MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„When High Earners Earn Less:
Effects of Artificial Intelligence and Machine Learning on the U.S. Labor Market. A Quantitative Study"

verfasst von / submitted by

Vlad-Andrei Mitroi

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 915
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Betriebswirtschaft
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Chem. Dr. Markus Georg Reitzig

Acknowledgements

I would like to express my utmost gratitude to Professor Markus Reitzig, University of Vienna, for his professional advice during the research of this Master's Thesis, for the continuous passion he shows for research, for being understanding and patient, and most importantly, for offering me the trust and opportunity to place my name near his in a scientific paper and in the German publication *"Der Spiegel"*.

Likewise, I would like to thank my family and friends for their continuous support and encouragements.

Thank you!

Table of Contents

1. Abstract

Novel technologies such as Artificial Intelligence (AI) and Machine Learning (ML) become more and more important each day, as companies start to understand their benefits and use them to gain a competitive advantage over their slower-to-react counterparts. The body of research is rich in theories warning about potential effects this technology could have in the economy and in the labor market. Many of these studies though focus on previous types of technological advancements, while the more recent literature argues about the nature and effects of AI or tries to infer past observed effects on the new technology. While the scholars started to notice already almost 10 years ago that these novel technologies have a different nature than their predecessors (Brynjolfsson & McAfee, 2011), there are many conflicting opinions about their long-term consequences: some envision an utopic future built with the help of technology, while others have a more dystopian approach, forecasting an economic fallout.

With some minor exceptions not fully related to artificial intelligence, there is a void of empirical studies to test the researchers' theories. The present Thesis aims to fill part of this void by using Frey and Osborne's (2013) computed probabilities of job computerization as independent variable in order to check if there are any observable effects in the labor market already. The results show that since the increase in AI and ML adoption in 2011, occupations which require lower skill levels in regards to social intelligence, creativity, perception and manipulation – regarded as artificial intelligence's today limitations – see a decrease in wages when compared to occupations which require a high skill level in said areas. The results further suggest that this effect gets more intensive in time and more interestingly, the highly paid occupations with low skill requirements in regards to social intelligence, creativity, perception and manipulation are more affected.

2. Introduction

The question of who will stand to gain and who to lose from the effects of novel technologies such as Artificial Intelligence (AI) and Machine Learning (ML) recently stands at center stage of many researchers and CEOs alike. Historically, technological changes have reshaped the industries and people's occupations alike. Schumpeter's (1975) creative destruction it a well-known concept by now. Each technological revolution of the past has left behind a clear mark, as some social categories benefited from them, while others did not. One common point throughout the technological changes was the fact the winners were usually the more educated social classes, whose high investment in education would not only protect them, but offer them skills complementary to the new technologies, which would further help the flourish in the new epoch (Acemoglu, 1999, 2002; Acemoglu & Autor, 2011; D. H. Autor et al., 2003).

There is a considerable amount of research on the effects of each technological revolution on the economy and the labor market. Some researchers propose that we could explain the effects of the present-day novel technology, the artificial intelligence and the machine learning, by looking at the repercussions of the past (Bloom et al., 2014). Other scholars noticed that AI may be a completely different form of technology, unlike any other we have encountered so far, in that it has the capacity to supplant highly paid high-skilled workers as well (Aghion, Bergeaud, et al., 2017; Aghion, Jones, et al., 2017; Brynjolfsson & McAfee, 2011, 2012).

Despite the vast body of research taking a more theoretical approach, there is a relative void of empirical studies to back up the theoretical assumptions. This Thesis' objective is to try covering part of this research gap and study if the AI has already begun shaping today's labor market, and if so, how. The results further presented are building on the broadly cited paper of Frey and Osborne (2013), who argue that AI has the power to fulfil non-routine tasks (as opposed to the position Autor et al.'s (2003) take), and that the skills that safeguard workers from AI today are those more closely related to social intelligence, creativity, or perception and manipulation. Their argument is that we have to take a closer look at artificial intelligence's (at least present) limitations to know what would a technology-complementing skill of tomorrow look like.

This paper's results show that artificial intelligence has already placed a statistically significant mark on our society, and that it is indeed a different type of mark from what we have encountered before. Some occupations requiring high standards of education seem to be the most prone to be affected by AI. More clearly, those highly paid occupations which require lesser levels of social intelligence, creativity, or perception and manipulation skills. As further developed in the Conclusion section, these effects could have broader further implications on the future generation's motivation to invest in education.

In order to provide a clear guideline, the structure of the paper will follow a typical form of an empirical research, which will be hereafter shortly described. Section 3 will cover the typical historical effects of technological advancements, a short introduction into the history of AI and its present day capabilities and limitations. A succinct presentation of the main ideas arising in the literature shall be presented further. Sections 4 and 5 cover the hypotheses and the methodology used throughout the analysis, while section 6 presents an interpretation of the results. Towards the end, the paper presents a discussion and a conclusion reiterating the main theories and results of the study.

3. Related literature

3.1. Historical effects of technological advancements on labor markets

Almost all important technological changes of the past came with disruptions in the labor market: changing or destroying some jobs while creating new ones. The Austrian economist Schumpeter (1975) defines this process as the creative destruction *("die schöpferische Zerstörung")* and argues it is a means of revolutionizing the economic structure from within. Acemoglu and Robinson (2012) describe the iconic case when in 1589 Queen Elizabeth I refused to grant William Lee the patent for his invention, the knitting machine, out of fear that the new technology would render Her subjects working as knitters unemployed. Novel technologies appearing in the 19th century were mainly skill-replacing ones: the skilled artisans were replaced by factories employing many un-skilled workers (Acemoglu, 2002). People fearing for their jobs in the face of novel skill-replacing technology is thus not particular to recent technological advancements.

Historically, the lower echelon of employees was typically affected by the advent of novel technologies because its relative lack of technology-augmenting skills, while the top echelon mainly stand protected in the face of such technological disruptions (Acemoglu, 1999, 2002). Acemoglu (2002) further notes that the last several decades of the 20th century were mainly defined by advancements in skill-biased technologies, which benefit the high-skilled employees. He even notes an acceleration of said effect in the years closer to the 21st century. Skill-biased technological advancements are defined as changes in technology which favor the skilled workers over their low-skilled counterparts by increasing the productivity and thus, demand of the former (Violante, 2008).

Autor et al. (2003) further argue that it was still the top echelon (i.e. high-skilled workers) who stand protected in front of the early stages of computerization revolution as well. After these early stages, organizations start to exhibit new behavior: they adopt technologies capable to replace skills which were once highly regarded, well-paid, and required high time- and material- investments in education, i.e. analytical skills, pattern recognition skills or even driving skills (Brynjolfsson & McAfee, 2011, 2012). Aghion, Jones, et al. (2017) also note that AI proves to be capable of automating non-routine and cognitive tasks, traditionally performed by high-skilled and well paid workers.

To understand the root of such changes, it is important to understand how recent technologies such as artificial intelligence or machine learning are significantly different than those in the early stages of computerization, and why do they have the chance to affect highly-educated workers, which were typically shielded from such risks.

3.2. Artificial Intelligence and Machine Learning

The term *"Artificial Intelligence"* was coined by John McCarthy in 1956, who began researching this field in 1948. He discovered that *"each aspect of learning and other domains of intelligence can be described so precisely that they can be simulated by a machine"* (Lischka, 2011; McCarthy et al., 2018, p. 1; SAS Institute Inc., 2017; Wisskirchen et al., 2017).

In the 1960s, the U.S. Department of Defense started to research AI and tried to program computers to mimic basic human reasoning. They used early AI technology to help street mapping in the 1970s and even produced intelligent personal assistants as early as 2003, long before mainstream companies such as Apple or Amazon (SAS Institute Inc., 2017).

Artificial Intelligence is currently defined by the Miriam-Webster (2018) dictionary as *"the capability of a machine to imitate intelligent human behavior"*. This capability relies on machine learning technology which mimic how the human brain functions in order to enable computers to learn from experience, to adapt to new variables, or to perform tasks previously reserved for humans. These feats are accomplished by using the sheer force of computational power to process large amount of raw data and to recognize patterns in it (SAS Institute Inc., 2017).

The literature recognizes more types of artificial intelligence. AI is split into two main categories: Artificial General Intelligence (AGI), referred to also as Strong AI, and Artificial Narrow or Weak Intelligence (ANI). ANI merely simulates intelligence, while AGI is able to conduct self-learning, to *"understand"* and to optimize its own behavior based on previous experiences (Wisskirchen et al., 2017). Wisskirchen (2017) underlines that this effect can be exponential when the computer is networking with other machines, therefore learning from others' experiences as well.

Although the narrow AI managed to experience many breakthroughs in the last period, it is built with a specific task in mind. ANI outperforms humans in its niche of expertise, but it is far from outsmarting humans at this point, as it runs under a narrow set of constraints and doesn't fully mimic human behavior and intelligence. Some instances of present day narrow AIs are IBM's Watson, Apple's Siri, Amazon's Alexa, Google's self-driving cars, or different medical diagnosing tools (IBM Watson, 2014; Reece, 2020).

The current AI technology is able to perform almost any task that a person can do with less than one second of thought, but it is limited to narrow, clearly defined tasks (Healy et al., 2017; Ng, 2016). Even if this does not sound as much, it means it can already fully perform tasks such as detecting suspicious behavior using video surveillance cameras, filtering abusive online posts, creating transcripts from audio or video files, translating different languages in real time, assessing whether a potential bank client will likely repay the loan or default, targeting product ads specifically to an interested audience, identifying cancer on MRIs, reading X-rays, etc. (Ng, 2016; SAS Institute Inc., 2017).

