The Economic Impact and Ethical Implications of AI

by
Christian Gabor

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Electrical and Computer Engineering
(Honors Scholar)

Presented March 14, 2019
Commencement June 2019

AN ABSTRACT OF THE THESIS OF

Christian Gabor for the degree of Honors Baccalaureate of Science in Electrical and Computer Engineering presented on March 14, 2019. Title: The Economic Impact and Ethical Implications of AI.

Abstract approved:_____

Amol M. Joshi

Artificial Intelligence has gained resurgence in popularity through machine learning methods. This thesis investigates the current use of AI, the impact it will have on the economy and the ethical considerations for developing this technology. In the coming decade AI could increase global economic output over $10 Trillion. AI will improve data analytics, lead to wide scale deployment of intelligent devices and enable tools for scientific discoveries. In contrast, AI may create income disparities between occupations and job volatility for up to half of workers. Popular debate centers on long term consequences of advanced AI; however, current technology poses widescale ethical risks. Policies need to address transparency with the public sphere on the current implications of AI technology.

Key Words: Artificial Intelligence, Economics Impacts, Occupational Changes, Ethics

Corresponding e-mail address: chrisgabor8@gmail.com

The Economic Impact and Ethical Implications of AI

by
Christian Gabor

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Electrical and Computer Engineering
(Honors Scholar)

Presented March 14, 2019
Commencement June 2019

Honors Baccalaureate of Science in Electrical and Computer Engineering project of Christian Gabor presented on March 14, 2019.

APPROVED:

_____

Amol M. Joshi, Mentor, representing Oregon State University College of Business

_____

Karl Mundorff, Committee Member, representing Oregon State University Advantage Accelerator

_____

Bill Smart, Committee Member, representing Oregon State University School of Mechanical, Industrial and Manufacturing Engineering

_____

Toni Doolen, Dean, Oregon State University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, Honors College. My signature below authorizes release of my project to any reader upon request.

_____

Christian Gabor, Author

Contents

1 Introduction

The 1984 science fiction Terminator and Amazon Alexa are both icons of Artificial Intelligence (AI) despite sharing little similarity. Business leaders like Elon Musk warn that AI is currently the largest existential crisis to humanity (Clifford, 2018), others like Yann LeCun, Facebook's director of AI research, say that AI is nowhere close to warranting fear (Vincent, 2017). Experts holding opposite opinions on the outcomes of AI can lead to general misunderstandings on how technological progress will impact society. Machines able to outperform human intellect warrant existential fears of job loss, loss of personal safety from robotic warfare or privacy concerns with algorithms peering into our personal lives. With such large risks at stake, divergent views on future outcomes muddle the public's understanding and could either hinder beneficial progress or could overlook real dangers. The high cost of overlooking these issues merits the discussion of this topic.

Despite concerns of AI taking over our lives, various mainstream technologies, such as advertisement segmentation, movie recommendations and search engine results all use AI as a core technology (Alpaydin, 2016). In recent years, AI has received a high volume of media attention surrounding a new paradigm known as deep learning. A likely reason for the large media coverage comes from optimism that deep learning could lead to algorithms that find cures to diseases, make scientific discoveries, automate tasks in jobs, and give intelligence to everyday machines such as autonomous vehicles. Deep learning involves training neural networks with deep hierarchy of layers how to represent a large volume of data with the goal of generalizing to new examples. For instance, a large (depth wise) neural network could learn to understand objects in images after training from hundreds of thousands of example images with labels. The modern digital landscape with large volumes of data and powerful computational infrastructure enables the deep learning class of

algorithms and research in this domain has surged in recent years for various applications across domains and improving techniques.

The rapid rate of technological advancement in deep learning has caused excitement for future applications and lead to a mysticism on potential societal outcomes. Many hold concern that AI will take away jobs (Solon, 2017). When machines can automate a job at a lower cost than a human worker, workers in that occupation may fear unemployment or reduced wages. Fear of job replacement has existed since the start of the industrial revolution with the Luddite movement (Conniff, 2011). The concern centers on whether increased capital investment in AI technology will increase or decrease the value of human labor and whether the growth of new jobs will outpace job substitution. While current AI has a very slim likelihood of leading a complete robot economy depending entirely on capital income, deep learning could automate specific tasks that could impact a large portion of workers in the economy and lead to widescale displacement. Economists such as Brynjolfsson argue that digital technology including AI will potentially lead to larger societal impacts than the Industrial Revolution (Brynjolfsson & McAfee, 2014). If Brynjolfsson's predictions hold true, new types of roles and jobs may open and inequalities of wealth and income would exist as a transitionary phase, as was seen from a societal change in the industrial revolution. Otherwise, if job growth stagnates while investment grows in technology capital, our developed societies could face wider inequalities in wealth and income.

Some may be unconcerned about job automation due to the belief that humans will always outperform the technological capability of computers. In the past fifty years, the field of AI has undergone multiple periods of excitement, heavy investment, but later fallen short of investor expectations. These periods have been known as AI winters, giving some the

belief that AI will fall behind again. However, the field of AI has undergone a paradigm shift with the use of deep learning on computer hardware orders of magnitude better than decades ago. The past shortcomings do not diminish the capabilities of current and future AI as modern algorithms now perform tasks such as image recognition and robot actuation that could impact many occupations.

Separate debates exist for the future outcomes of AI, such as the impacts on society and jobs or the dangers of AI in weapons. The ethical concerns for AI span multiple dimensions, including risks involved with technology expected to exist now or in the near future and the risks involving uncertainty in the dangers of long-term powerful AI. With present technology, the ethical implications of AI involve risking human safety, privacy, autonomy and freedom, as the malfunction of autonomous robots could harm people and privacy issues increase as algorithms improve at monitoring citizens through data collection and image recognition. If AI technology improves in the future the point of superseding human intelligence, AI could fight humans for resources and could carry out tasks misaligned with our wishes. Debate as to whether this concern should merit current fear exists between some philosophers and researchers, but often captures wide attention from the public in the portrayal of AI in movies, novels or the projection of immediate societal risks involved with technology.

This paper explores the societal impacts of AI involving two questions. First, what is the economic impact of AI? Second, what are the ethical implications of large-scale AI adoption? For this study, the definition of economic impact centers around value added to industries and effects on labor. The study of ethical implications assesses the short-term and

long-term risks associated with AI adoption. This paper examines these two questions in detail in two separate sections following an overview of AI technology.

2 Technological Overview

The term “AI” may evoke imagery of human thoughts emulated on computers or super-intelligent, free-thinking robots. In popular media, this appears in the personification of androids that serve as characters in movies or shows. While science fiction may offer insight for the ethical implications of machine intelligence, the formal study of AI represents itself differently.

AI research mainly explores the design of artificial agents that behave rationally in their environments. Russell and Norvig define AI as, “an artificial system that carries out tasks intelligently” (Russell & Norvig, 2009). To be considered intelligent, the agent must make rational actions based on information available, such as sensor readings from an environment, or user input. This definition allows AI to be implemented as simply or as complicated as needed. Research typically aims to find the most effective systems that can perform well at given tasks. While this definition may seem broad, the practice of building systems with this aim results in algorithms that work over a general domain. For instance, research on expert systems in medicine resulted in the development of expert systems for financial industries, physical sciences, and other industries (Yang, 2006). Research on neural networks has led to image recognition, speech recognition and advanced reinforcement learning agents, as described in section 2.2.

The definition of AI does not depend on whether a machine thinks like a human, as this metric would be subjective and difficult to quantify. A potential misconception for is that, “AI does not exist yet because humans are still smarter than machines.” The issues of this statement come from anthropomorphizing the definition of intelligence. This would

imply that intelligence requires human thought. Following this reasoning, we would consider any other non-human animal as unintelligent.

One thing that currently separates humans from AI algorithms is that humans possess the ability to think rationally in many situations and relate concepts together across diverse fields. For example, learning mathematical concepts from physics classes also plays over into understanding concepts in other science subjects. Mathematical concepts themselves are abstractions from real observations. This sort of abstract thinking allows humans to possess superior intelligence and understand the general nature of the world. If an AI could do this, we would consider it having the ability to perform tasks over wide range of diverse environments, carrying over knowledge between environments. Deep learning, a popular approach today which involves complex neural networks, has held promise for pushing us closer to this goal, but more progress is required before we can reasonably judge humans and AI together.

Despite human superiority in generalizing across tasks, AI has been useful for performing specific tasks very well. In some cases, these algorithms can reach abilities superior to humans, such as in the game Go (DeepMind, 2017). Focusing on task-based intelligence provides more utility than mimicking human thought, as task-based algorithms allow for systems to expand from the confines of human frailties and complement our abilities. In general, the academic definition of AI refers to a class of algorithms used to perform tasks rationally in a given environment and stimulus.

Practices used in AI have changed over the last seventy years. Section 2.1 explores the history of AI and how paradigms have changed in response to various factors. Section 2.2

looks more in depth at the deep learning algorithms developed in recent years and the inspiration for their use.

2.1 Brief Historical Overview

AI has origins as far back as the 1940s just as the first electronic computers were created. Neural networks, which are a popular research focus today, were proposed by Alan Turing and others over seventy years ago (Copeland & Proudfoot). Yet, despite the theoretical groundwork at the time, computer scientists could only work within the limits of the machines available. After a history of exponential improvements in the computer industry, known as Moore's Law, the algorithms of AI have caught up with past challenges such as giving AI the ability to interpret images from raw pixels (Gopnik, 2017). Only within the last decade have neural networks reached superiority over older paradigms in AI applications.

An apocryphal summary of AI history is that of experts saying, "a computer can never do X," then as hardware and algorithms improve, AI researchers show "computers can now do X!" Armstrong et al. show that of common AI predictions, the time between prediction and arrival ranges from six to over seventy years with the most frequent occurring after 16-25 years (Armstrong, 2014). Critics may dismiss applications as impossible since the time lag between proposals and implementation can span beyond the lifetime of those proposing the AI system. AI often holds a mystical view between future capabilities that computers may someday achieve, such as Alan Turing's ideas of neural networks in computers (Copeland & Proudfoot), and formalized methods for computing, such as many of the standard machine learning algorithms originating from statistics. A trope in the AI community is that as time goes on, some people stop calling this software intelligent but see it as just an algorithm, for example, search algorithms used in standard game playing systems. Whether algorithm or futuristic hope, AI has served as a foundation for computer

science since the dawn of digital computers, representing our search for the limits of computer applications and algorithms for software to build on.

Before the formal induction of AI in the 1950's, research drew inspiration from biological systems. In 1943 Warren McCulloch and Walter Pitts created an artificial neuron inspired by the neurons in animal brains (Roberts, Clabaugh, Myszewski & Pang, 2000). This researched occurred only seven years after Alan Turing proposed his idea of the universal computer in 1936, and still over a year before the invention of the first large scale computers, the Colossus in 1943 (Copeland, 2004) or the ENIAC in 1946 (Martin, 1995). In 1948 Alan Turing proposed training neural networks to carry out tasks as a potential computer architecture (Copeland & Proudfoot). Aside from inventing the modern computer, Alan Turing is famous for the Turing Test, which was one of the first tests to determine a machine's level of intelligence. In this test, a machine is considered to possess human-level intelligence if it can trick another human into believing that he or she was talking to a real person. To avoid the human from having visual bias, the machine must sit behind a terminal and exchange conversation through text responses. Researchers today use more quantitative metrics to rate the capability of AI algorithms since Turing's Test was an informal and anthropocentric assessment of AI. The test relies heavily on the human interrogator and there are tricks to make software use colloquial language to appear human in nature. Based on Alan Turing's proposals, both the design and testing of AI was human inspired in the 1940's.

