclusion, there exists lots of different definitions of AI and there are advantages and disadvantages of most. It was identified that those in the academic context define AI slightly more abstractly and including, compared to actors in society. With regard to Sub-RQ2, although no clear patterns were found in this study, this was not the main purpose of the study and it can not be disregarded as a hypothesis. Investigating how the definition of AI from a societal or economic perspective influence what is identified as risks and obstacles is of high interest. Understanding this connection might help create a better public discussion of AI.

5.2. Social obstacles

One key aspect to understand which social obstacles are considered major, is to identify social obstacles. Sub-RQ3 was stated to achieve this: “Which different types of social obstacles exists?” Through the pre-study, four major themes were identified as possible
social obstacles, **Unemployment, Increased polarization of wages and power, Gender** and **Individual fears**. Apart from the gender perspective, these have support in the theoretical framework. Unemployment have been the focus of several other studies, as presented in the literature review. [Dietterich and Horvitz (2015)] view socioeconomic impacts as a real threat in the near future, which connects to both unemployment and increased polarization of wages and power. Individual fears are found as an issue by several surveys, to varying extent, and should be considered a possible obstacle. Individual fears are also highly relevant looking further into the future, for instance the issues relating singularities [Fast and Horvitz (2017)].

During the empirical study these obstacles reoccurred in most interviews, although sometimes in slightly different form. The issue of biases in data and in the developing environment was viewed as an obstacle, somewhat connected to the issues raised with regard to gender. What was viewed as an obstacle here is the fact that there is a need to address the biases. All of the interviewees argued that there is a need to not simply reproduce the biases. This in turn forces someone to make a decision of what is considered to be unbiased. Further, we as society need to agree upon which rules should be applied, which types of biases can be considered fine to reproduce. Interviewee K raised an interesting example with regard to gender variables. If there is an AI system that is supposed to decide an insurance policy, gender could be a considered variable. We might decide that gender should not influence how good an insurance policy you get, therefore removing this variable from the data. Yet, there is a probability that the gender can be learned from the other data inputted, perhaps even creating a latent variable to some extent corresponding to gender. This latent variable can then be used to evaluate the insurance policies in a gender biased way. The question then is if this is problematic or not, which is something we as a society need to agree upon. Another perspective of the same issue is that of how to handle plain text, as it contains human biases that an AI system will learn [Caliskan-Islam et al. (2016)].

Another trend in the empirical study was that of moral and ethical issues. Regardless of which issue, there was an agreement that ethical issues are currently overlooked. There is the often mentioned example of how an autonomous vehicle should act in case of an accident. Should it protect the passengers in the car or should it try to minimize the general damage? These types issues are of a very specific nature but still raises big ethical questions. It is also likely that similar questions will be asked in several different fields. Therefore, it could be of interest to reason about underlying ethical question in a general way, to be able to apply the answer in several different settings. If this on the other hand is done separately, let us say for instance that each autonomous vehicle developer is allowed to make this decision on its own, it might lead to a situation where companies compete on an ethical ground. Taking it to the extreme, slogans like “*Our car will try to save YOUR life to any cost!*” might be used. Secondly it is also the bigger issues, with regard to how we should arrange our society. How should we handle unemployment, reskilling and an increased polarization that might occur due to AI? As argued above, these are obstacles that we need to handle in some specific way, but there is a need to agree upon how we should handle them.

Lastly it was found that most interviewees thought that there is an uncertainty about
the disruptive nature of AI. In short, individuals are uncertain about what will happen due to the introduction of AI in society. As found by Petronio (2018) 73 percent of Americans believes there will be a net lose of jobs, yet only 23 percent believed their job was at risk. This shows the sometime paradoxical views of AI by the public. Yet, this uncertainty is believed to cause a skepticism towards AI, as the future is unclear. This in turn might express it self as wish to stop the development. There will be a challenge to communicate the possibilities of AI in a nuanced way that the public understands and trusts.

In conclusion, seven broad categories of obstacles were identified in this study: Unemployment, Increased polarization of wages and power, Gender, Individual fears, Addressing biases, Moral and Ethical issues and Social uncertainty of the disruptive nature of AI.