Implementing narrow AI technology in an industrial setting enables organizations to use a huge amount of previously untapped data. An example is offered by Zilis and Cham (2016) in the health care industry, where AI issues diagnostic proposals based on patient data, medical images and genomic data. They argue that such technology will further be used in finance, transportation or agriculture, thanks to the volume of raw data available and their economic value.

On the other hand, general AI is capable of replicating human intelligence in such a way that it is capable to solve any kind of problem. It can think, understand, and behave such as to be indistinguishable from a real human being. This form of AI has not been developed yet, in part due to lack of computers with enough computational power and in part due to the lack of fully understanding the way the human brain works (Reece, 2020). Perhaps the most notable attempt of building a general AI is Fujitsu's K, a supercomputer with almost 83.000 processors built by Japanese and German scientists in 2011. Even with such computational power, it took the K computer 40 minutes to simulate just one second of the human brain's neuronal network activity (Hornyak, 2013).

3.3. Theories on innovation and technology's effects on economy

After a brief introduction in the historical effects of previous technological advancements and an overlook over the history and present state of artificial intelligence and machine learning technologies, the present paper will go into a more detailed analysis of the existing literature covering the AI topic and the mechanics behind technology's influence on the labor market and economy.

Researchers' crystal balls seem out of tune in regard to how a future shaped by artificial intelligence and machine learning might look like. There are some with a grim dystopian view of a future where we achieve skill-replacing artificial superintelligence – a computer more intelligent and capable than any human being, which will take over any task traditionally performed by a human. The outcome would be immense unemployment rates and decreased wages all over the skill spectrum (Knickrehm, 2018).

A similar dark future is foreseen by Stiglitz (2014), who argues that innovation led by artificial intelligence is creating inequality of wage and unemployment. He states that the skilled workers' gains would be the total opposite of Pareto efficient, these only seldom compensating enough of the unskilled employees' losses.

On the other hand, there are the utopians which envision a future with a more cheerful type of superintelligence, one which will take on previously human tasks but in the same time provide us with unparalleled wealth levels and spare time to pursue artistic endeavors (Knickrehm, 2018; O'Keefe, 2016). Some, as Ray Kurzweil, Google's director of engineering, even go that far as to claim we will achieve digital immortality (Cohan, 2013). Nordhaus (2015) proposes the idea that once we will reach AGI – the *"singularity"*, as he names it, our economic growth will increase exponentially.

One of the central frameworks of the literature regarding the mechanisms by which the supply and demand for skills is affecting the monetary returns to skills and the change of earnings inequality and polarization of income is the canonical model. This model assumes there are two types of skill needed for different, imperfectly substitutable task groups: high skill-tasks and low-skill tasks (Acemoglu & Autor, 2011). Further, as Acemoglu and Autor (2011) underline, technology – AI or any other previous technology – is considered to complement either one of the two worker groups: high-skill workers or low-skill workers, thus create shifts in demand for said skill groups.

Aghion, Jones, et al. (2017) and Aghion, Bergeaud, et al. (2017) notice a recent change in technology's effect on skill demand: historically, technology has substituted workers performing routine or low-skill tasks, while more recent technological changes – artificial intelligence – has the power to automate non-routine and cognitive tasks, typically performed by highly skilled workers.

Leavitt and Whisler (1958) anticipate that technology would help replace much of the middle managements tasks, and therefore the top management would take on larger proportions of non-substitutable tasks, such as innovation, planning, or creative functions. The middle management will be split into two – some employees moving up the ladder to top management, while others moving to the lower echelon. These ideas stand in line with Acemoglu's (1999) theory that when high-skilled workers are rare on the market and the productivity gap between them and low-skilled ones is small, it is more efficient for firms to open jobs suitable for both of them, as it is difficult to fill high-skill positions. Supply and demand laws dictate that with an increase in high-skilled employees' numbers comes a decrease of their wages. But as soon as the number of high-skilled workers in the economy reaches a certain threshold, it becomes efficient for organizations to split existing positions into high- and low-skilled ones. This leads to increased job polarization and mainly affects middle-skilled jobs (Acemoglu, 1999). These statements are empirically tested by Acemoglu and Autor (2011) for both the U.S. and the European Union's markets, where they find a simultaneous growth for both highly paid high-skill jobs and lower paid low-skill jobs.

On a much recent, but similar note, Tirole (2017) expects AI to further increase the wage gap between high-skill and low-skill workers, presuming that AI would more easily automate low-skill tasks. Further, he also anticipates that organizations would dispense of middle management, flattening their structure, in support of Leavitt and Whisler's (1958) or Acemoglu's (1999) theories. They are not alone in inferring the IT revolution's effects on AI. Bloom et al. (2014) also argue that the development of AI should continue the IT trend of eliminating middle-skilled jobs.

Krusell, Ohanian, et al. (2000) even consider AI as an extreme form of skill-biased technology which has the potential to increase the wage gap by both substituting low-skilled workers and creating additional demand for high-skilled ones, needed to install and exploit novel technologies.

Aghion, Bergeaud, et al. (2017) are more optimistic in their results, discovering some aspects missed by previous theories which were forecasting an increase in wage levels for high-skilled workers only. They ran an empirical study on labor market data from the United Kingdom and found that companies which invest heavily in R&D tend to employ more highly skilled workers, as per previous theories, but also to pay higher average wages across the whole spectrum of employees. Thus in the same time, the low-skilled workers employed by such firms are better off than their counterparts employed by less R&D-intensive firms, and this effect is also more pronounced for them than for the high-skill group. In other words, even when high-skill employees earn more than low-skilled ones no matter in which company, the wage relative to market mean ratio of unskilled laborers has a steeper slope for the latter group (Aghion, Bergeaud, et al., 2017). Working on these findings, the authors theoretically infer the effects from R&D

intensive firms to AI intensive firms under the presumption that AI would not substitute low-skill workers in their entirety.

They further argue that these changes make a remarkable and reliable but yet low-skilled employee even more valuable for an AI-intensive company. Therefore she would be better paid in order to create long term loyalty, to invest in a long term relationship. In the same time, it would be too expensive to hire a high-skilled worker to perform a low-skill task (Aghion, Bergeaud, et al., 2017). Furthermore, Garicano (2000) adds that such an employee would save valuable top-management time by passing up the chain less problems – only the really difficult ones. The paper states that an AI-intensive organization would have a bigger relative potential loss from an unreliable and less able low-skilled worker than a normal organization.

After illustrating the main body of research studying the mechanics by which innovation, computerization, or artificial intelligence and machine learning affect the labor market, the next section will focus on literature trying to quantify these effects.

3.4. Attempts to quantify the effects

One of the most widely cited and prominent study trying to assign figures to AI's effects on the labor market is the one of Frey and Osborne (2013). They step away from Autor et al.'s (2003), who considered a job has a low risk of automation if it included non-routine tasks. Building on the work of Brynjolfsson and McAfee (2011), Frey and Osborne assume that technological advancements such as Machine Learning and Mobile Robotics are substantially different from previous technologies, as they are now capable to take over tasks previously considered impossible to automate, such as cognitive, non-routine, or certain manual tasks. The only skills that cannot be replicated by a machine, according to Frey and Osborne, are those related to social intelligence, creativity, perception and manipulation. Another contrast between the two studies is that Frey and Osborne (2013) only take into account the technical feasibility of automating a task given the available technology, while Autor et al. (2003) consider if it would economically make sense as well (Arntz et al., 2016).

Frey and Osbourne's approach was to use the O*NET database on the task content of 903 detailed occupations from the United States and to ask ML experts to sort occupations as either automatable or not, based on the tasks needed to perform each job. They subjectively selected 70 jobs on which the experts were in full agreement of, and analyzed if they are related to any previously identified engineering bottlenecks. They further develop a model to predict the probability of computerization for the other occupations.

According to the results, 47% of the total U.S. labor market is at risk of computerization over the next *"decade or two"*, with Service, Sales, and Office and Administrative support jobs being at the highest risk (Frey & Osborne, 2013). As it will be explained in greater detail in the hypothesis and methodology sections, the present empirical study uses the probabilities of computerization computed by Frey and Osborne to assess how good these predictions reflect the reality five years after their publishing and as a means to discern whether an occupation require more or less skills which are more difficult to automatize, such as social intelligence, creativity, perception, or manipulation.

Alongside with great attention, Frey and Osborne's (2013) study drew extensive criticism worth mentioning. The most notable focusing on their (a) use of job-level instead of task-level predictions. Arntz et al. (2016) and Autor (2015) underline that tasks can usually be fully automated, and not jobs in their entirety, as a job's task structure is prone to changes in time. Other points of critique lie in the fact that (b) they based the study on subjective and possibly overestimated experts' opinion, that (c) they did not take into account neither that occupations might be adjusted by including more technology-complementing tasks, nor that (d) new occupations might emerge from the new technologies (Arntz et al., 2016). One particularity of the study, though, was never critiqued: namely the fact that a task requiring skills such as social intelligence, creativity, perception, or manipulation would be difficult to be performed by AI.

Addressing some of these critiques, Arntz et al. (2016) replicated the study and applied a task-level method to predict the probability of automation of jobs in 21 OECD countries and found that, on average across the 21 countries, only 9% of jobs are prone to automatization and the percentage is heterogeneous across countries.

Given the fact that there are so many papers on the future of the labor market, there are only a few studying empirical evidence on the effect of novel technologies. Among these, Acemoglu and Restrepo (2017) are analyzing the effect of industrial robots on the manufacturing sector using change in the employment to population ratio as outcome variable. They find that every *"one more robot per thousand workers [is] reducing aggregate employment to population ratio by about 0.18 percentage points (or equivalently one new robot reducing employment by 3 workers) and aggregate wages by about 0.25 percent."* (Acemoglu & Restrepo, 2017, p. 5).