AI diverged from neuro inspired computing in the 1950's when it was formalized to work on the hardware available. The newer formal work focused on creating systems that performed logical reasoning, which was well suited for general purpose computers at the

time. In 1956, AI became a formal study in the United States after John McCarthy[1], Marvin Minsky[2], Oliver Selfridge[3], Ray Solomonoff,[4] and Trenchard More[5] met at Dartmouth College with the intent to accomplish various tasks using AI (Dartmouth AI Conference, 2006). They proposed that in two months they could solve language understanding machines, neural networks, self-improvement algorithms, abstractions of sensory data, and creativity. Ultimately, these tasks were not accomplished at the conference and presented larger challenges than anticipated. The failure had computer scientists reassess what was possible to run on computers. As a result, researchers garnered the fact that computers were well suited to perform logical inferences faster than running neural networks or learning algorithms. Engineers could hand code computers with heuristics and rules of the environment to perform complex tasks like playing board games, a major feat at the time. For example, the value of pieces on a chess board can be assigned so that a computer can compute the value of taking an opponent's piece. In chess, taking the queen or another other high valued piece typically puts the opponent at a significant disadvantage. The program could follow a strategy to remove high valued pieces from the board and possibly look ahead at the opponents best move to develop a winning strategy, commonly implemented as tree search algorithms for games (Heinz, 2000).

One reason logic-based systems outperform any learning algorithms at the time was that computer hardware was significantly limited by today's standards. Over the development

[1] Notable computer scientist, developed Lisp programming language, coining the term "artificial intelligence" and received many awards including Turing Award in 1971.
[2] Famous cognitive scientist in domain of artificial intelligence, also known for his books such as The Society of Mind, recipient of Turing Award in 1969.
[3] Known as "father of machine perception" (Spark 2008), known for work on parallel computing and pattern recognition.
[4] Inventor of algorithmic probability, work leading up to Kolmogorov Complexity, and originator of machine learning (Rathmanner & Hutter 2011)
[5] Computer scientist who worked at IBM's Watson Research Center

of digital computers, data storage and CPU processing speed served as a primary constraint to computer applications. If a task could be done using a simple calculation or quick retrieval of information, this would greatly cut back on cost and time of the program. The cost of a single bit of computer memory in the late 1950s was $0.01 USD (CHM Revolution). Moore's law has observed the phenomena of computer components involving transistors, including memory devices, doubling in capacity every 18 months (Encyclopedia Britannica) which significantly reduces cost and improves performance due to the compounding exponential growth. To put this in perspective, a back of the hand cost of a standard four Gigabyte stick of RAM (~$20) used today for a laptop would cost $345 Million USD in the 1950s[6], or potentially on the order of $1.5 million in the 1970s (Adee, 2008). A neural network used for image recognition today involving 100 megabytes of memory to just hold onto the weights would run on memory worth $8.4 Million USD in 1950 or $36,000 in 1970. This example does not factor the cost of a modern high-speed CPU or GPU into consideration. Simply put, the applications possible on computers during the first wave of AI were of a different class than those possible today.

Another reason logic-based approaches were popular was that for the first time, computers could perform interesting software aside from pure number crunching. For instance, programmers found ways to assign lookup tables in databases and make chatbots that could attempt the Turing Test. For instance, if s human operator types a phase "hello", the chatbot could look this word up and extract a human crafted response such as "hello, how are you?" This program would be much easier to program than training a machine to learn language and build its own responses. To the user, this program seems intelligent enough to

[6] 8 bits x 4 gigabytes x (1024)^3 bytes/gigabytes x $0.01/bit

hold a conversation, even though it does not understand language beyond human programmed rules to pass the Turing Test (Hutchens, 1997). Despite the simplicity of systems like these, this type of computing could still offer novelty to users in the consumer space and bended the minds of how people viewed computer possibilities. In the personal computer revolution, instead of using computers for number strict crunching, consumers of video games or interactive applications may have found this minimalistic AI entertaining even if the software followed hard coded rules.

In the 1970s and 1980s, the cost of computer storage decreased to the point where researchers developed Expert Systems that would store large sets of knowledge on specific subjects. These systems behaved as "experts" on a topic and could answer questions such as, "if a patient has a lump on his neck, what illnesses could it be?" The system could respond with "it may be a cyst, a spinal hernia, ..." These systems store information from experts such as doctors and use hard coded rules to link questions to answers (Halim, 1990). The primary advantage these had over was individual was that these systems could combine a wide source of knowledge into one place. It had knowledge from not one, but many doctors. Before the world wide web, these systems were valuable in various settings. As one of the first expert systems, MYCIN stored knowledge pertaining to bacterial infection provided by the Stanford Medical Research team and could predict infections in patients. The system was able to recommend antibiotics to patients based on body weight and the infection present (Costea, 2016).

Large companies saw value in expert systems and began to adopt these in the 1970s and 1980s (Yang, 2006). Industries found use of these systems for health care, chemical analysis, credit authorization, financial management, corporate planning, oil and mineral

prospecting, genetic engineering, automobile design and manufacture and air-traffic control (Yang, 2006). Historically, breakthrough points in AI have beguiled firms to overestimate the capabilities of new algorithms, lending to swings in investor funding. On separate occasions of investment withdrawal, the field of research went into what was known as a “AI Winters”. One of these winters occurred in the 1980s during the time expert systems. During the cold war, the United States invested in developing text translation algorithms to cipher Russian Documents (Yang, 2006). After a combination of papers released underpinning the frailties of expert systems, many organizations, including the US government, cut funding in the expert system research (Yang, 2006). Pessimism grew as people believed that AI was doomed to never perform useful work; investors eventually turned down research that went under the label of AI. During this AI winter, the attendance at the National Conference of AI dropped by nearly 80 percent after the year 1986 to 2002 (Menzies, 2003).

Limitations of expert systems came from the difficulty of programming knowledge representation for subjects with unclear rules. For machine translation, colloquial phrases were a common failing point. An infamous example was the translation of, “the spirit is willing, but the flesh is weak." Expert systems of the time translated this to, "the vodka is good, but the meat is rotten.” (Russel & Norvig, 2009). By the 2010s, deep learning algorithms could translate text and recognize speech over multiple languages; yet, this required thirty years of computer speedups and paradigm changes to machine learning (Alpaydin, 2016).

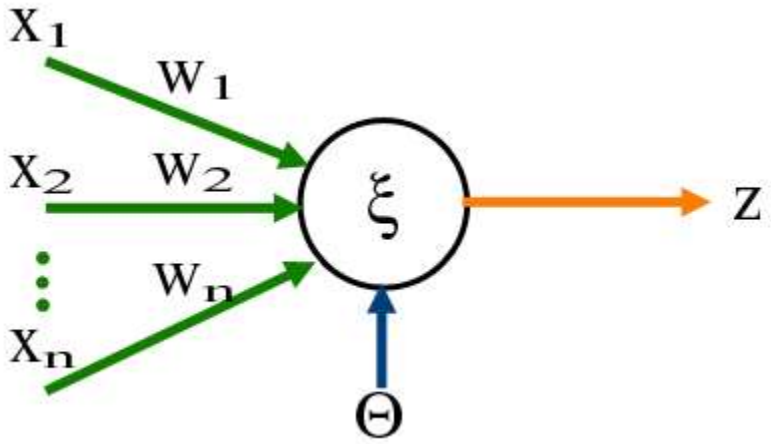
Researchers hid behind new discipline terms such as “machine learning” or “informatics” after the AI winter of expert systems (Yang, 2006). Researchers had to become more conservative on the promises they made for their technology (Bostrom, 2014). In

retrospect, the market correction over expert systems required researchers to rethink strategies to accomplish intelligent systems. In the recent years machine learning has moved the forefront of research today, offering solutions to the shortcomings of more hard coded systems of the past, such as machine translation or computational perception. It is possible that we will discover even better algorithms; however, as observed in the history of AI, the computational hardware improvements centered around computer infrastructure have allowed theory to find way into applications. Whether the future progress of AI will depend primarily on data resources, improved algorithms, cognitive science or improved hardware remains a debate (Müller & Bostrom, 2014); however, the approaches dominating the field of modern AI, as explained in the following section, would not be feasible without the hardware resources available for data, processing speed and memory. This point plays a key role in later arguments for societal change, namely Moore's law, in later sections of this paper.

2.2 Deep Learning

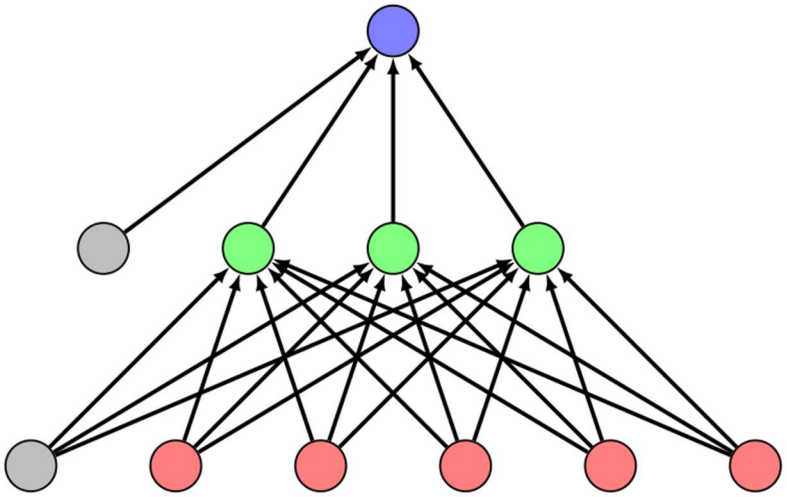
Artificial neural networks have been explored since the beginning of AI research but have not become mainstream until recently. Individual units, or neurons, figure 2-1, are implemented as functions that take a weighted sum of multiple inputs and produce an output. Training an artificial neuron involves updating the weights on the input connections to minimize the output error (LeCun, 2015). Building a neural network naturally follows by combining individual units into a larger network architecture, figure 2-2. The structure and representation of a neural network is not complex and early AI researchers have significant interest in this approach, yet there were seldom successes of early neural networks, likely due to memory constraints, limited access to data, training speed, uncertainty on best architectures and lack of optimization techniques, such as the credit assignment problem (Haykin, 1994). Algorithms using other approaches converged faster and performed better.

Figure 2-1 Perceptron, Artificial Neuron



X represents input values and weights (W) are adjusted through learning algorithm to change behavior of output (Z).

Figure 2-2 Combining into Simple Neural Network



A step towards improving neural network use was discovered in the 1980s when multiple research groups developed an algorithm known as backpropagation to train multiple layers of neuron connections (Roberts, Clabaugh, Myszewski & Pang, 2000). This algorithm depends on the properties of partial derivatives from multidimensional vector calculus and remains in use today for most deep learning algorithms. Researchers were then equipped with a mathematical algorithm to train a neural network. Despite this success, the AI research community was amidst a funding low in the 1980s and 1990s after the investment withdrawal of the AI winter. Other challenges remained for neural networks including slow training time, architecture design, limited training data and network sizing.