With regard to Sub-RQ4 “Which social obstacles can also be viewed as an opportunity by another stakeholder?” several insights were gained. As stated by several of the interviewees the increased efficiency gained through an increased use of AI can result in unemployment. Yet, they argued in the case of the Swedish public sector that there currently is a need for a higher productivity, mainly due to an aging population. Although unemployment in its own is only a risk, the cause of unemployment still can be viewed as an opportunity. Further, some thought that there is a wish to work less among parts of the population, thus viewing an reduced number of work hours as something positive. The same holds true for the discussion regarding gender and AI. Most interviewees identified both possibilities for increased equality and discrimination in society. For instance there is a chance to make juster judgments in several cases (courts, public cases, insurance policies, etc.). On the other hand there are also issues of biases and active discrimination through the AI systems. Although it was apparent that most of the interviewees wanted to address the biases in the data, it is important to remember that this have to be done by ”someone”. Depending on this ”someones” ethics and view of the world, this could be considered both a possibility and a real threat.

Regardless of which specific obstacle was discussed, what seemed to matter was not the technology it self but how we decided to use it. Or as expressed by interviewees B, F and H, it is the humans that cause the polarization, not the technology. Of the obstacles identified Individual fears, Moral and Ethical issues and Social uncertainty of the disruptive nature of AI were considered to be primarily obstacles without any possibilities. Yet, these could have positive effects as it forces us to stop and think of what causes the fear or issue. This in turn could lead to a development of sounder and more sustainable AI systems.

5.3. Which social obstacles were considered major?

The main research question of this study was ”What are perceived as major social obstacles for a continued wide implementation of AI solutions in the near future?” First it is worth stressing that non of the interviewees viewed Unemployment as a big obstacle in the near future. This was primarily motivated by two arguments. First, they argued that there is a need for a longer timeframe to allow the technology to mature fully and to
fully gain the benefits. In the timeframe of this study, of three to five years, the diffusion of AI will not have happen to any greater extent. Secondly, and also shifting focus to a slightly longer timeframe, most of the interviewees believed that labor market dynamics would mitigate the loss of jobs. How this was to be done depended on interviewee, but the two main thoughts were reducing the hours needed to work and creation of new tasks. Yet, it is important to not disregard these questions, because some individuals will lose their job. There is a need to be aware of the changes that will occur on the labor market due to automatization. Simply disregarding these changes, believing it will solve itself, most interviewees argued was not the way to go. As stated above, it is how humans use the technology that affects the society. Several also argued that as AI will develop in a global economy different parts might of be affected differently.

This view was even more apparent with regard to Increased polarization of wages and power, where it comes down to what and how we decide to use the new technology. One of the main issues with regard to this is how taxation need to change to address changes in who is working. If there is an increased automatization, more gods and services will be produced through non-human workers. As the current tax systems is based on human tax payers, this reduces tax incomes for the government. Therefore it is needed to change how taxation is done to be able to keep providing public services. Interviewee K also reasoned that this will have further effects on polarization. As the gains from production now is more closely connected to non-human workers, the owners of these non-human workers gains increased income. Further, non-human workers are traded simply using capital, while human workers to some extent is limited by working hours (and the fact that it is illegal to trade humans as gods). As capital have a higher tendency to distribute unevenly, this will increase the polarization.

In summary, issues regarding polarization of wages and power are closer in time than those of unemployment, thus making it a bigger obstacle in the near future. As these effects also are closely connected to those of digitalization in general, the development is all ready ongoing. Finding good ways to regulate this new technology without limiting the will to innovate will be a big challenge. Finding good ways to distribute power on a global scale is another huge challenge. Studies made from an economical point of view shows that there are possibilities to combine full automatization without extreme polarization [Aghion et al. 2017; Korinek and Stiglitz 2018].

Different aspects of Gender and AI were found interesting, but not necessary an obstacle for the development. Although the importance of this was stressed, there was no belief that AI stands or falls with this issue. Yet, as with the obstacles discussed above, it is important to start working with questions of gender and AI as early as possible. Thus, gender and other grounds for discrimination, should be a perspective when creating legislation surrounding how biases should be handled in regard to AI.