Contrariwise, the empirical study of Autor and Salomons (2018) finds automation not to be employment-displacing, but only to reduce the labor's share in the value-added. However, they focus on data ranging between 1970 and 2007, before significant implementations of AI or ML.

4. Research gap and Hypotheses

Despite the vast array of theoretical research on the technological advancements' effects on the labor market in general and the extensive attention received by the Frey and Osborne (2013) study, there is a relative void of empirical studies to test the theories or the probabilities of computerization proposed by Frey and Osborne (2013) or Arntz et al. (2016). The aforementioned study of Acemoglu and Restrepo (2017) limits its focus on effect of robotics in the manufacturing sector, while Autor and Salomons (2018) restrict their samples to the 1970-2007 period, too early to catch the effects of AI or ML.

This empirical paper aims to cover this research gap by providing a closer look on the effects of recent technological advancements and implementations on the labor market and to test the degree to which the probabilities of computerization proposed by Frey and Osborne can be used to bring light on the market changes.

Even when Frey and Osborne cautiously warned that their estimates are not an actual prediction of job losses, it is only logical that they should hold some forecasting power on the recent changes in the labor market, if not already on employment levels. Moreover, Frey and Osborne state that their probabilities of computerization aim for the next 10-20 years from the date of the calculation. Given the two above statements and the fact that the dataset used in this study is limited to the years between 2011 and 2017 – only 5-6 years after Frey and Osborne's initial calculations, the present analysis will be focused on wage-related effects rather than employment level effects, as the former are expected to arise sooner than the latter. To my knowledge the analysis' focus on wages instead actual employment levels is original in the published literature.

The expected result of the paper's analysis in regards to wages is that occupations requiring a high content of easily automatable skillsets – having lower skill requirements in regard to social intelligence, creativity, perception, and manipulation, as per Frey and Osborne (2013), should have a decrease in demand in the labor market, as they are more and more performed by artificial intelligence. This effect is already mentioned by Brynjolfsson and McAfee (2011), who signaled that even if the financial crisis ended in 2009, companies stopped laying off more workers, but did not rehire back new ones. Their argument is that companies preferred to buy new machines instead of hiring new people.

This decrease in demand in the labor market relative to jobs having high requirements for skills such as creativity or social intelligence would translate to lower relative wages and vice versa, all else being equal.

The analysis controls for the criticism received by Osborne and Frey for their probabilities of computerization by taking into account both the changes in skillsets required for each occupation and industry-specific effects, such as wage-level differences, specific industry growth rates, overall industry computerization level, or industry-specific remuneration practices.

Given all the aforementioned arguments, the paper proposes:

> H1: *After the advent of novel artificial intelligence and machine learning technologies, the lower the skill requirements a job has in regard to social intelligence, creativity, perception, and manipulation, the lower its wage levels when compared to jobs having high requirements in regard to said skill sets, all else being equal.*

The advancements in AI and ML technologies did not suddenly emerge as such in the period between 2011 and 2017, but have been developed continuously. Some researchers argue that the improvements in technology are even following an exponential trend, according to the Moore's law (Brynjolfsson & McAfee, 2011; Wisskirchen et al., 2017). Following this logic, the effect of a job's lower skill requirements in terms of social intelligence, creativity, perception, and manipulation shall have an increased level in 2017 relative to its 2011 level. In other words, the more time we allow for further implementation and development of AI and ML, the more intense the effect depicted in H1 will be. The paper thus proposes:

H1a: *After the advent of novel artificial intelligence and machine learning technologies, the lower the skill requirements a job has in regard to social intelligence, creativity, perception, and manipulation, the lower its wage levels when compared their levels prior to said technological advancements, all else being equal.*

Some papers discussed in the previous sections of the paper point out that the overall winners of AI and ML advents might be the high earners and the top management (Acemoglu, 1999, 2002; D. H. Autor et al., 2003; Bloom et al., 2014; Krusell et al., 2000; Leavitt & Whisler, 1958; Tirole, 2017). While this might hold true, the paper's focus is on the effect of a job's level of computerization on the job itself. Thus, the present thesis will analyze what is the effect on highly paid jobs requiring less skills involving social intelligence, creativity, perception and manipulation. Given that one of the more important goals of companies is to minimize costs while maximizing profits, when facing with the question of whom to replace through AI or ML, I argue that they will choose the option which is more efficient in cost minimizing: the higher paid employee, under the assumption that she can be fully replaced. In other words, when choosing between replacing a low-paid worker versus a high-paid one, companies would prefer choosing the latter, under the assumption that both are fully automatable, all else being equal. The following proposition follows the above arguments:

H2: *Highly paid jobs having lower skill requirements in regard to social intelligence, creativity, perception, and manipulation encounter a higher negative impact on their wage levels, all else being equal.*

5. Methodology

5.1. Data collection and compilation

The present paper will focus its analysis on the U.S. economy first and foremost because of the market's role as a pioneer in implementing novel artificial intelligence and machine learning technologies, and second for the availability and relative quality of the data available.

The dataset on which the present empirical test is built upon was compiled from three different sources.

The bigger part of the dataset was sourced from the census databases provided by the U.S. Bureau of Labor Statistics (abbreviated as BLS henceforth) on their official website (U.S. Bureau of Labor Statistics, n.d.-c).

The datasets are structured around both the North American Industry Classification System (NAICS) and the 2010 Standard Occupational Classification (SOC) system, thus allowing for merging data from different sources using the common classification framework as merging key.

The NAICS system replaced the older Standard Industrial Classification in 1997 as a structure for collection, presentation and analysis of the U.S. economy, and uses a six-digit hierarchical coding system to classify the economic activity in twenty industry sectors, as presented in Table 1 in the Appendix (U.S. Bureau of Labor Statistics, n.d.-b). The first two digits of the NAICS code reveal the industry sector.

The 2010 Standard Occupational Classification (SOC) system replaces the older 2000 edition and is designed to be used by the official U.S. agencies to classify jobs into occupational categories for statistical reasons (U.S. Bureau of Labor Statistics, 2010).

From the datasets available on the BLS website, only aggregated data ranging between the years 2011 and 2017 pertaining to the United States at a national level and industry level was used, while the detailed data regarding each state and each sub-industry was discarded due to higher inconsistencies across the dataset.

Even when using the aggregated data at national level and industry level, it was highly important to keep the full detailing pertaining to the SOC 2010 framework, as Frey and Osborne (2013) have calculated the probabilities of computerization at a job level of SOC 2010 (as opposed to job-group level).

Out of the whole array of variables offered by the Bureau of Labor Statistics, only the following were regarded as of interest to the present study and thus kept: NAICS code, NAICS title, SOC code, SOC title and SOC Group – to be used in merging and analyzing the datasets; Estimated total employment (rounded to the nearest 10 by the original authors), and the mean hourly wage of each occupation across industries – to be used as dependent variables in the analysis. Instances of inconsistencies due to spelling errors across NAICS and SOC titles were corrected prior to compiling the database for the present Master's Thesis.

As the BLS dataset does not contain any information regarding the skill composition required by each SOC 2010 occupation profile, the following information has been collected and merged from the O*NET Resource Center[1] (O*NET, n.d.-a) for further processing and analysis: the skills required for each job profile, their importance, and the proficiency level required for each skill.

The skill importance represents how significant a certain skill is in a certain occupation, and follows a grading system between 1 (minimum) and 5 (maximum). The level of a skill follows a scale between 0 (minimum) and 7 (maximum) and indicates the proficiency level required for each skill (O*NET, n.d.-c).

Given the fact that O*NET use their own O*NET SOC coding system, the official crosswalk to SOC 2010 (O*NET, n.d.-b) was used to merge the datasets. In instances where two or more O*NET SOC codes are translated to the same 2010 SOC code (one-to-many pairing), an average of the two or more

[1] For the compilation of the present database, one O*NET database for each year between 2011 and 2017 was used, namely: July 2011, July 2012, July 2013, July 2014, August 2015, August 2016, May 2017, as a *"July"* release is not available for each year.

coefficients was computed. Unfortunately, a weighted average taking into consideration the number of employees in each O*NET SOC code could not have been calculated because O*NET does not provide for this information.

Finally, Frey and Osborne's (2013) job-level probabilities of computerization were manually reproduced and merged onto the databased using the SOC 2010 as merging key.

It is important to note and explain why the present analysis is based on data pertaining to years 2011 and 2017.

Brynjolfsson and McAfee (2011) signaled that even if the financial crisis ended in 2009, companies stopped laying off workers but did not hire back new workers. They argue that this is because companies preferred to buy new machines instead of hiring new people. Furthermore, they argue that the advent of novel technologies will be[2] exponential, following the Moore's Law. For this reason the present study will focus on 2011 as an analysis starting point.

As explained in further detail earlier in this section, the SOC 2010 framework was used as a key method to compile the databases from different sources. BLS offers one database release for each year, thus seven databases (i.e. one for each year between 2011 and 2017) were downloaded and merged using SOC 2010 and NAICS as merging keys.

In 2018, a new SOC framework was adopted, therefore an official SOC 2010 to SOC 2018 crosswalk should have been used to merge the 2018 data. The crosswalk contained multiple instances of many-to-many pairings between the 2010 and 2018 SOC codes, which would have considerably lowered the data quality. For this reason, the present analysis focuses on 2017 and not on more recent database releases.

The final dataset consists of 9.643 entries containing data on hourly wage, number of employees, number of necessary skills for each job, skill importance, and required skill proficiency levels across all SOC

[2] Brynjolfsson and McAfee issued these ideas in 2011, so we can safely assume they meant the technological advent will follow an exponentially increasing trend after 2011.