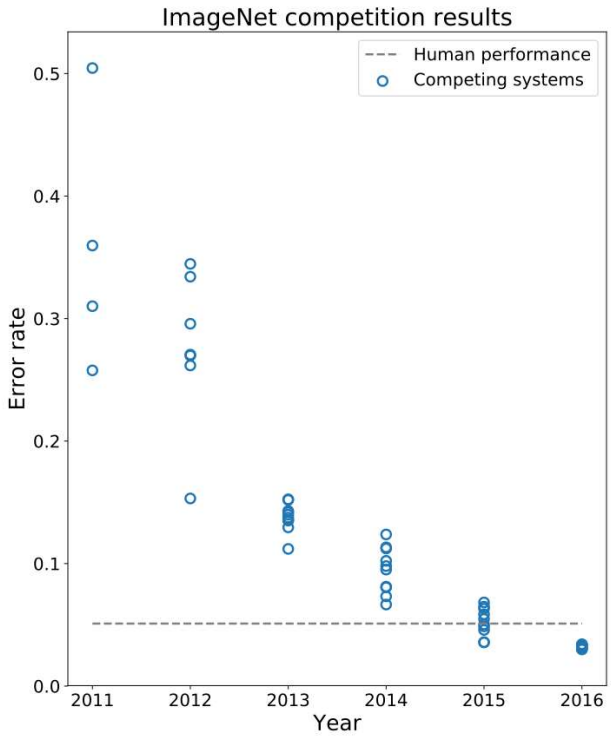
At the same time of the AI winter, one of the largest technologies that humankind has seen grew exponentially in the 1990s and 2000s. The World Wide Web caught on quickly and grew from 14 million to over 3 billion users between 1993 and 2015 (Julia Murphy & Max Roser, 2017). The emergence of the World Wide Web has enabled ease of collecting and sharing data. The amount of digitally stored data has grown in parallel with the world wide web, with a sea of images, video, audio and text files available across the world. This

large amalgamation of data from the Web fuels data driven applications such as machine learning since these algorithms involve statistical analysis of data. Today, the infrastructure of the Web also serves as a platform for cloud based deep learning applications.

In 2004, DARPA held the DARPA Grand Challenge to recruit teams to build autonomous cars. That year in 2004 no car drove more than 7.5 miles in the challenge before failing; however, after the prize pool was increase in 2005, five teams completed the 142 mile desert course (Thrun, et al., 2006). As an interesting historical turning point in AI, the winning team in 2005 relied heavily on machine learning to help the car stay on course (Thrun et al., 2006). The success of a machine learning approach to creating a robust autonomous vehicle for this challenge indicated to the AI community that a machine learning may lead to promising results. This was not the first machine learning application; however, the media attention of this event was a signal to many that a new spring could appear around the corner since the results were staggering between the two competition years.

Machine learning characterizes algorithms from linear fit algorithms to neural networks; yet, today deep learning comprises a large majority of research focus. Popularity for deep learning occurred after the 2012 ImageNet Large Scale Visual Recognition Challenge in which the winning algorithm achieved superior results among the contestants. Figure This algorithm used a convolutional neural network and accelerated the training time with a modern graphics processor unit (Russakovsky et al, 2015). Since then, deep learning involving neural networks has triumphed over previous held challenges in computer science causing a growth in research interest. Figure 2-3 shows the improvements from 2011 to 2016, indicating the rapid progress in this field with use of deep learning. The winner of 2012 outperforms the runner up contestants with close to half the error rate.

Figure 2-3 ImageNet Error Rates by Year



In 2016, a deep learning algorithm developed by DeepMind overwhelmingly beat one best players in the world at Go, the 9-Dan professional Lee Sedol in a match 4-1. Go was considered impossible to win with computers because of the 1.741×10^{172} game states of on 19x19 board (Number of Possible Go Games, 2016). Masters of Go claim that winning requires intuition and they will spend their lives from childhood practicing the game. Further improvements on AlphaGo resulted in a win against the best player in the world, Ke Jie in match of 3-0 (Cheng, 2017). In 2017 AlphaGo Zero improved from the original AlphaGo by learning the game without any human encoded knowledge (Silver et al., 2017). This version was able to beat the original AlphaGo 100-0.

The advantage afforded by deep learning is that the system learns entirely end to end. In contrast to expert systems, human operators do not tune the system after storing

information or rules; instead, developers create a model and adjust starting parameters before letting it learn entirely from data. Deep learning typically learns from raw values, such as pixel values in an image, and aims to match the output as close as possible to the correct label with the inputs provided. For instance, image recognition deep learning involves supplying hundreds of thousands of images containing desired labels and the system will use back propagation to reduce the error between the output from each image to the actual image label. In the Go example, billions of boardgame moves with a history of eventual outcomes are fed to the neural network for it to approximate which moves lead to wins. Once the network is good at predicting known outputs, the network can be used to predict over new inputs that follow similar input-output relationships as the examples in the training data.

Deep learning possesses the property where design architectures lead to multiple application domains with relatively small implementation cost. The upfront research of developing good architectures serves as a main challenge; but once designed, the knowledge representation depends on the data and computational resources. No human operator needs to manually update rules in the system when new data arrives. Specific neural network architectures, such as those targeting image recognition, need only data in the end user application without a redesign of the system model. A system model could be designed once, but then deployed in devices worldwide, continuously learning from data without human intervention.

The power of deep learning comes at a cost over older approaches like expert systems. While deep learning can automate knowledge representations, human understandability and testing has become more difficult (Tickle, 1998). Currently, a large challenge exists in tracking down the decisions of neural networks since the output depends

on many nuanced connections tuned during training time. This is analogous to the challenge of reading specific thoughts in the human mind by observing neuron activity. Another modern challenge of deep learning is the high volume of inputs needed to train neural networks which requires heavy computation requirements. Nevertheless, the large supply of data from the digital age on the Web today lessens the issue of data, and computer power has grown significantly since the time when speed of parallel computation was of main concern. In general, developers of deep learning may face various constraints: large data requirements, high cost in training time, importance of neural network architecture selection, importance of selecting training techniques, unpredictable behavior as black box systems, hardware constrained in some real time domains. At the current state, developers and engineers cannot throw deep learning at any problem and great results. While the research has improved and shown promising applications, these applications were the result of careful engineering practice put toward the competitions. While the term neural networks invokes imagery of human thought, these algorithms behave in a narrow domain and have limitations within the bounds of the tasks in which scientists and engineers design them. Moreover, the areas in which deep learning can find value will depend on the constraints of end applications and market adoption. Without taking heed to these remarks and promoting AI as a panacea to all problems, nothing limits AI from once again reaching an AI winter, as described in the previous section. The history of AI merits careful consideration of what capabilities exist among current technologies and to remain transparent on the limitations.

Apart from the constraints of modern AI technology, deep learning offers novel capabilities that previous techniques lacked. Namely, AI can learn end to end with minimal human oversight as well as learn to perform tasks that we do unconsciously that have been

historically difficult to formalize. The follow examples illustrate the various approaches in deep learning that give rise to machines capable to perform tasks in vision, language and taking actions that learn without humans determining the rules a priori. Whether these systems perform previous tasks at greater capability may matter less insofar as providing computers with novel capabilities that were previous reserved to human perception. Table 2-1 summarizes function of each deep learning techniques described in the following subsections.

Table 2-0-1 Deep Learning Techniques

Technique	Function	Examples	Novel Capability
Convolutional Neural Network	Computational perception, spatial data	Image recognition	Reuse of generalized representations
Recurrent Neural Network	Natural language processing, temporal data	Voice recognition, stock market predictions	Abstract memory representations
Deep Reinforcement Learning	Estimating value of states and actions, decision making	Robotic actuation, competitive decision agents	Abstract understanding of environments for AI decision making
Generative Adversarial Network	Generating novel structures, behaving creatively	Drug discovery, graphics generation	Abstract representation of task for novel proposals

2.2.1 Convolutional Neural Networks

Convolutional neural networks are a type of neural network typically used for image recognition but have found general use for spatial pattern recognition. Deep learning has become popular after the CNN AlexNet won the 2012 ImageNet Large Scale Visual Recognition Challenge (Russakovsky, 2015). This contestant achieved a 15.3% error rate compared to the winner of the year prior at 26.2% error (Krizhevsky, 2015). With a 40% decrease in error, AlexNet showed that deep learning holds superiority over previous methods. In the years following, the top contestants in the ILSVRC have used convolutional neural networks, reaching error rates as low as 3.57% in the 2015. (He, Zhang, Ren & Sun, 2016). What was once a difficult task, Convolutional Neural Networks have shown that computers are capable of high accuracy image recognition.

The accuracy rates of convolutional neural networks may be impressive, but the intrinsic value lies in giving novel capabilities to computers. Depending on the hardware used to implement the models, computers could be used to recognize objects in the real world at speeds orders of magnitude greater than humans. The visual reaction time for a human takes approximately 200 milliseconds on average (Jain, Bansal, Kumar, & Singh, 2015). However, convolutional neural networks programmed into an FPGA fabric have been shown to recognize 170,000 images per second (Ahn, 2015). This comparison implies that CNNs could recognize images over 30,000 times faster than humans. This example does not take into account sizes of larger networks and relies on specialized hardware; however, even at rates 10 or 100 times faster than humans, computer-based image recognition will lead to novel applications. Automated quality control in manufacturing facilities, high speed

document searching and collision avoidance in autonomous vehicles are several potential areas for this novel technology.

In addition to image recognition, convolutional neural networks can perform abstract pattern recognition tasks such as classifying sounds in audio spectrograms (Wyse, 2017). These spectrograms are typically two-dimensional plots of frequency vs time, where the intensity of the pixel represents the magnitude of the frequency at a time point. These may be bird calls or other sounds bytes that the neural network can learn to recognize. An added value of CNNs is that inference is not limited to visible light image recognition as they could potentially learn to recognize patterns in infrared lighting, ultra-violet or any other frequency domain. This may aide in the discovery of new planets in space data. In general, convolutional neural networks are finding use in domains where spatial relationships hold significance in each data instance.

2.2.2 Recurrent Neural Networks

In contrast to convolutional neural networks, recurrent neural networks are a class of neural networks used to find patterns in time dependent information. During inference time, the current input as well as previous inputs are used to determine the output of the neural network. Implementations of recurrent networks take multiples forms with some networks involving logic gates to determine the states to save. In the time domain in which convolutional neural networks may not perform effectively, recurrent neural networks have shown greater efficacy.

In previous years, RNNs have disrupted previous attempts to solve language translation. Some example applications of RNNs include language translation, speech recognition, genomic sequencing and financial predictions. Time dependency appears often in language; for example, the sentences “he loves orange” and “he ate an orange” contain the same spelling of “orange” yet have different meanings. The ability for previous words in the sentence to influence the final output aide in the prediction of the sentence meaning. While previous investment went into expert systems and other AI techniques for translating text, neural networks have shown the ability to form abstractions between languages to reconstruct sentences between multiple languages.