Individual fears and Social uncertainty of the disruptive nature of AI are closely connected, or rather the latter is a subset of the first. These relate to how easily AI will diffuse to society. Therefore this was considered one of the biggest issues in the near future. As AI, apparent from the discussion regarding definitions, is a broad and fuzzy term, so is the public’s relation to the technology. Although being both hard and abstract issues, this is something that can be rather easy to work with.
It is of high importance to increase the transparency of the AI system used and of the organizations using AI. Legislation is one way to solve this, but as most of the development is currently occurring in a global setting, there is no clear organization that can enforce the legislation. Although European Union's initiative with regard to GDPR is one example of an attempt to regulate what is allowed, this also creates a risk of EU lagging behind as the same regulations do not apply in China and the US for instance. Therefore it is important that developers themselves agree upon standards and clear communication, trying to achieve an as transparent market as possible. However, this might not be in their interest.

One way to address this is through education, and this is something that can be done on a local level. There is a need to change the education at all levels to increase everything from general computer knowledge to AI specific knowledge. Although technical knowledge of the systems is the best way to go, it is unrealistic to believe that everyone want and can learn this. Hence, there is a need to find a way to learn about these issues without being too technical. There is also a need to improve adult education, to allow for easier reskilling if your job is at risk. There is a need to create tools for individuals to use to solve their situation.

In summary, Individual fears and Social uncertainty of the disruptive nature of AI are obstacles that are highly present in the near future. These obstacles will limit the diffusion of AI, and there is a need to address them to achieve a smooth diffusion.

In relation to the discussion of transparency, organizations and companies also need to work with Addressing biases in the data and development of AI systems. As with the gender perspective, this is of importance to create trust in the system from the public. Although not necessary at this stage, it is easier to address issues regarding biases as early as possible. Further, this can also be used to lower the individual fears of AI, as it is easier to trust that the AI systems make fair decisions. But in the near future, addressing biases will not be a major obstacle.

Lastly are the obstacles related to Moral and Ethical issues. As argued before, this has relevance in specific solutions. As some of these solutions are already coming to the market, there is a need to address these issues. The quote of interviewees G and J that "We should invest one dollar into ethics for each dollar invested in AI" is an important reminder of this. Yet, there is the problem of who should invest this additional dollar. These types of investments generate little return in the short run, and the gains would be intangible to some extent. Therefore, companies are unlikely to make these, except out of good will and a sustainable approach. Governments and other official authorities on the other hand are just starting to invest in AI in general, then there is a low likelihood of them investing in ethical issues relating to AI. Interviewee A argued that in the case of Sweden, this might be an advantage that can be used. As Sweden probably will not be able to compete with countries such as the US and China in developing the new technology, Sweden should focus on some other aspect. Interviewee A further argued that historically Sweden have been good at creating social structures to allow for technical changes, and that Sweden should take the lead in implementing ethical solutions. But in general, these issues are not considered to be a major obstacle for a continued implementation of AI.
6. Discussion

This section is a reflection and discussion regarding some of the findings in the study. These reflections are made outside the scope of the research questions, and should not be considered major or more important than any of the other obstacles identified. This should instead be viewed as further analysis of two specific sub-fields: Gender and AI and nuances in public discussion. To some extent both these fields are of high interest for further research or public actions. Lastly a comparison to other historical technological changes is made, very briefly identifying similarities and differences.

6.1. Gender and AI

As new technologies emerges, such as AI, society has to adapt or relate to it in some way. These reactions in turn might have an effect on society in general and different groups in society might be affected differently. For instance, Wahl and Linghag (2013) finds that advances in robotics have increased the possibilities for women to work in industry. Another example is that of the washing machine. Baxter (1998) argues that although this could be viewed as an invention to ease woman’s workload at home, it also increased the expectations of how often the chores should be carried out. Therefore resulting in an increased workload. One interesting question is then how AI can affect women, men and their relations. In this section several different aspects of this will be presented. There is no claim to cover all possibilities, but rather start the discussion with regard to AI and gender. Further, although the focus in this discussion will be of gender, an intersectional perspective could be used as well, focusing on other grounds for discrimination.