2010 job profiles in each main industry sector, for the years between 2011 and 2017. Instances of randomly missing values in the raw datasets were encountered, but no systematic patterns were found.

5.2. Variables

Besides the variables imported as raw data and described in the above section, two further variables were computed, namely the Tech Index proposed by Cannon and John (2007) and the Skill Coefficient pertaining to each occupation; both will be further explained.

Dependent variable

The present analysis uses hourly wage data for 2011 and 2017 as dependent variables.

Independent variable

The independent variable used throughout the present study is Frey and Osborne's (2013) probability of computerization. As argued by Frey and Osborne (2013), the lower the skill requirements regarding perception, manipulation, creative, and social intelligence, the easier it is for a job to be realized by artificial intelligence or machine learning technology, thus the higher the probability of computerization index is.

Control variables

Industry Dummy

A set of industry dummy variables are deployed to strip variation related to each industry sector, as each of them has its specific growth rates, specific remuneration practices, specific computerization rates, etc., which are unrelated to the job itself and its proneness to being executed by novel technologies.

<u>Tech Index</u>

Tech Index is calculated as a proportion of each industry's employees working in Scientific, Technologic, Engineering and Mathematical (STEM) jobs out of the total workforce of each industry, in each year, as per Equation (1) below. The index was first proposed by Cannon and John (2007) as a method of describing each industry, thus it has the same value for different jobs within the same industry, but it varies across years.

For the purpose of this calculation, the list of all occupations included in the STEM group was sourced from the Bureau of Labor Statistics' website (U.S. Bureau of Labor Statistics, n.d.-d).

$$Tech\ Index_i = {\sum Total\ STEM\ employees_i * 100}\Big/{\sum Total\ employees_i} \qquad (1)$$

i = industry.

STEM occupations have predominantly higher mean wages than the U.S. average (Cover et al., 2011). Inclusion of the Tech Index as a control measure aims to strip off variation related only to the nature of the industry and not to the job's susceptibility to computerization as defined by Frey and Osborne (2013).

<u>Skill Coefficient</u>

The Skill Coefficient was computed according to the Equation (2) below as a concentration measure of job complexity and theoretically lies between 0 and 1. A higher Skill Coefficient means that the said occupation is more of an expert job. It is a measure describing each job, therefore it has the same value for the same job in different industries, but it varies across years.

$$\text{Skill Coefficient}_{\text{J,Y}} = \sum_{s=1}^{S}[(\text{Importance}_{\text{s,J,Y}}/5) * (\text{Level}_{\text{s,J,Y}}/7)] \Big/ \text{Number of Skills}_{J,Y} \qquad (2)$$

J = occupation;
Y = year;
S = number of skills necessary for each occupation J;
s = 1, 2, …, S.

The Skill coefficient can be regarded as an average proficiency across the full skill set required for a job in a certain year. The raw importance (level) coefficient was adjusted by 5 (7) in order to account for the different scales of each. As mentioned earlier, the importance (level) was measured on a scale up to 5 (7) points, where 5 (7) is the maximum attainable.

Skills with an Importance or a Level degree at the very bottom of the scale – not important and which require zero proficiency were dropped from the raw dataset before compilation. Thus, the measure Number of Skills used in equation (2) above is a count of all meaningful skills remaining in the dataset, for each job profile, for each year.

With the help of this last variable, the model accounts for differences in the skill composition between different job profiles, which are unrelated to their susceptibility to computerization, but can still influence the wage levels.

5.3. Model

In order to test the three hypotheses presented earlier, I estimate the following models:

Hypothesis 1:

$$wage\ 2017_{i,j} = \alpha + \beta * computerization\ probability_j + \gamma * Tech\ Index\ 2017_i + \delta * Skill\ Coefficient\ 2017_j + \zeta * Industry\ Dummy_i + \varepsilon_{i,j} \qquad (3)$$

Hypothesis 1a:

$$wage\ 2011_{i,j} = \alpha + \beta * computerization\ probability_j + \gamma * \ Tech\ Index\ 2011_i + \delta * Skill\ Coefficient\ 2011\ _j + \zeta * Industry\ Dummy_i + \varepsilon_{i,j} \tag{4}$$

Hypothesis 2:

$$wage\ 2017_{i,j} = \alpha + \beta * computerization\ probability_j + \gamma * \ Tech\ Index\ 2017_i + \delta * Skill\ Coefficient\ 2017\ _j + \varepsilon_{i,j} \tag{5}$$

As the next sections will discuss in further detail, the industry dummy variable had to be dropped from Hypothesis 2 model in order to exploit the richer picture of the phenomenon that quantile regression can offer over the standard linear regression.

6. Results

6.1. Dataset and descriptive statistics

Table 2 below depicts the descriptive statistics of the variables used in the models depicted in Equations (3), (4), and (5) above.

Table 2

Descriptive statistics

	Obs	Min	1st Qu.	Median	Mean	3rd Qu.	Max	NA's
wage_2011	7723	8.3300	15.4500	21.2800	24.6700	30.1500	118.7000	1920
wage_2017	7977	8.9000	17.2200	23.5900	27.4700	33.3300	128.1400	1666
prob_OF	8237	0.0028	0.1000	0.6400	0.5365	0.9000	0.9900	1406
Tech_2011	9643	0.0143	0.8671	3.8966	6.7043	10.6366	29.7883	0
Tech_2017	9643	0.0399	1.1064	4.7278	8.1802	11.2738	33.9494	0
SkillCoef_2011	8426	0.0796	0.1683	0.2149	0.2168	0.2623	0.4506	1217
SkillCoef_2017	8707	0.0796	0.1676	0.2140	0.2162	0.2628	0.3711	936

Note. Values presented in Table 2 are computed using RStudio version 3.6.2 (R Core Team, 2019).

As it can be seen, the hourly wage had an overall increasing trend between years 2011 and 2017, its mean (median) increasing by 11.3% (10.8%) over the course of 6 years.

The maximum value of the variable Skill Coefficient lies at the value of around 0.45 out of its theoretical maximum of 1. This is explained by the fact that for a job to score a skill coefficient of one, it has to require a proficiency score of 7 out of 7 points for all its comprising skills, and all said skill to have an equal maximum importance level of 5 out of 5. In other words, to require the highest level of proficiency and importance for all its required skills, which is virtually possible but practically not. This argument is further confirmed by the fact that no occupation in the dataset has a full score for both importance and level for all its comprising skills.

The highest probability of computerization given by Frey and Osborne (2013) to a job is 99% - theoretically certain to be replaced by technology in the future, while the lowest is 0.28% - no job is entirely safe according to their analysis.

Tables 3 and 4 below describe the correlations between the variables used in the models of the present study. Table 3 covers the variables for the year 2011, while Table 4 those for the year 2017.

Table 3

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3
1. wage_2011	24.67	13.01			
2. prob_OF	0.54	0.37	-.58** [-.59, -.56]		
3. Tech_2011	6.70	7.79	.14** [.12, .16]	-.01 [-.03, .01]	
4. SkillCoef_2011	0.22	0.06	.75** [.74, .76]	-.71** [-.72, -.70]	.03** [.01, .05]

*Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates p < .05. ** indicates p < .01. Table 3 is generated using the apaTables package (Stanley, 2018) in RStudio version 3.6.2 (R Core Team, 2019).*

Table 4

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3
1. wage_2017	27.47	14.57			
2. prob_OF	0.54	0.37	-.57** [-.58, -.55]		
3. Tech_2017	8.18	9.28	.14** [.12, .16]	-.01 [-.03, .01]	
4. SkillCoef_2017	0.22	0.06	.73** [.72, .74]	-.71** [-.72, -.70]	.03** [.01, .05]

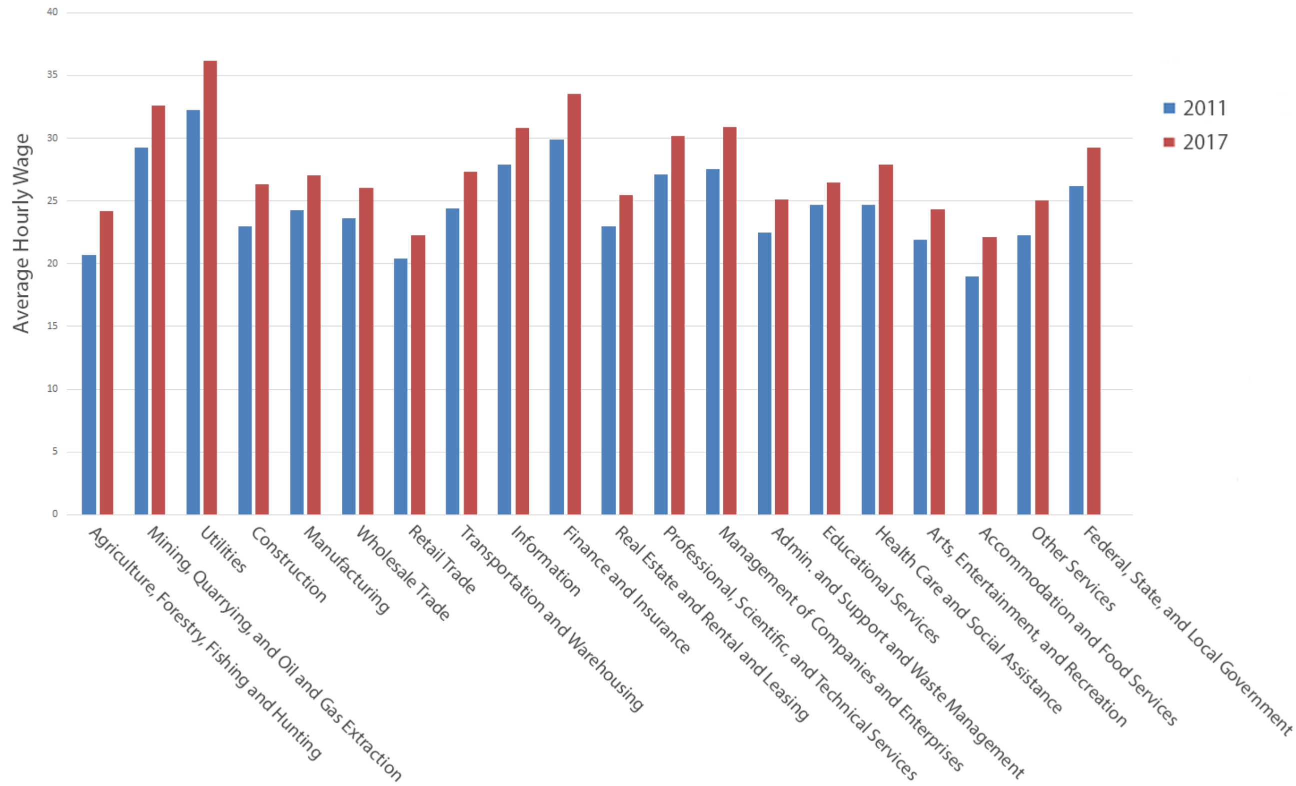
*Note. M and SD are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates p < .05. ** indicates p < .01. Table 4 is generated using the apaTables package (Stanley, 2018) in RStudio version 3.6.2 (R Core Team, 2019).*