2.2.3 Deep Reinforcement Learning

The previous two neural network architectures involve learning from ground truth examples, such as labeled images or pre-translated sentences. This process is known as supervised learning. Reinforcement learning takes a separate approach to learning through the process of exploring actions and observing rewards. Inspiration for reinforcement learning comes from animal psychology research. Early demonstrations have shown the ability for rats to pull levers to receive food through reinforcement learning (Yale University, 1948). In the real world, many problems may require complex planning and abstract relationships; actions that animals can learn but computers often struggle with. Reinforcement learning narrows in on these fundamental challenges of AI.

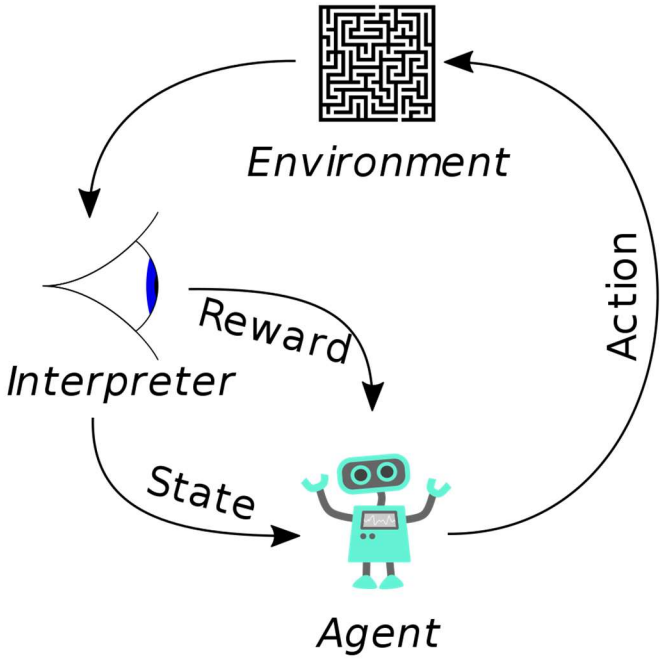
Reinforcement learning branches from dynamic programming, a frequent paradigm of computer science algorithms. The fundamental principle derives from the Bellman equation which involves a recursive function that computes the value of a state accounting for the maximum future payoff that can follow. A modified adaption of the Bellman equation is given for reinforcement learning in equation 2-1. This equation is simply for illustrative purposes, while an entire class of algorithms centers around a technique called Q-learning. Russel & Norvig or Sutton & Barto (provide an in-depth exploration of Q-learning relating to the Bellman equation (Russel & Norvig, 2009), (Sutton & Barto, 1998).

Equation 2-1 Simplified Bellman Equation

$$Value(state, action) = reward\_now + Value(next\_state, best\_next\_action)$$

The equation guarantees that the absolute optimal action at any state is the action that maximizes the right-hand side. This means that if an agent possesses the correct values for this equation at every state, it will be able to always choose the best action at any time. The issue is that the right side of the equation requires a recursive computation of the *Value* function for the states that follow. Figure 2-4 gives an example of an agent taking actions in the world after it observes its state and reward, which are part of the function. Its goal is not to maximize the immediate reward, but the cumulative reward over its series of actions. For this reason, the agent needs information on all the possible following actions and take the path of actions that will result in the highest sum of rewards.

Figure 2-4 Agent Taking Actions



Finding the *best_next_action* can become challenging if the number of future states or action possibilities is large. Dynamic programming can solve this equation for the optimal action by saving each state as a lookup table; however, saving each state requires a tractable

number of state action pairs. When there are many possible states, the use of dynamic programming can be intractable, which occurs for most real-world tasks.

In a game of go, the 19x19 board has over 10^{172} possible states that could be encountered (DeepMind 2019). At each player's turn there are potentially over a hundred other moves and a win may occur over a hundred moves later. This number of states dwarfs the number of the 10^{80} atoms in the universe, preventing any useful attempt to store all states in a computer. The game of Go has been said to be a game of intuition; experts can look at the board and visually recognize good moves from bad. As visual recognition serves as a dominant strategy for Go, it logically follows to leverage the recent success of neural networks to recognize valuable board states. Deep learning can use a neural network as a function approximator for the value of each state. In turn, this method replaces the right-hand side of the bellman equation given above. The more accurate the network can approximate the value function, the more likely it can choose the best actions.

Game playing may seem to be a trivial pursuit for AI; however, many real-world tasks can be modeled in the same fashion as a game. For instance, the controls of a manufacturing robot include states, actions and rewards; therefore, reinforcement learning can be used to play to maximize the efficiency of the robot as if it were playing a game (Russel & Norvig, 2009). The innovation of neural networks in reinforcement learning enables function approximation in tasks that were previously intractable due to a high number of states. It is likely that deep reinforcement learning will help bridge the gap from AI existing in controlled, discrete environments to more complex environments of the real world.

2.2.4 Generative Adversarial Networks

Deep learning is not solely limited to inference and function approximation. Clever utilization of neural networks has led to systems that generate novel creations. Generative adversarial networks (GANs) exemplify this technique and have shown success in generating new images made from raw pixels that look like real photographs to the human eye. The principle of generative adversarial networks involves using a convolutional neural network to recognize good images from bad images proposed by the generating agent, acting as the adversary (Goodfellow, 2014). A separate module learns to create fake images that get past the adversary that appear real. NVIDIA's generated face results (Karras, Aila, Laine, & Lehtinen, 2017), using progressively trained generator and adversary networks in tandem.

While the technique of generating novel images and media could provide interesting applications, this approach has found applications in other fields. A computational biology research group has used GANs to generate candidate oncology drugs to treat cancers (Kadurin, et al., 2017). Generative algorithms could be used to generate candidate designs for researchers to explore, reducing cost of research and accelerating the pace of scientific discovery. This approach could be extended into physics, materials science, engineering and various other domains where novel designs derived from data may lead to promising results.

2.3 Summary

The field of AI covers many algorithms, each involving tradeoffs that lead to better performance in some domains than others. Currently, no one algorithm fits every task. Nevertheless, approaches in AI to solve one problem can impact many domains. For instance, the approach of a game playing AI could also solve a problem in a business application. Expert systems rely on engineers to encode knowledge into the systems; however, the expert systems provide use to multiple industries. Machine learning offers an alternative technique over expert systems through training automatically with various tradeoffs in cost, challenges and complexity. Like expert systems, the algorithms to train a machine learning model can work across various types of data, allowing a machine learning algorithm to fit to many applications. A significant reason for transition to machine learning and data driven AI stems from improved computer resources to allow complex and dynamic learning systems.

Deep learning has attracted attention to AI in recent years with the ability to perform tasks previously intractable for computers, such as Go and accurate computer vision. Each form of deep learning comes with unique advantages over previous techniques that could redefine the potential uses of these technologies. Convolutional Neural Networks (CNNs) gained fame through accomplishing special tasks; however, a more abstract novelty of CNNs involves the ability to reuse features extracted from data across domains. For example, the lower layers of CNNs learn basic edges, shapes and work across multiple domains, whether image recognition of animals or lesion detection in medical settings. The ability to retrain these networks quickly through transfer learning could lend to faster development of other AI applications in various industry domains. Recurrent Neural Networks (RNNs), express the

unique ability to encode abstract memory representations that could prove important for complex intelligence and various tasks in the real-world involving time dependent information. Deep reinforcement learning opens potential for AI systems to function in the real world instead of simple environments, with the ability to draw reasonable interpretations of new situations. The robotics field involving actuation and decision making in complex environments may improve in tandem with deep reinforcement learning, allowing machines to perform tasks previously reserved to humans. Generative Adversarial Networks (GANs) provide one example of a deep learning approach to exploit abstract understanding of data and then use these abstractions to create proposals for new creative works. For instance, the deep learning system could automate the generation of various proposed structures for a later agent or human to pick from. While each described AI still acts within tight bounds of narrow tasks, these examples illustrate the novel capabilities of AI possible on modern computational resources. Future applications may involve various combinations of these principles in new methods of AI or find use in applications across various domains.

The deep learning examples come with drawbacks and barriers for adoption that can impact the feasibility of implementing these approaches. These include the black box nature of neural networks, the high cost in computational resources and example data, and challenges in inventing optimal network architectures. We have not created general purpose, superior AI systems simply through use of neural networks. It may take a convolutional neural network hundreds of thousands of images to learn to recognize objects; on the other hand, humans can learn new skills or knowledge only after viewing few examples. Potential research directions may investigate improving the ability of deep learning systems to transfer knowledge across domains. Other methods to explore are generating an optimal neural

network architecture for an arbitrary task instead of humans designing the network architecture. Nevertheless, the ability for deep learning algorithms to be distributed across many domains and learn automatically from the data can lead to novel applications. Models that have been trained on large cloud computer servers with large datasets can be downloaded on small devices and perform tasks across the world.

The deployment of widespread machine learning systems will more likely result in many specialized systems instead of one standalone intelligent system. These systems may learn to interoperate to accomplish complex tasks, acting collectively to form a more abstract system. The concept is not too unfamiliar to human society. Humans form into similar patterns in society, as most specialize in specific tasks and domains. There is no human that has mastered every domain, field and job; yet as a large collection, everyone working together in the world on their own tasks enables a thriving society. A large collection of specialized agents could perform complex tasks to some extent as well as a single intelligent system. As deep learning enables systems to have greater perception of the world, interoperating agents could reshape our world of technology.

3 Economic Impacts

AI entails the creation of systems that rationalize and perform logical actions. While some examples may serve as novelties, such as in AI played games, the end applications involve the automation of valuable tasks. Since current AI systems perform specific tasks and do not offer a one size fits all solution, the value gained from AI depends on the ingenuity of developers to find and fit the need of people or businesses. After firms overcome barriers of adoption and find target applications, the potential impact could be on the order of trillions to tens of trillions in improved world output. According to the 2018 McKinsey Global Institute report, AI will increase gross world product by 16 percent by 2030 which would amount to an increase of $13 trillion to GWP (Bughin et al, 2018). This group expects that 70 percent of firms will adopt some form of AI technology (Bughin et al, 2018). Gillham predicts that the economic impact may add an additional yearly GWP of $15.7 trillion by 2030 (Gillham, 2017). An older estimate claimed that AI could lead to a total impact of nearly $20 trillion including the impact of knowledge workers, manufacturing jobs, robot control and autonomous vehicles (Manyika, et al., 2013). This analysis also claims that 230 million knowledge workers and 320 million manufacturing workers may be impacted AI by 2025 (Manyika, et al., 2013). Frey and Osborne looked at the jobs most vulnerable to substitution and predict that 47 percent of jobs could become automated from AI (Frey & Osborne, 2014). Whether these predictions play out in the next decade, firms across domains may find use of AI and this could impact organizational structure, technological innovation and labor.

Since the formal inception of AI, investment has focused on the creation of toolkits and methods to automate tasks. This aims to free up human resources and enable new capabilities for machines. For instance, the United States military was invested in the

translation of Russian documents during the cold war to fill in the demand for translators and reduce time to decipher codes. Even the simplest tools result in profound societal impacts such as reducing the cost to produce products or services, increasing the supply of products, enabling humans to take on new roles, and enabling the creation of new products and more tools. The more widely applicable the tool, the larger the impact on society.