As mentioned above, and in several of the interviews, there is the issue of biases in the data. This raises a key question with regard to what type of biases should be allowed. At first glance it would be easy to simply answer that no biases should be allowed, but using an example I argue that this might not be a given answer.

Consider car insurance premiums for instance. These are, or have been, based on several different factors. Some factors of a more technical nature, such as driving patterns and type of car. These have limited amounts of biases as they are of a more formal nature. Yet, there might be biases such as cars from country X are considered more safe, thus reducing the premium for those cars. Some factors are on the other hand more connected to the driver, for instance gender, profession, age and driving history. These on the other hand can be highly biased. A recent UK study showed that young adult male drivers (age 17-19) crash causality were 11.4 time higher than those aged 30-59 and 1.9 times that of young adult female drivers (Jones, 2017). Thus, insurance companies wants to charge younger drivers a higher premium. Further, they might also want to charge young male drivers a higher premium than young female drivers. The latter is in an European setting illegal since 2012[^5]. Yet, it has been argued that it is possible to achieve similar gendered outcomes by variables of professions (assuming that profession correlates to gender to some extent) and engine size (assuming that men are

[^5]: See the case of C-236/09 (Test-Achats) from European Judgment of the Court
more likely to buy cars with bigger engines) (Jones [2017]). This is also in line with the latent variables mentioned by interviewee K. So what types of discrimination, and uses of biases, should be considered fair?

Avraham (2017) summarizes a framework to decide whether a characteristic is considered to be fair or not to use for discrimination when deciding an insurance premium. In short he argues that there is a need to determine if the features of a characteristic is: controllable or immutable, changing over time, causing or correlating with the risk, the characteristics’ predictive value, perpetuating negative stereotypes, and if the characteristic defines a socially salient group. Although his argument is made within the specific case of insurance, this framework should be easy to extrapolate to AI solutions. The first two features and the last are with regard to the characteristic itself, thus possible to decide regardless of context. The other four features on the other hand is highly dependent on context. This creates a setting where we as a society can agree on three out of seven features in general, regardless of which AI solution it is to be implemented in. The rest needs to be decided from AI system to AI system, thus reducing the benefits of having general discussions of how biases should be handled. Gender biases might be acceptable in certain systems, while not in others. It is also important to remember that even if gender is not used as a characteristic, the outcome might still be highly gendered. Hence, there might be a need to do a similar analysis of the outcome as of the characteristics. Using the features suggested in the framework could be of help when analyzing the outcome as well.

A part from issues of how AI systems should act in relation to gender biases, there are also several different scenarios in which AI can effect how we view genders and questions related to gender. On one hand we have the new possibilities created through AI. Chatbots and personal assistants that could replace lots of the human interactions we have, in the long run perhaps having an AI as our closest friend.

Or the increased amount of services possible at home, possibly redefining gender roles. One could argue that an increased robotization of our homes through, for instance robot vacuum cleaners, further liberates women. But another argument is that this just creates new task and expectations, thus redefining the gender roles (Baxter [1998]).

And how should we relate to smart sex-robots, and how does sex-robots effect our gender roles and sexuality? Sex-robots could be viewed as a tool to reduce human trafficking and sexual violence. Yet, there is a high risk that it reproduces skewed views of sexuality, and creates even worse conceptions of sexuality.

On the other hand there are also the aspects of how an introduction of AI will occur on different parts of the labor market at different times. As the labor market currently is rather segregated, it is likely that men and women will be affected differently. As interviewee I argued, it is probable that men will be both the big winners and losers. The primary gains from AI will be earned by those who either have capital and can invest in new AI solutions, or those who develop the systems. Both of these categories are today dominated by men. On the other hand some of the biggest losers will be those doing rather simple, repetitive tasks without human interaction (Fölster [2014], Frey and Osborne [2013]). Examples of occupations that falls within this description are within transportation and production in industry which are male dominated. These
occupations are often of a nature which allows for unskilled workers. Unskilled women on the other hand more often work in care or public administration, which to greater extent require human interactions. Even if public administration was identified as one of the sectors where AI can have the biggest effect, this could be used to increase the quality of the services. Increasing the quality in this case would involve a higher amount of human interaction and personalization. Therefore, it is is not unlikely that men will be the biggest losers due to AI.