It is noticeable that the highest correlation level lies between the skill coefficient and wages and between the skill coefficient and the probability of computerization. A relatively strong positive correlation between the skills required by a job and its wage level is to be expected and is self-explanatory. A relatively strong negatively correlation between skill coefficient and a job's probability of computerization is again to be expected and to some extent indicates the correct calculation of the coefficient – the more skill intensive a job is, the more difficult it is for it to be replaced by a non-general artificial intelligence.

Figure 1 exhibits the wage trend across industries. The wages in the U.S. Market follow a clearly increasing trend between the years in focus for the present study both overall and for each industry sector apart.

Figure 1

Wage development between 2011 and 2017 across Industries

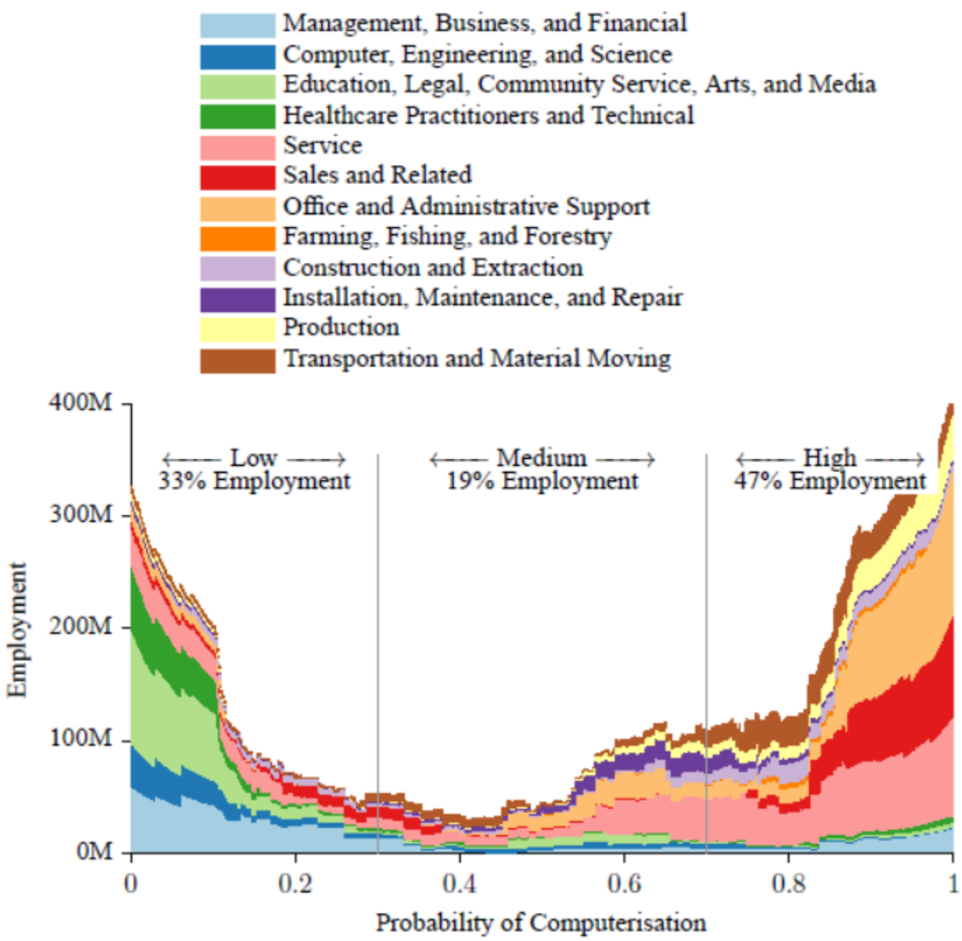


Data source: (U.S. Bureau of Labor Statistics, n.d.-c)

Figure 2 below reflects the distribution of the total U.S. employment over the probability of computerization, as calculated by Frey and Osborne (2013). It clearly shows that an important share of employment is at risk of substitution (47% of U.S. employees have a high risk of computerization) and the most affected occupations are those active in the services, sales, and construction areas.

Figure 2

U.S. employment by probability of computerization



Source: (Frey & Osborne, 2013, p. 37)

6.2. Results for Hypothesis 1 and 1a

Table 5 presents the regression model for Hypothesis 1, as described in Equation (3). As per Frey and Osborne (2013), the lower requirements a job has in regard to social intelligence, creativity, perception, and manipulation, the easier it is for it to be accomplished by a machine, thus a higher probability of computerization index it scores.

The results of the linear regression analysis portrayed in Table 5 confirm Hypothesis 1 and indicate that a higher risk of computerization (prob_OF) tends to lower the hourly wage by 3.61 US $, all else being equal. In other words, an occupation with lower skill requirements in regards to social intelligence, creativity, perception, and manipulation is worse off in terms of wage when compared to a job demanding such skillsets. The prob_OF coefficient bears a strong statistical significance, as well as the model, which manages to explain 58% of the variation in wages.

The Skill Coefficient (SkillCoef_2017) estimate is relatively high (149.15), reason for which is the fact that SkillCoef_2017 index has a maximum value of 0.37 as per Table 2 above, value which tempers down its seemingly high estimate. This coefficient confirms the intuitive fact that an expert-job is better remunerated than one which requires lower proficiency.

Table 5

Regression results using wage_2017 as the criterion

Predictor	*Estimate*	*Std. Error*	*95% CI* [LL , UL]	*t value*	*p value*		Fit
(Intercept)	-5.01	1.18	[-7.33 , -2.69]	-4.23	0.00	***	
prob_OF	-3.61	0.42	[-4.44 , -2.78]	-8.51	0.00	***	
Tech_2017	0.37	0.12	[0.13 , 0.60]	3.02	0.00	**	
SkillCoef_2017	149.15	2.54	[144.17 , 154.12]	58.72	0.00	***	
NAICS21	4.03	0.80	[2.46 , 5.60]	5.05	0.00	***	
NAICS22	3.50	1.34	[0.88 , 6.12]	2.62	0.01	**	
NAICS23	2.34	0.90	[0.58 , 4.10]	2.61	0.01	**	
NAICS31-33	-0.89	0.75	[-2.37 , 0.58]	-1.18	0.24		
NAICS42	-0.55	0.59	[-1.71 , 0.61]	-0.93	0.35		
NAICS44-45	-0.78	1.01	[-2.76 , 1.20]	-0.77	0.44		
NAICS48-49	3.01	0.95	[1.15 , 4.87]	3.17	0.00	**	

NAICS51	-3.29	2.27	[-7.74 , 1.16]	-1.45	0.15	
NAICS52	3.02	0.71	[1.62 , 4.41]	4.25	0.00	***
NAICS53	1.21	0.97	[-0.70 , 3.13]	1.25	0.21	
NAICS54	-8.44	3.31	[-14.91 , -1.96]	-2.55	0.01	*
NAICS55	-1.90	1.46	[-4.76 , 0.96]	-1.30	0.19	
NAICS56	-0.60	0.68	[-1.93 , 0.73]	-0.89	0.38	
NAICS61	-2.10	0.63	[-3.34 , -0.87]	-3.33	0.00	***
NAICS62	0.15	0.94	[-1.69 , 1.99]	0.16	0.88	
NAICS71	-0.06	1.01	[-2.04 , 1.92]	-0.06	0.95	
NAICS72	-1.47	1.11	[-3.64 , 0.70]	-1.33	0.18	
NAICS81	-0.19	0.89	[-1.93 , 1.55]	-0.21	0.83	
NAICS99	NA	NA	NA NA	NA	NA	
						R^2 = .58***

Note. LL and UL indicate the lower and upper limits of a confidence interval, respectively.
One coefficient not defined because of singularities.
*Signif. codes: 0 '***' ; 0.001 '**' ; 0.01 '*' ; 0.05 '.' ; 0.1 ' ' ; 1*
Residual standard error: 9.123 on 6904 degrees of freedom
Multiple R-squared: 0.5792, Adjusted R-squared: 0.5779
F-statistic: 452.5 on 21 and 6904 DF, p-value: <2.2e-16
Values presented in Table 5 are computed using RStudio version 3.6.2 (R Core Team, 2019).