Because tools possess the intrinsic property of automating tasks from human labor, concern arises for the outcome of jobs. This sentiment exists throughout history during times of technological development, such as during the introduction of textile machines and the response of the luddite movement. This also appears in stories that symbolize the fall of man to machine, such as in the folktale of John Henry, a folk hero who attempted to outpace a machine for railroad nail driving and died of exhaustion in the process. When tools automate labor, the skills specific to these tasks lose value. The concern now for AI comes from the fact that both knowledge work and manual work is at stake for automation. The present outcomes depend on the areas developers and businesses target and how those affected will respond. With the development of new technologies and roles, new jobs will open; however, the issues on those displaced may be complex.

Evident in the past AI winters, many problems in the domain of AI have non-trivial solutions. Some of the promises of the 1980s have only been recently solved more than thirty years later. To some this marks the end of the AI winter since machines can now do what was not possible before: translate sentences between languages, recognize images, win games that involve intuition. The previous economic shortcomings of AI arose from brittle, expensive systems; yet, while the deep learning algorithms today get around this issue, using deep learning in the 1980s would have not been possible. The exponential increases in

computer power has been a driving factor to enable the research focused today on machine learning. Household computers can now run billions of operations a second and handle gigabytes of data at a time, an unimaginable size in the 1980s. Without knowing the capabilities that computers would reach, one would struggle to predict the possible applications. Previous academics claimed computers would not perform tasks that involve object recognition or sentence processing (Brynjolfsson & McAfee, 2014). Years later this prediction has been proven wrong with Convolutional Neural Networks and Recurrent Neural Networks.

Moore's law has observed the exponential improvement of computer hardware over the course of decades (Schaller, 1997); however, the observation makes no prediction on the fundamental uses of computers as they improve (Brynjolfsson & McAfee, 2014). Exponential improvements in technology fundamentally change the end applications. Take for instance a car doubling in speed each year. The first generation of car would travel at one mile per hour, the second generation at two miles per hour, the next at four, and so on. The first several generations of this car would have no use, as people could easily walk or run faster. However, after just 30 years, in the same lifetime as many people this car was first made, this car would travel at 10.7 million miles per hour, faster than the speed of light. A car traveling this fast may have broad implications for our society as it may transform how we value distance and time. Predicting applications of a technology at this level would have been hard to quantify observing just the first slow iterations. Moore's law is not a concrete law since eventually the laws of physics prevent further improvements; however, the historical growth of computer performance has disproved previous long-term predictions of end computer applications, such as intractability of image recognition (Brynjolfsson &

McAfee, 2014). Section 2 explained the paradigm changes of AI resulting from improvements in computer resources, such as the exponential improvement in cost of computer memory, processing power and the wide volume of datasets now available on the world wide web. In the early years of AI, deep learning and neural networks were considered intractable, but the exponential changes in computers have enabled new classes of algorithms to find use in real applications.

Brynjolfsson and McAfee argue that the development of computers and digital economy puts us in a time of economic restructuring of equal or greater magnitude to the industrial revolution (Brynjolfsson & McAfee, 2014). During the 2008 recession, companies in the United States recouped losses quickly but primarily invested in equipment instead of labor. Despite the growth in GDP, the average income of workers did not increase, and unemployment persisted (Brynjolfsson & McAfee, 2014). Other economists hypothesized the stagnation of middle-income earnings to a lack of technological progress since our old labor models correlate technological progress to middle income earnings. However, Brynjolfsson and McAfee attribute the phenomenon to rapid technological progress indicative of large-scale economic restructuring. The hypothesis that technology has lagged contradicts the observation of Moore's law which shows the opposite phenomenon of compounding exponential growth over the past decades (MacCrory, et al., 2014). If a digital revolution explains the economic changes of the coming decades, AI could play a large role in providing novel applications and value to firms invested in digital technology.

Section 3.1 explores how AI can impact firms and section 3.2 provides examples of industries that may adpot AI technologies. Section 3.3 explores how AI may disrupt or improve occupations.

3.1 The Value of AI

The economic impact of AI may exceed upwards of adding $15-20 trillion to GWP through several estimates (Bughin et al, 2018), (Gillham, 2017), (Manyika, et al., 2013). An estimated 70 percent of firms may adopt some form of AI technology (Bughin et al, 2018). This number represents various changes to industries and relies on firms to adopt AI technology to fit specific needs. As explained in section 2, AI software cannot directly replace human knowledge work; rather, AI can automate specific tasks. Fitting AI to these tasks requires engineers and developers to delegate AI and configure systems to perform as intended. These tasks require knowledge of the systems and often significant computational resources. The adoption strategies, market fit, internal usage and digital resources all play into the value added to specific firms. The main innovations of modern AI that add value to firms, described in this section, involve efficient data analytics, widescale deployment capability and potential to accelerate scientific discovery. In today's digital economy, these factors play a large role in enabling novel applications and efficient workflow processes.

The most recent innovations of AI involve machine learning and deep learning, explained in the case studies of the technology overview section. The difference between deep learning systems over expert systems involves automatic learning and adaptability rather than hard coding information into systems. This difference enables improved data analysis, reaching deployment across many devices, and uncovering insights for scientific discoveries. For example, training for deep learning image recognition may occur on the cloud through pooled data and computational resources, but the deployment of the image recognition software may be deployed across many devices such as smartphones or

autonomous vehicles. As introduced in section 2, researchers seek the use of deep learning for scientific discoveries such as cancer treatments (Kadurin, et al., 2017).

The use of AI will rely on the creativity of developers and entrepreneurs. Even though deep learning can learn from raw data, this does not mean it will automatically learn what we want it to do nor what we of ask it. To create a valuable product, developers will need to create model architectures that fit the application requirements. No matter what capabilities AI achieves, it will still require humans to work out what the tool does and how it fits customer needs. The power of these tools will enable novel applications but will require humans to decide the outcomes.

Gilham finds that by 2030, the consumer side impacts may account for most of the economic value (Gillham, 2017); however, the initial beneficiaries of AI technologies will likely include larger companies (Bughin et al, 2018). Gilham shows that the production side of the economy will experience the greatest cumulative sum of economic impact averaged between 2017 and 2030, but this is due to the supply side of the economy experiencing more immediate efficiency gains (Gillham, 2017). Larger companies could find the lowest barrier to adoption due to existing large volumes of data, computational infrastructure and efficiency improvements from analytics on data exceeding human readability. In general, businesses may see benefits sooner than consumers from this technology since businesses tend to collect more data about everyday operations than average individuals. Many of the algorithms and tools developed for research purposes may transfer directly to business data analytics since data may already exist and fit the correct formats. Example applications include improved marketing and customer analytics, operational cost analytics and other internal tools. This requires marginal effort to implement since businesses can invest into employees or

consultants to build out tools using well known algorithms to match company needs. The applications will serve primarily to offer complimentary, time saving analysis for firms and managers to make better decisions. Executives and analysts will still need to interpret the information to make business decisions; nevertheless, this will add value to various industries even if in a passive way.

Firms that collect large volumes of data during normal business operations will find value in AI sooner than other businesses. These companies may include web-based companies that gather data from customers for advertisements or predictive services, such as Amazon.com, Netflix, Facebook, Google or other companies that gain value from analyzing the data of consumers on their platforms. Due to the large volumes of data they already possess, they do not need to change data collection patterns before finding value of adopting AI tools. Besides digital companies, other larger corporations may generally find more justification for the cost of data collection and AI analytics with economies of scale. The data analyzed can extend beyond customer information to areas such as large-scale production or warehouse efficiency, employee patterns or financial analysis.

Firms that have the capital to invest in robots and automation for business operations will gain value from AI developments. Improvements to the technology will enable more complex robotics and efficient operations. The 2013 Robotics-VO roadmap determined that the greatest limiting factors of robot deployment include software design, control, planning and perception (A Roadmap for U.S. Robotics, 2013). Since then, deep learning algorithms have offered solutions to these issues; for instance, convolutional neural networks solve problems in the perception domain and reinforcement learning can enable control and planning. The software that runs the robots serves a large technical constraint, indicating that

AI can fill this need and innovate robotics. Firms that develop and deploy automation will find that AI integrates into the technology stack of these firms, leading to cost savings in manufacturing, agriculture and transportation industries. Larger firms will have greater economies of scale to afford the machinery, large volumes of data and more computational resources develop robust systems with deep learning; thus, the initial integration of AI will occur within large firms. These developments will serve mainly as internal tools to reduce operational costs. These observations can be seen today as large web-based companies collect data from billions of users and use AI tools to automate applications, such as Facebook, Google, Amazon and other online giants. This trend will continue as more firms decide to collect data for AI analytics.

Data analytics for internal businesses will play as the immediate cost savings of AI; however, the deployment of deep learning models can lead to innovations that will have a much larger impact. The prospect of deploying intelligent algorithms on tens of billions of devices worldwide implies that widescale growth may occur for applications in AI. Each device may retrain locally after deployment through collecting new data and adapting to the environment, uplifting the cost of human oversight for adaptation. Potential applications span from learning personalized settings on user devices to adapting to local tasks in work environments. Some techniques exist to exploit this ability such as uploading data back to a central location or learning individually on devices, but this area will likely evolve over time as it finds more value across devices. This property was not possible before deep learning since the models were rigid and human developers were required to update features when the environments changes. This innovation still requires human ingenuity and imagination to

capture; however, the impacts will change how we interact with and view technology in society.

Smaller firms will find immediate value of deployable models with the ability to integrate and modify pre-trained networks into new applications. Once a neural network architecture has learned from a general dataset, developers can retrain the network to fit a more specific local task, known as transfer for learning. This means that the smaller firms do not need to gather more data to train a general model nor release proprietary data so long as the firm can acquire the pretrained network. For instance, a small biotechnology company may download a network pre-trained on drug molecule structures and then use this network on proprietary data for new applications. Transfer learning requires fewer computing resources to retrain the network and works for local, proprietary data. This decouples experts from building model architectures for small companies, reducing costs to develop products. Companies may be able to build an initial viable product with little data using a pre-trained model. Over time the model can improve as more data comes in from many devices. Additionally, the firms themselves may find use in retraining models to fit local data for internal tools. This trend will gain speed as more models become available on the public domain or for leasing purposes.

The prospect of wide deployment will lend be applicable to consumer-facing applications. Deployable AI could reach personal devices, transportation, customer services and operational equipment that people interact with daily. This will spur growth in the technology sector spanning from features added to new consumers devices to complex decision agents working alongside humans. Consumers may customize product orders over online transactions and communicate with devices for personalized experiences. Modern

examples include Amazon's Alexa, which has been forecasted to add $11 billion to this firm's revenue by 2020 (Kim, 2016); yet, the devices of the future need not be limited to voice command home assistants and could someday include actuated robots in homes. The digital revolution complements AI technologies in the consumer space as models trained once can be downloaded onto billions of consumer devices worldwide over the Internet. The models can uplink data to a central location for further improves across the mass of deployed models. Even if end applications have narrow scopes, such as recognizing images, the prospect of reaching billions of devices indicates a large economic impact. The results ultimately depend on how developers find use of this property to make valuable applications to end users.

Beyond analytics and features on consumer devices, the largest economic impact will come from intelligent systems able to perform tasks that humans could not otherwise implement in a machine. Some tasks we perform we can do with ease, but we have difficulty explaining how we do it. Examples include moving our limbs, recognizing objects, or speaking. This has prevented humans from creating machines able to perform many of tasks that we find necessary to operate in our world. AI will revolutionize many technologies in the real world through enabling sensor interpretation and complex behavioral planning. This will lead to vertical markets that integrate collections of sensory information and actuation into machines that work robustly in real working environments. These services could look like Waymo's autonomous vehicle taxi service, integrating transportation, robotics, AI and a service into a consumer app (Korosec, 2019).