In response to losing your job you can either try to get a job that requires a higher amount of skills or find another job at a similar skill level. The latter will be tasks which is harder to automate, and one great deal of these tasks will be those involving high amounts of social competence and human interactions. Abilities such as empathy and care will be valued higher. As these are closely connected to the female norm, females will either be valued higher or the male norm will shift towards the female norm.

Viewing the issue from a power perspective, it could be argued that male dominated sectors will not be affected as much as female sectors, due men being in charge and making the decisions. It would be easier for men to raise concerns regarding what is happening than it would be for females (Wahl and Linghag 2013). It is also possible to argue for the case that the male dominated sectors will be more exposed to an increased automation. This can be done by reasoning that it is more profitable to automate high wage occupations. As many of the low-wage occupations currently are female dominated, this might cause these sectors to be less exposed to AI. On the other hand, this also means that women will not be as likely to get the benefits from AI.

In conclusion, this reasoning is of a highly speculative nature and it is hard to say what will happen for sure, but shifts along these lines are likely. Yet, it is important not to forget to have a gender perspective on the development of AI.

6.2. Nuances in public discussion

Something that became present during the interviews conducted in this study is the need for nuances and definitions, something further supported by Brynjolfsson et al. (2017). As those spoken to in this study were representative for very different organizations and social interests, it is striking how similar their views and opinions were. In part, this might be due to the framing of the questions asked, but this shows the need for a more nuanced public debate. Either it is the general view that AI will solve everything and that there are no problems at all, or it is the view that we will get a superintelligent AI that will control the world.

The discussion regarding unemployment shows this rather well. Some talk are of AI or robots taking all the jobs. "One in two jobs will be automated within 20 years" to translate the title of Fööster (2014) study. Comparing that to the McKinsey and Company (2017) report that conclude that there will be an increased demand for labor in 2030. Yet, they still find that 44 percent of current work hours are at risk of automation, which shows a nuanced view. But the general message from reports like this is that there is no real problem. As stated by interviewee F, journalists often try to find and present these slightly more chocking findings. But, as stated, it is because of the public demands,
there is no demand for the more nuanced stories. There is a need for responsibility from everyone here, to present a nuanced picture and to ask for it.

But there is also a need to speak clearly. Most of the discussions regarding AI are done lacking a coherent definition of what is talked about. When Elon Musk, Tegmark (2017) and Bostrom (2017) argue that AI should be treated with care, it is from an existential perspective with regard to AGI. When Frey and Osborne (2013) and Fölster (2014) argues that every other job is at risk, it is as a result of specific AI solutions. Further, they do not, nor do they claim, to account for all aspects that could impact it. When Google is talking of AI they define it as:

AI is making it easier for people to do things every day, whether it’s searching for photos of loved ones, breaking down language barriers in Google Translate, typing emails on the go, or getting things done with the Google Assistant. AI also provides new ways of looking at existing problems, from rethinking healthcare to advancing scientific discovery. (Google AI, About)

Yet, when discussed these distinctions are often not made, especially by the public. As a result you either believe Google want to create a superintelligent AI to control the world, or you believe Musk, Tegmark and Bostrom want to stop people using translation services. Although I can not be sure, I strongly doubt this is the case. To solve this, it is important to be clear about your definition. And when referencing someone else, be clear about their definition. Otherwise it simply creates a demagogic discussion, resulting in a loss for the shared interest in AI.

6.3. Compression to other technological changes

As technological changes occur, it is always of interest to compare them to similar historical changes. So which historical changes are similar to that of AI. I argue that AI contains aspects making it rather unique, but there are similarities and comparisons that can be made with regard to several. Here I briefly touch upon some of them, but no claim to be an exhaustive comparison is made.