While Hypothesis 1 states that a high risk of computerisation exhibits a negative effect on wage levels, Hypothesis 1a argues that the effect described in H1 got more intense in time, the more artificial intelligence and machine learning technologies were deployed across the U.S. Market.

For this reason, I replicated the previous model for the year 2011. Table 6 below reports the results of the analysis outlined in Equation (4) and confirms Hypothesis 1a. The estimate of prob_OF is -2.82, meaning that a high prob_OF index beared a lesser impact on wages in 2011 than it does in 2017. When adjusted to inflation, 2.82 USD in 2011 is equivalent with 3.11 USD in 2017[3]. Thus, a hypothetical job bearing a 100% risk of replacement by artificial intelligence would have a negative hourly wage impact in 2017

[3] Using the official inflation calculator of the Bureau of Labor Statistics, available at https://data.bls.gov/cgi-bin/cpicalc.pl, last accessed: 14.02.2020. Used dates: January 2011 to January 2017.

with 0.5 USD higher than in 2011. This translates to a monthly wage impact of 84 USD when accounting for 8 working hours per day for 21 days a month. The results confirm Hypothesis 1a.

The 2011 model is statistically significant as well and succeeds to explain 61% of the variation in hourly wages in the U.S. market in 2011.

Table 6

Regression results using wage_2011 as the criterion

Predictor	*Estimate*	*Std. Error*	*95% CI [LL , UL]*	*t value*	*p value*		Fit
(Intercept)	-5.81	1.02	[-7.82 , -3.80]	-5.67	0.00	***	
prob_OF	-2.82	0.37	[-3.54 , -2.10]	-7.65	0.00	***	
Tech_2011	0.38	0.13	[0.13 , 0.63]	2.97	0.00	**	
SkillCoef_2011	137.94	2.17	[133.68 , 142.20]	63.45	0.00	***	
NAICS21	3.18	0.87	[1.48 , 4.88]	3.66	0.00	***	
NAICS22	2.79	1.18	[0.47 , 5.11]	2.36	0.02	*	
NAICS23	1.94	0.78	[0.41 , 3.47]	2.48	0.01	*	
NAICS31-33	-0.86	0.78	[-2.39 , 0.67]	-1.11	0.27		
NAICS42	-0.57	0.57	[-1.69 , 0.55]	-1.00	0.32		
NAICS44-45	-0.05	0.87	[-1.77 , 1.66]	-0.06	0.95		
NAICS48-49	2.74	0.82	[1.13 , 4.34]	3.35	0.00	***	
NAICS51	-0.54	1.32	[-3.13 , 2.05]	-0.41	0.68		
NAICS52	2.86	0.62	[1.65 , 4.07]	4.62	0.00	***	
NAICS53	0.94	0.87	[-0.76 , 2.65]	1.08	0.28		
NAICS54	-7.45	3.09	[-13.50 , -1.39]	-2.41	0.02	*	
NAICS55	-1.60	1.17	[-3.88 , 0.69]	-1.37	0.17		
NAICS56	-0.25	0.64	[-1.51 , 1.02]	-0.38	0.70		
NAICS61	-0.85	0.54	[-1.92 , 0.21]	-1.57	0.12		
NAICS62	0.22	0.83	[-1.40 , 1.84]	0.26	0.79		
NAICS71	0.00	0.89	[-1.74 , 1.73]	0.00	1.00		
NAICS72	-1.35	0.98	[-3.26 , 0.57]	-1.38	0.17		

NAICS81	-0.35	0.80	[-1.91 ,	1.22]	-0.43	0.66
NAICS99	NA	NA	NA	NA	NA	NA

R^2 = .61***

Note. LL and UL indicate the lower and upper limits of a confidence interval, respectively.
One coefficient not defined because of singularities.
*Signif. codes: 0 '***' ; 0.001 '**' ; 0.01 '*' ; 0.05 '.' ; 0.1 ' ' ; 1*
Residual standard error: 7.896 on 6869 degrees of freedom
Multiple R-squared: 0.6088, Adjusted R-squared: 0.6076
F-statistic: 509 on 21 and 6869 DF, p-value: <2.2e-16
Values presented in Table 6 are computed using RStudio version 3.6.2 (R Core Team, 2019).

6.3. Results for Hypothesis 2

For the test of Hypothesis 2, I have employed a quantile regression analysis, developed by Koenker and Basset (1978). The quantile regression models the relationship between the independent variables and the conditional quantiles of the dependent variable, as opposed to the standard linear regression, which models the relationship between IVs and the conditional mean of the DV. This type of analysis gives a more detailed view of the effects of the IVs on the DV across the quantiles of the latter. Other advantages of the quantile regression lie in the fact that the median regression is more robust to outliers than the classic linear regression (Katchova, 2015).

The analysis was implemented with the help of the quantreg package (Koenker, 2019) in RStudio version 3.6.2 (R Core Team, 2019). One of the downsides of using this analysis method is that the NAICS variable had to be dropped, as using dummy variables was not possible. For this reason, the linear regression used further in Table 7 for comparison differs from the one previously used for testing hypothesis 1 in that it drops the set of industry dummy variables. I believe that this is a concession worth making given the advantages of forming a richer picture of the effects across quantiles.

According to the linear regression model in Table 7 below, the mean hourly wage of employees with a 100% probability of being replaced by AI is 3.07 US $ lower than that of employees with a 0% chance of being replaced by artificial intelligence or machine learning, all else being equal.

The quantile regression estimates in Table 7 provide a more detailed perspective than the OLS estimates, and indicate that a score of 100% replacement probability has a larger negative impact at the higher quantiles of hourly wage, i.e. on the high-earners who also have a high substitutability rate, all else being equal.

The employees at the 20^{th} (80^{th}) quantile of hourly wage who also have a high probability of replacement have 2.28 US $ (5.70 US $) lower wages than those with a low substitutability index. Thus, the high earners' wages are more severely affected than the low earners' by the prob_OF index. The difference in intensity between the 20^{th} and 80^{th} quantile for the described effect is as high as 3.42 US $ per hour, meaning that a hypothetical 100% replaceable job at the higher quantile has a monthly income 574.56 USD lower than a 100% replaceable job at the lower quantile.

The linear regression model underestimates this effect at the higher quantiles and overestimates it at the lower ones.

It is important to mention that the regression coefficient prob_OF is statistically different from zero in both the linear regression and the quantile regressions. Besides, it is also statistically different from the linear regression at all the three quantiles presented in Table 7, as the OLS estimate for the probability of substitution lies outside the confidence interval of the prob_OF estimate at every one of the three instances calculated.

Table 7

Regression results using wage_2017 as the criterion

Predictor	*Linear Regression*		*Quantile Regression*					
			20th Quantile		*50th Quantile*		80th Quantile	
	Estimate	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
(Intercept)	-4.87***	0.75	0.21	0.48	1.57*	0.60	-0.23	0.91
prob_OF	-3.07***	0.43	-2.28***	0.32	-4.84***	0.42	-5.70***	0.60
Tech_2017	0.16***	0.01	0.10***	0.01	0.15***	0.01	0.18***	0.01
SkillCoef_2017	152.49***	2.59	96.26***	1.69	120.78***	2.25	165.99***	3.34
		*R^2 = .56****						

*Note. Signif. codes: 0 '***' ; 0.001 '**' ; 0.01 '*' ; 0.05 '.' ; 0.1 ' ' ; 1*
Linear Regression:
Residual standard error: 9.327 on 6922 degrees of freedom
Multiple R-squared: 0.559, Adjusted R-squared: 0.5588
F-statistic: 2924 on 3 and 6922 DF, p-value: < 2.2e-16
Values presented in Table 7 are computed using the quantreg package (Koenker, 2019) in RStudio version 3.6.2 (R Core Team, 2019)

Moreover, the ANOVA analyses presented in Table 8 compare the coefficients from the 20th, 50th, and 80th quantiles and prove that there are statistically significant differences between all three of them.