The autonomous vehicle market acts as one of the largest short-term growth opportunities for AI, freeing time for passengers in traffic during commutes and enabling

autonomous transportation of goods. The economic impact will not only affect passengers: internet service providers will evolve to keep pace with higher demand for wireless internet access; autonomous delivery vehicles will reduce the cost and time of shipping for goods to homes and to stores; efficient algorithms may assist cars to reduce overall traffic delays; vehicle fleets may use electric storage and renewable energy to provide rideshare services. There is a strong likelihood that the autonomous vehicle market itself will transform technology in a way that impacts many aspects of human life. Similar markets to autonomous vehicles include markets that can integrate robotics to reduce costs for operations. A wide deployment of self-learning robots similar to autonomous vehicles could change industries and create new markets opportunities. AI addresses many of the challenges that robots face before reaching the robustness to perform valuable work, mainly through elements that humans cannot easily program into machines.

AI will not only change the infrastructure and improve technology around us but can also assist with scientific discovery. Scientific researchers can find value either through using robots for exploration purposes or through testing the insights of data automation. The results could include medical breakthroughs, robotic space exploration, high energy physics and new discoveries in materials science. Early results have already been found using deep learning algorithms. These include drug discovery (Seq2seq Fingerprint: An Unsupervised Deep Molecular Embedding for Drug Discovery), finding treatments for fighting cancer (The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology) or discoveries in genetics (Alipanahi, Delong, Frey & Weirauch, 2015). Other researchers have used AI to analyze data in particle physics (Baldi, Sadowski, & Whiteson 2014) Each of these research areas could correspond to an economic impact of

billions of dollars; for instance, the American Cancer society reported that U.S. spending on cancer-related healthcare was $87.8 billion. (The Costs of Cancer, 2017). Furthermore, enhancing scientific research could lead to profound impacts that improve the wellbeing of humanity and fueling our curiosity for discovery.

AI will have a profound economic impact on society as it reduces costs for firms, reaches billions of devices, creates new vertical markets and augments scientific discovery. Nevertheless, applications will require human effort to build and deploy after accounting for limitations in current technological capabilities. Deep learning performs specific tasks well but has multiple implementation barriers. Training base models requires a large computational infrastructure and large datasets on top of expertise for developing optimal model architectures. Nevertheless, the continuing growth cloud computing infrastructure and large volumes of data on the web can mitigate this barrier. Aside from computability, one of the most pressing issues facing deep learning stems from the inability to interpret when the network will make a wrong decision. This hampers the deployment of deep learning in safety critical and high stakes environments. Research developments in the domain of understandable AI could break the barrier for market adoption of many products. Consequentially, the deployment of AI may take many years to reach maturity, requiring research in many subdomains and the human development of applications to fit consumer needs. Nevertheless, as these barriers are overcome, there may be wide adoption of applications on devices worldwide.

Large firms may see the greatest initial value from AI technologies (Manyika et al, 2018), when considering these firms can benefit from improved efficiencies and have lower barrier to entry with data and computational resources. However, as AI technology becomes

more available to smaller firms or innovators, the consumer space of applications may improve rapidly and lead to novel applications that capture great value in the long run (Gilham, 2017). These factors involve the portability of AI with deep learning, which can be trained remotely from consumer devices and deployed across a wide number of devices, as well as the ability for small groups to retrain and repurpose AI models for specific applications that could lead to innovative technologies and scientific discoveries.

3.2 Example Market Impacts

Bughin et al. state that upwards of 70 percent of firms may adopt AI technologies by 2030 (Bughin et al, 2018). The 2013 McKinsey Global Institute Disruptive Technologies Report claims that the combination of advanced robotics and machine learning will augment knowledge work and substitute physical tasks (Manyika, et al., 2013). This report predicts the automation of millions of knowledge and manual labor tasks: up to 230 million knowledge workers and 320 million manufacturing workers may be impacted AI by 2025 (Manyika, et al., 2013). When considering the role of deep learning in autonomous vehicles, AI may work its way into over one billion vehicles as the technology overcomes barriers. Considering the factors of robotics, automation of knowledge work and autonomous vehicles, the total impact of AI on these industries may result in a global impact of over $20 Trillion by 2025 (Manyika, et al., 2013). MGI describes the economic impact as the value added to people's lives with these technologies such as cost savings for otherwise previously expensive goods or services. AI ranks as one of the largest disruptive technologies by 2025 due to both the large economic impact and the ways in which it can transform occupations and technology.

Automation of knowledge work goes in line with data analytics which will offer cost savings and profit increases for large companies. The deployable technologies including large scale robotic automation and autonomous vehicles may take more time for developers to reach market fit due to design challenges and constraints. Nevertheless, AI can impact many diverse areas ranging anywhere from healthcare to marketing. The end technologies will depend on the applications work on. The following examples explain the potential applications that could be developed in various markets according to the MGI 2013 report.

3.2.1 Automation in Call Centers

Recurrent Neural Networks have shown significant improvements over previous algorithms involving time series data processing. This type of learning improves the ability for software to understand what sentences are spoken by customers speaking to call centers. Deep learning could be used to generate synthesized speech that sounds more realistic to humans. This could lead to full creation of sentences to answer questions as opposed to premade responses. Additionally, meaning extraction could be used to categorize questions and redirect the caller to a qualified representative. These applications would provide cost savings to companies and make call center experiences more enjoyable for customers. One source claims chatbots that communicate with customers could reduce annual costs of firms by $8 billion (Gilchrist, 2017).

3.2.2 Recommendation Systems

This broad category includes specific applications such as recommendation services for customers of online platforms. Recommendation services make user experience more enjoyable by filtering unwanted items from customers. Machine learning enables these services from collecting data to make predictions on user preferences. Companies that adopt these algorithms may find their customers invest more time into provided services and match customers with relevant advertising to increase profits. Data collection and cookies will improve services for web browsers seeking information, increasing customer satisfaction. Recommendation systems provide a large value to digital companies that rely on algorithms to interact with billions of customers, such as Amazon, Netflix, Google and others, where up to 35 to 70 percent of revenue comes from products recommended to customers (Vemuri, 2018).

3.2.3 Predictive Services

Predictive services could be used for a wide variety of tasks, such as predicting the need for maintenance on systems, power outages, success of new products or sizing advertisements on websites. Predictive services may help with risk mitigation by providing insight for future events. This could save costs and improve quality of operations across businesses. With an estimated 70 percent of firms adopting AI technologies by 2030 (Bughin et al, 2018), predictive services could provide value towards the initial investment of AI intregation.

3.2.4 Education

AI could augment education through enriching learning experiences of people of all ages. Data gathered on the success rate of students can provide feedback on how a course should be formatted to optimize the ability of each student to learn. Custom learning experiences that cater to learning styles can improve each student's success. AI could be used to understand which skills a student is stuck on to assist their learning process. It is likely that information technology will augment education in a way that enables teachers to facilitate more personalized connections with students instead of providing a general lecture template for each student to follow. Massively online open courses have enabled people to access education who otherwise may be busy with jobs, lack financial resources, or have geographical barriers preventing them from receiving a formal education. AI enhanced educational platforms could offer more people the ability to learn necessary skills in a changing digital economy. The current E-learning market reached $160 billion in 2016 (Statista, 2019) and one estimate claims the global E-learning market could reach $275

billion by 2022 (Costello, 2017). AI technologies could serve as a core technology to enable dynamic learning experiences in this market.

3.2.5 Healthcare diagnostics

Expert systems have previously been used to aid physicians in diagnosing patients. This trend will continue through advanced data analytics using deep learning approaches. Diagnostic services may function through raw data analytics collected on patients, either to find customized predictive services or to find general trends among a larger population of people. Another application may be through natural language processing. As medical research papers come out, AI may use natural language processing to point doctors to the most relevant information. In larger systems, AI may be used to automate clinical visits in the form of kiosks or at home services that document patient symptoms and notify doctors about abnormal trends. AI could provide personalized healthcare keeps track of various features each day and can notify the patient if they show abnormal behavior. At home services could be particularly helpful for patients in age ranges that show higher risks of illnesses. Ultimately, this service could catch issues before they grow into worse conditions and could provide minimally invasive preemptive care. In the case of stroke, which accounts for the leading cause of death in China, the global cost of medical expenses reached $689 billion (Fei et al., 2017). Reducing the cost and improving quality of care with AI in this domain could add value on the order of billions of dollars per year.

3.2.6 Speeding up scientific discoveries

Scientific research has focuses on the collection and analysis of empirical data. Computational based automated workflows can provide ways to accelerate the pace of

analyzing data and discovering trends (Gil et al., 2007), indicating that AI technologies that improve workflows could accelerate scientific discoveries. Using generative adversarial networks and reinforcement learning, researchers may find ways to generate new cures to diseases, discover new molecular structures, develop nanotechnology, sequence genomes, analyze space imagery and discover particles from high energy collisions. DeepMind currently researches deep learning and reinforcement learning for understanding protein folding, which could help us understand the fundamentals of biology and treating various diseases (DeepMind, 2019). The American Cancer society reported that U.S. spending on cancer-related healthcare was $87.8 billion. (The Costs of Cancer, 2017). Given a goal and a way to test out the search space, deep reinforcement learning could be used to find specific ways to develop structures that may be valuable to humans. It is likely that AI will advance the pace of scientific discovery and provide unpredictable benefits to society from these endeavors.

3.2.7 Debugging and optimization of software

The practice of AI in software development has existed for decades (Rich & Waters, 1986). AI could aid software development though debugging and optimization. While programmers are designing software, AI can determine if any issues may arise in the code and notify the programmers to save time in finding bugs. AI could also use algorithms such as reinforcement learning to optimize algorithms, such as optimizing other deep learning models to provide the highest accuracy or fastest convergence speeds. Eventually AI may be used to recursively optimize its own algorithms such that it can find more powerful optimization techniques. AI will likely play a large role in technology development as it will

help streamline the design and development of new tools and may even be used to discover new methods to carry out a task.

3.2.8 Project management

In 2017, project managers claimed that managing project costs, organizing work and sharing information across teams were the largest challenges (Sauer, 2019). In complex environments with many variables, AI will provide strategies to manage resources for projects. Optimization algorithms generally take exponential time to solve which means traditional computer algorithms cannot find perfect solutions in reasonable time. Wu et al. describe various state of the approaches to improving risk management through intelligent systems (Wu et al., 2017). AI could determine a reasonable solution that may be suboptimal but still better than what a human could determine. Additionally, AI could provide real time project management strategies such that when a project becomes delayed, the system could make quick decisions on the next best plan to finish the task under its constraints. It is likely that AI will play a role in project management as platforms come out to augment the ability of managers to make quick decisions for businesses. Deeper insight into project progress could be found in collecting and analyzing data gathered from workers or inventory. This analysis could determine where roadblocks may be occurring or even predict if something may come up in coming weeks. This insight could improve foresight of managers to make better decisions for business efficiency.