First of all we start with a comparison with that of the first industrial revolution. One of the main similarities between the two are that of productivity gains. Hence, several of the social obstacles connected to the two changes are of a similar nature. Issues of unemployment were present during the first industrial revolution, albeit not being an issue that became realized to any greater extent (Deane, 1979). It is a fear that is present due to AI as well. During the first industrial revolution the most extreme form of resistance was that of the Luddites, who sabotaged the machines. However, that something like this would occur with regard to AI is unlikely. Mainly because AI is a rather abstract concept, and it would be unclear of what to attack to achieve similar outcomes. Yet it is probable that the same attitudes will rise, but the anxiety and anger need to be expressed in some other way. Further, due to the development of internet,

---

6https://ai.google/about/ Retrived: 18:19 16/05/2018
it is common that opinions tend to get extreme, and that you get stuck in your “filter bubble” (Pariser 2011). How will this anxiety and anger then express itself?

As AI is so abstract there is a need to find something else to attack, and one actor at risk is the public sector. This could be seen as the increased distrust towards politicians (Dyck et al. 2018) or a general distrust towards authorities (Schmitter 2015). In turn, this might have bigger consequences on society than those of the Luddites. It is important to remember the difference of tangibility, that AI is often abstract and hard to grasp. Especially for a novice.

A second technological change that could be considered similar is the introduction of internet. Compared to the industrial revolution, it is closer in time and both builds on digital technology. To some extent, both the internet and AI is a part of digitalization. But that is were the similarities stops. One of the major differences between the two is that internet has always been viewed more as a tool or an infrastructure, rather than a replacement of human abilities. AI is strongly connected to an anthropocentric world view, the goal is to mimic or improve human abilities. Although AI can be considered a complement to human activities, for instance as a support in decision making, it also makes greater claims. This greater claim is something that the introduction of internet lacks. The claim made when introduced was to connect society (Leiner et al. 2009), thus not threatening what it meant to be human. Hence, most of the comparisons between the two technologies lacks support. Yet, one key lesson can be learned from the case of internet. Although the claim made was simply to connect society and individuals, it have had effects that few could predict. This outcome is also possible for AI.

Thirdly is the environmental movement, or more specific how to relate to the threats associated with AI in a similar way as that of climate change. Interviewee A argued that both of these threats are abstract, remote and disputed. This matter because the environmental movement have struggled historically to make their case. Being abstract and remote makes it hard for individual to grasp the issue, it seems either far of in time, space or of a too speculative nature. Hence, it will be hard to create a greater movement around the issues, and it is likely that you will be viewed as somewhat skeptical of progress. Further it is also the fact that it is highly disputed, as apparent from the discussion with regard to nuances in the public discussion. Experts are currently, it seems at least, in no or little agreement on the progress of AI, and their views and predictions vary widely (Müller and Bostrom 2016). In general I believe this comparison to hold true for many cases, and that several lessons can be learned. Yet one important distinction is important to make, namely that of the advantages connected to AI. The comparison to climate change only applies for the risks and issues connected to AI, not the technology as whole. Climate change in turn could be viewed as one of the issues related to industrialization.
7. Conclusions

Based on material from the empirical study conducted, there are few social obstacles
considered major for a continued implementation of AI in a near future, at least in a
Swedish setting. The most relevant obstacle in the near future is that of Increased
polarization of wages and power. The main reason for this obstacle being most
relevant is that it is closely connected to the general development of digitalization. As
the digital technologies emerges, there is a need to address issues of taxation, distribution
and societal structure to account for the changes in how we work. AI, viewed as a
part, or the next step, of digitalization, thus need to address this issue. It can further
be argued that in contrast to earlier steps in digitalization, AI presents the first clear
cases for disruption in the labor market. At this point it is also worth stressing that
non of the interviewees viewed unemployment as an obstacle in the short run. As the
interviewees were representatives of different parts of society, this provides a rather
interesting contrast to the views of the academics.

Another obstacle identified to be likely to affect the development is that of the in-
dividual fears. This connects to a fear of the new technology and an uncertainty of
how to relate to it. In short this issue is connected to the fact that there is a need for
customers of AI services. Someone need to use the systems developed. This fear can
be reduced through education and transparency. Another individual fear is that of the
disruptive nature of AI, or how will it affect me as an individual. Among the public there
is a belief that many jobs will be lost and from an individual perspective this results
in a fear of being able to put food on the table. These fears can be addressed through
good social structures, through welfare solutions, easy ways to reskill or a combination
of both.