Table 8

ANOVA analysis – inter-quantile coefficient differences

```
> anova(rq_wage_17.q20, rq_wage_17.q50)
Quantile Regression Analysis of Deviance Table

Model: wage_2017 ~ prob_OF + Tech_2017 + SkillCoef_2017
Joint Test of Equality of Slopes: tau in {  0.2 0.5  }

  Df Resid Df F value    Pr(>F)
1  3    13849  134.72 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(rq_wage_17.q20, rq_wage_17.q80)
Quantile Regression Analysis of Deviance Table

Model: wage_2017 ~ prob_OF + Tech_2017 + SkillCoef_2017
Joint Test of Equality of Slopes: tau in {  0.2 0.8  }

  Df Resid Df F value    Pr(>F)
1  3    13849  297.66 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(rq_wage_17.q50, rq_wage_17.q80)
Quantile Regression Analysis of Deviance Table

Model: wage_2017 ~ prob_OF + Tech_2017 + SkillCoef_2017
Joint Test of Equality of Slopes: tau in {  0.5 0.8  }

  Df Resid Df F value    Pr(>F)
1  3    13849  129.75 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

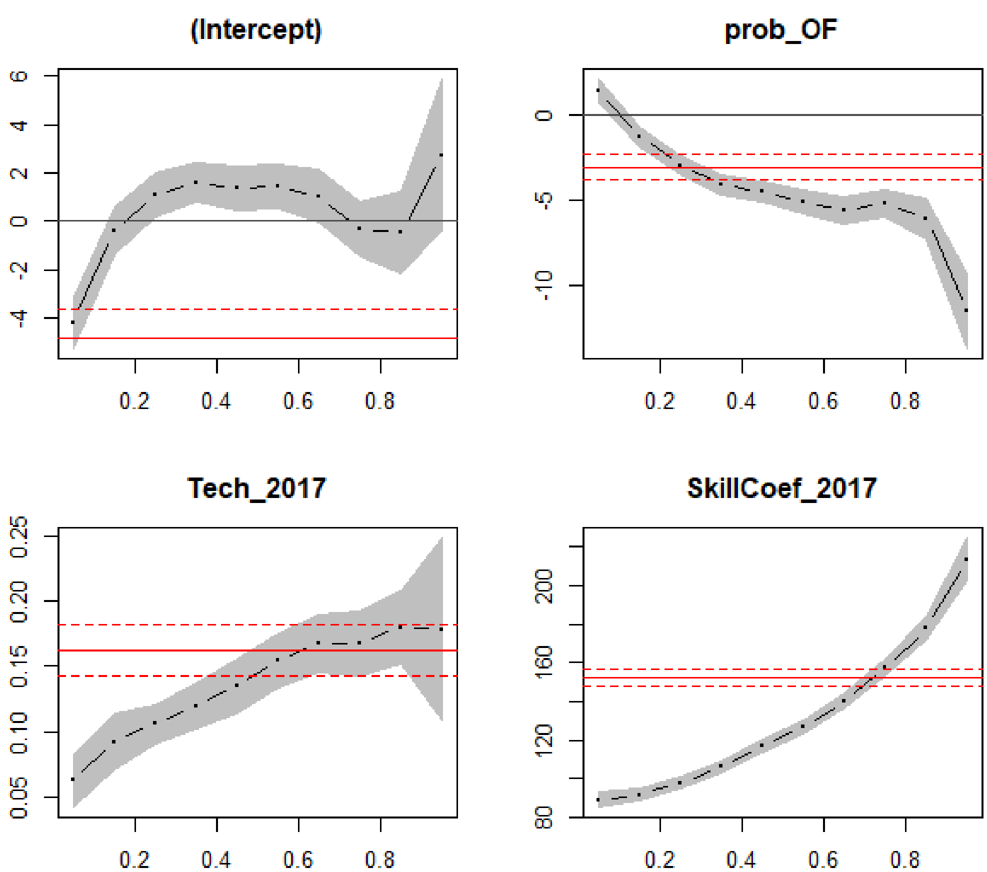
Note. Table 8 is generated using the quantreg package (Koenker, 2019) in RStudio version 3.6.2 (R Core Team, 2019).

The diagnostic plots in Figure 3 depict the quantile regression following the same model as above, but calculated for a sequence of quantiles between 0.05 and 0.95 by an increment of 0.05, in order to give a more detailed graphical perspective than one would get by using only the 20th, 50th, and 80th quantiles. The OLS regression is depicted by the continuous red line, while its confidence interval is depicted by the two dotted red lines in each graph. The values of the quantile regressions are represented by the black lines and points, with their confidence intervals as the grey area.

The prob_OF plot in Figure 3 confirms the calculations above and graphically shows that the quantile regression is statistically different from the linear regression (from zero) with a sole exception at around the 25th (10th) quantile.

Figure 3

Hypothesis 2 diagnostic plots



Note. Figure 3 is generated using the quantreg package (Koenker, 2019) in RStudio version 3.6.2 (R Core Team, 2019).

7. Discussion

First and foremost the analysis presented above shows that, despite all the critiques received by Frey and Osborne (2013), the probabilities of computerization they have calculated do hold some predictive power, which shows even earlier than the timeframe they have estimated. This does not mean that the critique received by the study is redundant, but only that it is the correct approach to focus on skills such as social intelligence, creativity, perception, or manipulation as the key to determine which tasks can be performed by artificial intelligence and machine learning, at least while AI has the current form and limitations.

Second, the results show that, in accordance with studies such as Brynjolfsson and McAfee (2011) or Wisskirchen et al. (2017), the effects of artificial intelligence and computerization do get more intensive in time. Presumably, this increase is due to the fact that (1) the technology itself needs time to develop further and to cover more and more niches, and (2) companies as well need time to adapt to the new technologies, to test it, and to actually adopt artificial intelligence software.

The results of the quantile regression are perhaps the most interesting ones. The researchers are mainly in agreement with the fact that if somebody stands to lose from the advent of artificial intelligence, these are the low-skilled workers. Historically, they were the most affected by new technologies, thus it was expected for this trend to continue. The difference now is that the artificial intelligence is not just an improvement of old technologies, but a new type of technology, one that has the ability to perform highly remunerated non-routine tasks as well, as argued by Brynjolfsson and McAfee (2011), Aghion, Jones, et al. (2017), or and Aghion, Bergeaud, et al. (2017). The results confirm the studies stating that the middle management will be affected: we see that performing a job with low requirements in regard to social intelligence, creativity, perception, and manipulation skills has a clear negative effect on wages across all the quantiles. Contrary to the today's general expectations, this negative effect gets even more intensive the higher the wage level. Thus, the highly paid employees who do not possess the right skillset stand to lose in the face of artificial technologies.

7.1. Possible limitations of the study

The most important limitation of the present analysis is perhaps that the dataset does not contain a measure clearly indicating how AI-intensive the companies in the United States are. Such a variable, at best released by the U.S. Bureau of Labor statistics or some other official entity, would have been critical to the regression models presented in the previous sections. It could have underlined a direct connection between the results and the related technology.

Another important information to have would have been a measure of AI development and implementation across the United States. Given that the present NAICS framework was released in 1997, it does not have an industry classification pertaining to companies active in the artificial intelligence or machine learning fields, i.e. those who produce and distribute such software. These companies are assimilated in the broader Sector 51 – Information, but it is impossible to discern between classical ITC companies and novel AI ones. Such a measure would have helped as well to pinpoint the results to advancements in AI as opposed to the normal IT developments.

8. Conclusions

The question of who will benefit and who will suffer from the effects of artificial intelligence or machine learning technologies is a widely reiterated question today, as technologies which are not fully understood by everybody stand to emerge. As history shows, it is only natural for people to worry about how such a technology will shape the future and whether or not they will stand to gain or to lose.

Artificial intelligence is not just an upgrade of old technologies, but a new kind of technology, one that can perform high-paid non-routine tasks as well as routine ones, as Brynjolfsson and McAfee (2011) seem to have noticed almost ten years ago.

It is now critical for researchers to try to understand the true capabilities and limitations of artificial intelligence and machine learning, and even more so, their consequences on the economy and labor market. One of these effects seems to be that it affects people thought to be protected against technological changes thanks to their high time- and financial- investments in educations. Historically, highly paid occupations were normally associated with high investments in education, and highly educated employees were traditionally shielded from being rendered jobless or receiving wage cuts (Goldin & Katz, 2009).

The canonical model mentioned earlier assumes there are only two groups of tasks, ones requiring high skillsets to fulfil, and others requiring low skillsets; technology is considered to complement either one of the two (Acemoglu & Autor, 2011). Following the present results, we seem to witness a further division of the high-skillset: automatable high skills and non-automatable high-skills. As the literature shows, one cannot state that an entire occupation can be so easily entirely automated, but rather tasks alone – see the critiques on Frey and Osborne's (2013) focus on jobs rather than tasks, but effects on wages can already be observed. If we factor in the observed increasing trend of these effects, one could speculate that at some point in the future we could see not only a decrease in remuneration levels, but actual drops in the employee numbers for the automatable high-skill occupations as well.

In this case, when the wages of workers possessing automatable high skills already get affected by technological developments, the long-term consequences should be taken into consideration. Besides the obvious possibility that the future generations would lose the incentives to specialize for the affected high-skill jobs specifically, they might lose motivation to invest in education altogether, seeing that schooling cannot protect them anymore in the face of change. In other words, they might come to realize that investing in their education is not as profitable as it used to be. In this case, what should the governments and the education institutions do, to mitigate the consequences? This is a question for future research. For now, we can only see that in the face of artificial intelligence, even the high earners end up earning less if they do not possess the right skillsets for tomorrow.

9. References

Acemoglu, D. (1999). Changes in unemployment and wage inequality: An alternative theory and some evidence. *American Economic Review*, *89*(5), 1259–1278.

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, *40*(1), 7–72.

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics* (Vol. 4, pp. 1043–1171). Elsevier.

Acemoglu, D., & Restrepo, P. (2017). *Robots and Jobs: Evidence from US Labor Markets.* 90.

Aghion, P., Bergeaud, A., Blundell, R., Griffith, R., & Market, A.-B. L. (2017). *The innovation premium to low skill jobs*.

Aghion, P., Jones, B., & Jones, C. (2017). *Artificial Intelligence and Economic Growth* (No. w23928). National Bureau of Economic Research. https://doi.org/10.3386/w23928

Arntz, M., Gregory, T., & Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis* (OECD Social, Employment and Migration Working Papers No. 189). https://doi.org/10.1787/5jlz9h56dvq7-en

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333.