3.2.9 Fault analysis

In complex systems with possible points of failures, AI could analyze patterns to predict where faults may occur. This could take the form of looking at individually designed

parts to entire power grids (Bhattacharya & Sinha, 2017). The ability for AI to take in data and predict the locations of faults will save money and time to those developing and maintaining these systems. This will prevent user inconvenience when systems fail. Fault analysis could also be applied to dynamic systems, such as to determine congestion in internet traffic or traffic on roads. The annual cost of traffic could rise to $293 billion by 2030 (IRNIX, 2014), and AI solutions to mitigate these expenses could contribute to billions in economic impact.

3.2.10 Forecasting trends in data, including financial analysis

Today most financial trading systems use algorithms to quickly buy and sell stocks. In 2005, algorithmic trading accounted for 25 percent of trading volume but increased beyond 75 percent in 2009 (Glantz & Kissel, 2013). The ability for algorithms to make accurate decisions on the value of trade increase the return on capital in these financial markets. AI will continue this trend in providing deeper insights and predictions on the values of stocks and other financial decisions. This will help large trading firms achieve higher growth. Financial analysis could also be used for other purposes such as determining the price of property, valuation on a business or an individual's estimated salary based on education, work experience and skills. Data accessibility may be blocked behind proprietary barriers but the algorithms for these applications will likely augment these areas for those with data.

3.2.11 Legal document searching

Image recognition and natural language processing can be used to search through documents, either electronically written or scanned into a computer. This would improve the

ability for lawyers to extract relevant documents to help in a case. Additionally, document searching could be used by many firms to find a datasheet or important information from document archives. This will save workers time from searching and retrieving information because AI will be able to automate this task. In 2014 the global market for eDiscovery was $1.6 Billion and was forecasting to reach $3.8 Billion by 2018 (MarketWire, 2014). Numerous companies have grown to fill this demand will likely continue to gain use across other domains.

3.2.12 Robot-human augmentation

In robotic systems that have many degrees of freedom, there are numerous paths to reach the same position, which can be challenging for a computer to determine the best path to take. Deep reinforcement learning algorithms could be used to help program dexterous limbs used for prosthesis such that these systems are able to reach the goal of grasping objects in reasonable time. Another way AI could be used to improve prosthesis could be through scanning a person's neurological activity to determine if they intend to do a specific action. Eventually the human robot interaction would converge to a point where a person could simply think about the action to take, such as grasping an object, and the robot would carry out the task. This augmentation could be used for people with disabilities or people who are performing strenuous tasks in an occupational setting. The direction of how these technologies evolve an play into our lives could take multiple paths. The estimated market value of exoskeletons is $2.5 billion by 2024 (Bay, 2018) while the estimated market value of augmented and virtual reality could reach $209 billion by 2022 (Gee, 2018), where AI could act as a core technology for applications and integration.

3.2.13 Industrial robots

AI will have a large impact on improving the efficiency of industrial robotics. AI algorithms will determine the best ways to carry out a task and can learn to perform tasks given a goal. With AI enabling the actions of industrial robots, automation of many tasks will drive down the prices of most manufactured goods. In 2017 the global revenue for indsutrail robots was $39.3 billion but is expected to reach $498 billion by 2025 (Robotics market revenue worldwide 2017/2025 | Statistic, 2019).

3.2.14 Surgical robots

With surgical robots, minimally invasive procedures provide the best rates of recovery of patients. The actuation of surgical robots will be augmented by AI that plans the best ways reduce invasiveness invasive during surgery. These systems will learn how to apply the minimum pressure necessary grasp and perform tasks in this domain using high accuracy sensors and feedback for the system controls. These systems could also learn to avoid damaging a patient is the surgeon makes a sharp move on accident. The market value of surgical robots was $3.9 billion in 2018 and is expected to reach $6.5 billion by 2023 (Singh, 2019). It is likely that AI will augment the work of surgeons, allowing them to heal patients more effectively.

3.2.15 Personal and home robots

The improvements of natural language processing are enabling home robots to understand human intentions and carry out tasks. As home robots become more complex and may move around, perception from deep learning techniques will improve personal home robots to avoid interfering with objects in a home. These robots may also learn trends about

daily patterns of people in homes and provide routine services such as cooking meals or preparing coffee in the morning automatically. Personal robots may also act as security systems for homes if they watch over the area to make sure people do not attempt to break into a house or steal mail. It is likely that AI will continue to make its way into consumer households in the form personal assistants as developments continue to improve. One estimate puts the market at domestic robots at $4.4 Billion by 2025 (Murphy, 2017), while home assistants may drive greater value to firms, as Amazon's Alexa is forecasted to add $11 billion to revenue by 2020 (Kim, 2016).

3.2.16 Commercial service robots

Robots will use AI to automate tasks in service sectors. These tasks range from preparing food, shipping packages and taking orders. All these tasks carried out by robots will require advanced perception, action planning and natural language processing to understand customer desires. As these field of AI improve, robots will automate many of these tasks and reduce the costs to perform these tasks. A current prediction expects the market size to grow to $34.7 Billion by 2022 (Gunjan, 2015). Services may become more profitable as these services may be performed faster and at lower cost.

3.2.17 Autonomous cars and trucks

Autonomous vehicles will rely on AI to carry out perception, decision making, navigation, and actuation of the car. As autonomous vehicle become more common, a more abstract decision-making algorithm my evolve to determine the best actions of all the cars on the roads to prevent traffic congestion. Autonomous vehicles could reduce environmental impact as they could run on electric charge and autonomously navigate to charging stations

that run on renewable energy. These vehicles may carry passengers or provide delivery services to retails stores or home delivery. The cost effectiveness of these vehicles will reduce prices on transportation and delivery services while also reducing environmental impact. An estimate predicts that the global market for mobile robots may reach $54.1 Billion by 2023 (Singh, 2018). Algorithms involving perception of pedestrians and avoid accidents will need to improve for mass adoption.

3.3 AI Impacts on Labor

AI is projected to automate many tasks, substituting some jobs and complimenting others. Frey and Osborne predict that 47 percent of occupations in Europe and North America may be substituted (Frey & Osborne, 2013). The OECD puts this number lower to 9 percent of jobs expecting full substitution, attributing the affect of automation to substituted tasks in other occupations rather than full job substitution (Gillham, 2017). This lower bound does not necessitate that the income distributions with highly substituted tasks will not become volatile or experience lower real earnings, which could be the case where automation substitutes tasks in occupations that increases supply and lowers the demand of skills in that occupation. A geographical assessment puts North America and Europe at a higher substitution rate between 26 percent and 76 percent, while Asian and other developed countries may experience a lower rate of 11 and 29 percent (Gillham, 2017). Bughin et al. acknowledge that while growth may increase rapidly for adopters and workers with high skills, those performing repetitive tasks could experience lower incomes and lags in growth (Bughin et al., 2018). While the previous section found a large growth in GWP, the income distribution between occupations will unlikely rise equally. Those in the upper decile of skills and income could see growth, while middle income could decrease. Furthermore, low skill jobs involving narrow tasks or repetition could see widescale substitution of tasks resulting in job volatility and reduction in real earnings. The impact of AI on labor does not necessitate job loss; rather factors between job responsibilities and skills could impact distributions for many occupations. This section looks at the factors that Occupational changes brought from AI depend on individual factors, the job responsibilities pertaining occupations and the market demand for products enabled with AI.

On one dimension, the response of individuals learning high value skills will lead to higher earnings and job prospects; on the other hand, differences in motivation may cause some people to respond slower to change and face poorer outcomes. This dimension has many complex factors involving individual backgrounds and values, constituting rigidity to which we cannot expect easy change. For instance, it may be unreasonable to expect many to go back to school to relearn skills not relevant to previous occupational work. Rigidity in values may factor into an increased demand for education of non-relevant skills. Some may go to institutions to learn non-relevant skills later to find poor job outlooks and low earnings due to remaining skill gaps, not spending time learning relevant skills because they were not interested in learning those skills that correspond to higher earnings. On the second dimension, the automation of tasks will cause some occupations to become more productive and valuable while others counterproductive and less valuable. For instance, one who uses new AI tools to automate time consuming tasks will find higher daily productivity for the set of assigned responsibilities. In contrast, in a situation where one's primary responsibility involves performing a task identically to a machine, hiring a person instead of a computer for this task would be counterproductive and have a higher cost to the firm. Because of different responsibilities between occupations, AI will compliment labor for many jobs while substituting labor for many others. The third dimension involves the elasticity of demand for markets in which AI automates jobs. In industries where elasticity of demand is high, AI will enable more job openings and bring large benefits to society. In jobs with low elasticity of demand, AI will likely automate many jobs while driving prices to low levels that provide an overall decrease in labor and lower marginal benefit to society. The large-scale outcome for labor will depend on individual values, the set of responsibilities for specific occupations as

well as the elasticity of demand for products in the markets adopting automation. During the digital revolution, someone who values learning new skills and finds enjoyment in complex, high responsibility work settings will see better outcomes than one who does not value learning important skills and works within in a narrow set of responsibilities.

Previous sections have observed the need for large amounts of data for the high performance AI algorithms. Occupations that share many similar tasks among a large set of employees may allow for large enough data collection for automation. For example, call centers require employees to understand common questions and provide answers; yet, machine learning allows for call centers to automate this process and reduce the number of employees needed in call centers. This observation implies that jobs that occupy a larger set of workers have a higher change for automation, overall impacting more workers. Frey and Osborne predict that 47% of occupations risk future substitution from computerization in coming decades (Frey & Osborne 2013). Jobs in the service, transportation, manufacturing and construction industries will experience automation due to the ability to use robots and data collection. These industries contain the large percentage of jobs at risk for automation since so many workers fill roles in these industries. Frey and Osborne make the argument that future developments in AI will break the trend of non-routine tasks complementation and routine task substitution. Previously computers could not automate non-routine jobs; however, this notion has changed with deep learning able to learn complex tasks at a superhuman level given enough training data. Computers alone can perform routine tasks superhuman level such as crunching numbers at superhuman speeds. Now the realm of computable tasks has migrated to non-routine tasks like driving cars. They argue that computers will automate any task given enough data for training.

Other researchers such as Autor and Acemoglu have observed a historical growth in occupations that involve non-routine work (Autor & Dorn, 2013). Nevertheless, this growth was in terms of low skill service jobs and high skills jobs, where many of the routine middle-income jobs were automated with computers. In the context of Frey and Osborne's prediction, jobs with non-routine tasks, but are common among millions of people will become automated. Non-routine tasks in terms of uniqueness to specific individuals will be the tasks not automated with computers soon. This characteristic shields jobs such as management in individual firms for automation. Nevertheless, the implications that occupations with a higher volume of workers have a higher risk for automation means that many people may lose jobs and need to relearn skills. Workers that need to relocate may face lower lifetime earnings or transition into a lower income occupation without learning new skills.

As mentioned at the start of this section, one of the most individualized factors determining jobs outlooks depends on one's skills, traits and values. Current research investigates the effects of personal traits on individual incomes and job stability. Brynjolfsson et al. have found traits that correlate to earnings in various occupations involving information technology (Brynjolfsson, MacCrory, & Westerman, 2015). Their findings show that interpersonal and physical skills correlate with lower earnings, while equipment and initiative skills correlate with higher earnings in technical fields. In occupations with less computer use, these skills have the opposite effect. Workers in lower skill service jobs experience better earnings with physical labor skills and interpersonal skills. The skills correlated with the highest earnings across all occupations were adaptability and initiative. If anything, this data provides further evidence for the digital revolution

argued by Brynjolfsson; adaptability and initiative would allow people to take advantage of a changing economy when new lucrative opportunities open during change.