Moral and ethical issues were identified as important to develop good solutions,
but should not be considered a major obstacle. Although stressed as important, there are
no clear actors to raise these questions in the short run. Thus, it is unlikely to influence
the development to any greater extent. At the time of writing the conclusion an article
in Wired was published stating that several of the big actors on AI (Microsoft, Facebook,
Google) are pursuing ethical guidelines surrounding their AI solutions. These initiatives
were in part motivated by ethical reasons, but there were also a business case present.
One key remark of these are that of transparency and the fact that it is strange for
those who develop the solutions should regulate themselves. Although welcome, there is
a need for a more public discussion with regard to the issues. This still shows that some
actors start to take moral and ethical responsibility.

Even if this study has been made in a Swedish setting, some of the finding should be
generalizable to an international setting as well. As AI is developing in a global setting,
it is unlikely that actions and thoughts of single countries or regions will have any greater
impact. The exception of this might be the biggest actors such as China and the big US
tech companies. Thus the findings hold a higher relevance for countries or regions of a

---

7Tech Firms Move to Put Ethical Guard Rails Around AI by Tom Simonite -
09:13 18/05/2018
similar position to Sweden.

Looking forward there is a need for much more research in the intersection between AI, politics, society and philosophy. Preferably this should be done in an as cross-sectional way as possible. Although studies from an economic perspective have been made (Aghion et al. 2017; Brynjolfsson et al. 2017; McKinsey and Company 2017), there is a need to start being more specific with regard to policy making. Ethical guidelines and frameworks need to be developed, or tools to be able to develop these in the specific situations. With regard to the technical research there is a need to present pedagogical tools explaining AI, with the main purpose to reduce the fear and uncertainty of the new technology. There is also a need to address several of the technical transparency issues, to be able to explain the decisions being made.

Lastly, we need to start talking about AI. As stated by interviewee A “Even if AI does not develop any more from today’s technology, it will be a revolution”. If this holds true, we need to find a way of relating to this new technology, and how to act in this revolution. Although an important academic discussion, there is a need for the public and society to start talk about it as well. As experts, we need to help ease this discussion by speaking clearly and pedagogically. This is one of the most important issues to overcome if we look a bit further ahead!
8. References

References


A. Pre-Study Questions

As all the interviews were conducted in Swedish, the questions were prepared in Swedish. Below both the original Swedish version (in italic font style) and a translated English version is presented.

Technology

- How do you view AI and the surrounding technology today? *Hur ser du på AI och den kringliggande tekniken idag?*

- How do you think the technology will change within five years? What do you think will be possible? *Hur tror du att tekniken kommer vara om 5 år? Vad kommer vara möjligt?*

- Where do you think the greatest possible uses are of AI within your sector? *Vad ser du som det största möjliga användningsområdena inom ditt fält?*

View of AI


- What do you think are the greatest risks of implementing AI? *Vad ser du som det största riskerna med att implementera AI?*
B. Empirical Study Question

As all the interviews were conducted in Swedish, the questions were prepared in Swedish. Below both the original Swedish version (in italic font style) and a translated English version is presentet.

Definitions of AI

• How would you define AI? Hur skulle du definiera AI?

• Can you give some examples of AI used in everyday life? Kan du ge några exempel på AI i vardagen?

Possibilities and obstacles

All of the following questions were asked within a time frame of three to five years. If asked for a clarification this was further specified as technology currently considered the current edge of academic research.

• What do you think is the greatest possibilities from a wide implementation of AI? Vilka är de största möjligheterna du ser som följd av en vid implementation av AI?

• What do you think is the greatest risks of a wide implementation of AI? Vilka är de största riskerna som en följd av en vid implementation av AI?

• What do you think is the greatest obstacles for a wide implementation of AI? Vilka är de största hindren du ser för en vid implementation av AI?

Specific obstacles and issues

The following four topics were raised as possible obstacles or issues: Unemployment (Arbetslöshet), Increased polarization of wages and power (Löne-/maktskillnader), Gender (Könsperspektiv) and Individual fears (Individuella rädslor). The interviewees were asked to freely associate with the respective term, arguing if they viewed it as an obstacle or not. After an initial reflection, or if the interviewee asked questions, the interviewer clarified the topic and presented some additional perspectives.