Autor, D., & Salomons, A. (2018). *Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share* (No. w24871). National Bureau of Economic Research. https://doi.org/10.3386/w24871

Bloom, N., Garicano, L., Sadun, R., & Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, *60*(12), 2859–2885.

Brynjolfsson, E., & McAfee, A. (2011). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee.

Brynjolfsson, E., & McAfee, A. (2012). Winning the race with ever-smarter machines. *MIT Sloan Management Review*, *53*(2), 53.

Cannon, A. R., & John, C. H. S. (2007). Measuring environmental complexity: A theoretical and empirical assessment. *Organizational Research Methods*, *10*(2), 296–321.

Cohan, P. (2013, June 20). *Google's Engineering Director: 32 Years To Digital Immortality*. Forbes. https://www.forbes.com/sites/petercohan/2013/06/20/googles-engineering-director-32-years-to-digital-immortality/

Cover, B., Jones, J. I., & Watson, A. (2011). Science, technology, engineering, and mathematics (STEM) occupations: A visual essay. *Monthly Labor Review*, 13.

Frey, C. B., & Osborne, M. A. (2013). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. *Journal of Political Economy*, *108*(5), 874–904.

Goldin, C. D., & Katz, L. F. (2009). *The race between education and technology*. harvard university press.

Healy, J., Nicholson, D., & Parker, J. (2017). *Guest editors' introduction: Technological disruption and the future of employment relations*.

Hornyak, T. (2013, August 5). *Fujitsu supercomputer simulates 1 second of brain activity*. CNET. https://www.cnet.com/news/fujitsu-supercomputer-simulates-1-second-of-brain-activity/

IBM Watson. (2014). *Does your business think?* https://www.youtube.com/watch?v=_Xcmh1LQB9I

Katchova, A. (2015). *Quantile Regression—Econometrics Academy*. https://sites.google.com/site/econometricsacademy/econometrics-models/quantile-regression

Knickrehm, M. (2018, January 24). How Will AI Change Work? Here Are 5 Schools of Thought. *Harvard Business Review*. https://hbr.org/2018/01/how-will-ai-change-work-here-are-5-schools-of-thought

Koenker, R. (2019). *quantreg: Quantile Regression*. https://CRAN.R-project.org/package=quantreg

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.

Krusell, P., Ohanian, L. E., Ríos-Rull, J., & Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, *68*(5), 1029–1053.

Leavitt, H. J., & Whisler, T. L. (1958). *Management in the 1980's*. November.

Lischka, K. (2011, October 25). John McCarthy: Der Vater der Rechner-Cloud ist tot. *Spiegel Online*. http://www.spiegel.de/netzwelt/web/john-mccarthy-der-vater-der-rechner-cloud-ist-tot-a-793795.html

McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (2018). A proposal for the Dartmouth summer research project on artificial intelligence (1955). *Reprinted Online at Http://Www-Formal. Stanford. Edu/Jmc/History/Dartmouth/Dartmouth. Html*.

Merriam-Webster.com. (2018, February 5). *Artificial intelligence*. https://www.merriam-webster.com/dictionary/artificial+intelligence

Ng, A. (2016, November 9). What Artificial Intelligence Can and Can't Do Right Now. *Harvard Business Review*. https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now

Nordhaus, W. D. (2015). *Are we approaching an economic singularity? Information technology and the future of economic growth* (No. 0898–2937). National Bureau of Economic Research.

O'Keefe, A. J. (2016, August 11). Ray Kurzweil: The Future Offers Meaningful Work, Not Meaningless Jobs. *Singularity Hub*. https://singularityhub.com/2016/08/11/ray-kurzweil-the-future-offers-meaningful-work-not-meaningless-jobs/

O*NET. (n.d.-a). *O*NET® Database Releases Archive*. Retrieved November 12, 2018, from https://www.onetcenter.org/db_releases.html

O*NET. (n.d.-b). *O*NET Crosswalk Files*. Retrieved November 12, 2018, from https://www.onetcenter.org/crosswalks.html#soc

O*NET. (n.d.-c). *O*NET OnLine Help: Scales, Ratings, and Standardized Scores*. Retrieved November 12, 2018, from https://www.onetonline.org/help/online/scales

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reece, S. (2020, January 31). *What Are The 3 Types Of AI? A Guide To Narrow, General, And Super Artificial Intelligence*. Codebots. https://codebots.com/artificial-intelligence/the-3-types-of-ai-is-the-third-even-possible

Robinson, J. A., & Acemoglu, D. (2012). *Why nations fail: The origins of power, prosperity and poverty*. Profile London.

SAS Institute Inc. (2017, July 27). *Artificial Intelligence – What it is and why it matters*. https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html

Schumpeter, J. A. (1975). Capitalism, socialism and democracy (1942). *J. Econ. Literature*, *20*, 1463.

Stanley, D. (2018). *apaTables: Create American Psychological Association (APA) Style Tables*. https://CRAN.R-project.org/package=apaTables

Stiglitz, J. E. (2014). *Unemployment and innovation* (No. 0898–2937). National Bureau of Economic Research.

Tirole, J. (2017). *Economics for the common good*. Princeton University Press.

U.S. Bureau of Labor Statistics. (n.d.-a). *List of NAICS sectors*. Retrieved November 12, 2018, from https://www.bls.gov/oes/current/oessrci.htm

U.S. Bureau of Labor Statistics. (n.d.-b). *North American Industry Classification System (NAICS) at BLS: U.S. Bureau of Labor Statistics*. Retrieved November 12, 2018, from https://www.bls.gov/bls/naics.htm

U.S. Bureau of Labor Statistics. (n.d.-c). *Occupational Employment Statistics*. Retrieved November 12, 2018, from https://www.bls.gov/oes/tables.htm

U.S. Bureau of Labor Statistics. (n.d.-d). *OES Publications by Topic*. Retrieved November 12, 2018, from https://www.bls.gov/oes/topics.htm

U.S. Bureau of Labor Statistics. (2010). *SOC 2010 User Guide*. U.S. Bureau of Labor Statistics. https://www.bls.gov/soc/soc_2010_user_guide.pdf

Violante, G. L. (2008). Skill-biased technical change. *The New Palgrave Dictionary of Economics*, *2*.

Wisskirchen, G., Biacabe, B. T., Bormann, U., Muntz, A., Niehaus, G., Soler, G. J., & von Brauchitsch, B. (2017). Artificial intelligence and robotics and their impact on the workplace. *The International Bar Association Global Employment Institute (IBA GEI)*.

Zilis, S., & Cham, J. (2016, November 2). The Competitive Landscape for Machine Intelligence. *Harvard Business Review*, 41.

10. Appendices

Table 1

List of NAICS Sectors

Sector 11 - Agriculture, Forestry, Fishing, and Hunting
Sector 21 - Mining, Quarrying, and Oil and Gas Extraction
Sector 22 - Utilities
Sector 23 - Construction
Sectors 31, 32, and 33 - Manufacturing
Sector 42 - Wholesale Trade
Sectors 44 and 45 - Retail Trade
Sectors 48 and 49 - Transportation and Warehousing
Sector 51 - Information
Sector 52 - Finance and Insurance
Sector 53 - Real Estate and Rental and Leasing
Sector 54 - Professional, Scientific, and Technical Services
Sector 55 - Management of Companies and Enterprises
Sector 56 - Administrative and Support and Waste Management and Remediation Services
Sector 61 - Educational Services (including private, state, and local government schools)
Sector 62 - Health Care and Social Assistance
Sector 71 - Arts, Entertainment, and Recreation
Sector 72 - Accommodation and Food Services
Sector 81 - Other Services (except Public Administration)
Sector 99 - Federal, State, and Local Government, excluding state and local schools and hospitals and the U.S. Postal Service (OES Designation)

Note. List source: (U.S. Bureau of Labor Statistics, n.d.-a)

11. Deutsche Zusammenfassung

Neuartige Technologien wie Künstliche Intelligenz (KI) und Maschinelles Lernen (ML) werden von Tag zu Tag wichtiger, da Unternehmen beginnen ihre Vorteile zu verstehen und sie zu nutzen, um sich einen Wettbewerbsvorteil gegenüber ihren langsamer-reagierenden Kollegen zu verschaffen. Die Forschung ist reich an Theorien, die vor möglichen Auswirkungen dieser Technologie auf die Wirtschaft und den Arbeitsmarkt warnen. Viele dieser Studien konzentrieren sich jedoch auf frühere Arten von technologischen Fortschritten, während die neuere Literatur über die Art und die Auswirkungen der KI diskutiert oder versucht, die vergangene beobachtete Auswirkungen auf die neue Technologie zu schließen. Während die Wissenschaftler bereits vor fast 10 Jahren bemerkten, dass diese neuartigen Technologien einen anderen Charakter haben als ihre Vorgänger (Brynjolfsson & McAfee, 2011), gibt es viele widersprüchliche Meinungen über ihre langfristigen Konsequenzen: Einige stellen sich eine utopische Zukunft vor, die mit dem Hilfe der Technologie gegründet wurde, während andere einen eher dystopischen Ansatz verfolgen, und einen wirtschaftlichen Niederschlag prognostizieren.

Mit einigen geringfügigen Ausnahmen, die nicht vollständig mit künstlicher Intelligenz zusammenhängen, fehlen empirische Studien, um die Theorien der Forscher zu testen. Die vorliegende Masterarbeit zielt darauf ab, einen Teil dieser Lücke zu schließen, indem die von Frey und Osborne (2013) berechnete Wahrscheinlichkeiten der Computerisierung als unabhängige Variable verwendet werden, um zu überprüfen, ob bereits beobachtbare Auswirkungen auf den Arbeitsmarkt vorliegen. Die Ergebnisse zeigen, dass seit der Zunahme der Einführung von KI und ML im Jahr 2011 Berufe, die ein geringeres Qualifikationsniveau in Bezug auf Soziale Intelligenz, Kreativität, Wahrnehmung und Manipulation erfordern - die heutigen Einschränkungen der KI - einen Rückgang der Löhne im Vergleich zu Berufen verzeichnen, die ein hohes Qualifikationsniveau in diesen Bereichen erfordern. Darüber hinaus deuten die Ergebnisse darauf hin, dass dieser Effekt von Jahr zu Jahr intensiver wird und interessanterweise die hochbezahlten Berufe mit geringen Qualifikationsanforderungen in Bezug auf Soziale Intelligenz, Kreativität, Wahrnehmung und Manipulation stärker betroffen sind.