Furthermore, the jobs that have not been automated by machines yet, such as service jobs, may be experiencing growth due to substitution of once middle-class jobs, accounting for the value of interpersonal skills. Those with initiative or adaptability to learn skills migrate to high skilled labor while others face migration to lower income occupations. Bessen has pointed out that middle skill workers commonly invest in learning new skills to attain higher earnings, reflecting the traits of adaptability and initiative that account for growth in high income jobs (Bessen, 2016).

The main observation to extract from Brynjolfsson et al. is that of the increased earnings from adaptability and initiative across all occupations. Those that take the initiative to learn new skills will find skills complementing work and see increased earnings, more so today than before the digital revolution. Brynjolfsson's research showed that the current trend correlates physical labor and dexterity to increased earnings. The value of other skills, such as physical skills, only indicates that jobs not computable yet garner higher earnings. The observation that these skills lead to lower earnings in computerized occupations indicates that trend will last over a short term as an indicator of higher earnings in some jobs. As noted in the economic value section and by Frey and Osborne, AI will reach deployment across billions of devices worldwide and substitute tasks whenever data is available. Once computers reach more markets involving physical labor or other skills, humans will no longer add value with this skill except for in niche applications. Predictions about the skills not computable in coming years should be met with caution. Few could look at computer technology decades ago and predict cars driving themselves today. The same goes for vision

and speech recognition. Future improvements in AI, such as reducing the time for convergence with deep reinforcement learning, could enable billions of robots to automate physical tasks. In the time of the digital revolution, those that adapt to changes to learn new skills will see complementation in occupational earnings while those that fail to adapt will face migration to lower skill and lower income jobs. Physical skills could be valuable now but could lose value is future technologies mature.

Adaptability and initiative may lead to higher earnings, but this observation does not provide an innovative insight to the job market. In a free economy, initiative to learn valuable skills often leads to success. It is not reasonable to expect all people to possess the same level of adaptability and initiative due to various background factors and situations. Furthermore, some in specific occupational settings may possess an easier means of learning new skills than others. Bessen explains that high skilled workers more often invest in learning computer skills that complement current occupational work; whereas, those in occupations without computers see less incentive and return in learning computer skills. As a result, high skill employees may choose to invest in new skills and earn higher wages, while low skill employees do not experience wage increases (Bessen, 2016). Bessen explains these choices between different workers leads to wage disparities and polarized incomes. In addition to different incentives, those with high earnings may have a stronger educational background, disposition or financial means of furthering education. For this reason, not everyone faces the same benefits when new technology opportunities occur.

Nevertheless, while polarized incomes may occur from initiative and adaptability differences, industries may change to where computers become commonplace with lower skilled jobs and require less experience to learn or operate. Even those with less educational

background or financial means can still find resources available among the vast wealth of information on the world wide web. Those that have the initiative to pursue technology enabled businesses with low startup costs can find more earnings than at a minimum income job as well as employ other workers. Thus, despite the observation that those at higher skilled jobs experience a path of less resistance, overall the wave of new technology from the digital revolution can equip people from all backgrounds with the ability to see higher earnings given they have initiative to learn new skills and adapt to change.

Aside from the personal traits that motivate occupational changes, the nature of specific occupations factors into the outcome of substitution or complementation. AI serves as a tool for humans to automate tasks and will not automate all occupations (Bessen, 2016). An occupation may contain a diverse set of tasks, where technology may only automate a faction of the tasks involved. In the case where technology substitutes small a subset of the tasks involved in an occupation, the overall efficiency of the occupation improves, and technology complements the job (Autor, Levy & Murnane, 2003). AI will not automate a job unless it automates every task within the job. In many cases, AI will increase the value of occupations through automating menial tasks that detract from the core value. For example, medical doctors may experience occupational improvements as new technology assists communication, diagnosis, prescription and reviewing research. AI could automate tasks such scanning medical images and searching through lists of prescriptions, giving doctors more time to interact with patients or see more people. In general, technology creates new jobs through reducing the effort of performing mundane tasks. Before the industrial revolution, none could expect to work as an airline pilot, software engineer or automotive

manufacturer. Throughout history, technology has automated tasks and engendered the rise of higher income occupations.

Acemoglu and Autor find that the demand for both workers in low and high skill tasks has increased because computers could not substitute for non-explicitly defined tasks (Acemoglu & Autor, 2011). In contrast, computers substituted workers in occupations where the tasks could be accomplished by computers, primarily in middle income occupations. Autor and Dorn have observed the substitution of second quartile has created a polarized workforce with a U-shaped curve (Autor & Dorn, 2013). Second and third quartile jobs have decreased in employment share narrowing the middle class. Labor has shifted to low skill and low-income jobs and high skill high income jobs. Employment on both extremes of low and high skill jobs has increased in response. The increase of high skill jobs can be attributed to technology complementing these workers. Autor finds that computers substituted jobs in non-routine tasks and complemented jobs in non-routine tasks (Autor & Dorn, 2013) Both first and fourth quartile jobs require non-routine jobs that computers could not perform well. This observation follows Bessen's explanation of middle skill workers investing in new skills to reach higher earnings. People that once held jobs in middle skill occupations either learned new skills to transition into high income jobs or did not learn skills and took on lower skill jobs with lower earnings. This observation means that any occupation that had computable tasks pushed workers into adapting to these changes. What was previously non-computable became area of disruption once computers had the technical capability to automating the core tasks of the occupation. Other research has shown technological substitution of core value reducing earnings even in higher skill third quartile jobs (Deming, 2017)

Acemoglu finds that with a surplus of middle skill employees and lack of middle skill jobs, college graduates with mid-tier skills often migrate to low skill occupations (Acemoglu & Autor, 2011). While the number of low skill jobs have grown, the supply of workers seeking low skill jobs has reduced the real earnings of low skill workers who would normally take these jobs. When the average skills of workers in a labor market increase, those that do not pursue the same level of education as competing employees experience reduced earnings. This trend can be seen in a reduction in earnings for employees who received a high school education as the highest level of education (Acemoglu & Autor, 2011). MacCrory finds that over the last thirty years, earnings of college graduates have increased while those with less education have experienced lower earnings (Brynjolfsson, MacCrory, & Westerman, 2015)

From the observations of technological substitution, disruption will occur for jobs with narrow sets of tasks as AI enables automation of less explicit tasks. Even if the task itself is complex, if the occupational responsibilities comprise a small set of tasks then the core nature of the occupation is at risk for automation. Automation may occur if enough data can be collected to automate the core task involved in the job. Diversification of job responsibilities mitigates the risk of automation undercutting the core value of the job. Jobs such as truck or taxi drivers may experience substitution from autonomous vehicles since they have a narrow set of tasks. In contrast, workers with less clear roles designing or managing fleets of autonomous vehicles will experience earnings increases, as automation of a subset of tasks will increase working efficiency. In general, AI will have a higher chance of disrupting occupations with narrowly defined responsibilities even if single tasks are complex. As seen with AlphaGo, even the most intuitive and complex decision-making tasks can be automated, but only in a narrow scope.

In the short term, a greater division of income will grow between those of different skills. These disparities may be a natural consequence of technological revolution. Imagining the lives of poor children working in factories after coming from farms during the industrial revolution gives perspective on how a cultural lag can lead to economic disparities. Some may often not have the option to remain in a longstanding occupational setting once technology substitutes core tasks. Those coming from the farm during the industrial revolution may not have been able to keep up with machines that drove up competition for farm work and reduced the cost of food. Technological revolutions may have a pattern of displacing once middle-income workers while enabling new industries. As a result, large groups of people must migrate to new work, either accepting a lower relative income or adapting to high skill occupations. Even though the overall quality of life in society improves from technological innovation, those that do not adapt do not see as great immediate returns. As the digital revolution grows, not all will find ease transitioning into higher skill occupations without proper education background. Those researching the digital revolution have suggested a need for improving education to keep up with change (Brynjolfsson & McAfee, 2014). Even those with today's public education in the United States may still complete secondary education and not overcome the skill gap necessary to enter relevant work, facing low income jobs after graduation from college (Acemoglu & Autor, 2011). It may require a new generation of learning skills from an early age before income disparities and skill gaps level out and for the middle class to grow.

Although task automation changes the nature of occupations, the response of markets to costs determines the outcome of the occupation itself. In some situations, the automation of jobs replaces workers in this occupation, resulting in an obsolescence of that occupation.

For instance, elevator operators have been automated with robust microcomputers. Nevertheless, some occupations grow even when the core task becomes automated. Bessen points out the example of the growth of bank tellers after the installment of ATMs. Although ATMs automated the core task of handling cash, the rapid growth of local bank branches opened more jobs, and the people that filled the new jobs were the past bank tellers (Bessen, 2016). Those with previous experience in the banking industry took on new responsibilities for the tasks not feasible with computers. Adaptability and initiative barriers were lower than would be needed for entire occupational migration; consequentially, the people in these jobs had the ability to gain higher earnings from adapting to new responsibilities despite job disruption. Bank tellers could not cling to old tasks or else they would have experienced positive outcomes from substitution and change. Their prior knowledge served as a catalyst to reduce the energy needed to acquire relevant skills. We may see a similar pattern happen for taxi, but or truck drivers, where these workers adapt to new skills while working in the same industry. Bessen argues that the elasticity of demand of a product to determines the impact of technology on a labor market for each industry (Bessen, 2016). If the elasticity of demand is high for a product, automating jobs in the given industry creates surplus demand that outpaces job loss.

If the cost of a product goes down in half, but four times as many people buy the product, the total revenue of the industry doubles. Jobs open since two people now oversee four robots instead of one person working by hand. For elastic demand of products or services, automation may increase jobs, albeit requiring workers to take on new roles. In contrast, automation will cause job loss if elasticity of demand for a product low. Cost reductions for inelastic products does not increase gross product, thus one robot will take the

job of multiple people. Bessen has documented this trend for textiles, steel and automotive vehicles (Bessen, 2018). When automation reduced the cost of automotive manufacturing, many more people purchased cars who previously could not afford this product, indicating high elasticity of demand. Jobs grew rapidly in the automotive industry as automotive demand increased. Years later when most owned one or more cars, the elasticity of demand dropped, and further automation resulted in massive job loss. Few people needed more cars in the second wave of automation.

Looking through the lenses of elasticity of demand, those in occupations where AI automates core tasks, those that take on broader responsibilities may see overall job and earnings growth. These responsibilities will be those where little data exists to create robust systems or where human values and judgement holds precedence. Those that have long term experience in an industry that experiences automation will still find good outlooks if the industry experiences a high elasticity of demand shift for consumers. The products and services that few people can afford now may become the fast-growing industries in the future, employing many workers. Nevertheless, those working in low elasticity of demand industries will likely find job substitution or layoffs, as fewer new jobs will open. Instead, those with higher skills and initiative may take over these industries using new means of meeting consumer demand, just as how a far smaller percentage of people in the United States grow food crops than before the industrial revolution.

Through observing the research of technological change and jobs, the factors that will determine job outcomes during AI adoption will involve personal traits, the nature of the occupations set of tasks as well as the elasticity of demand for the product or service. Those with adaptability and initiative will find opportunities to learn high value skills that lead to

large earnings in the digital revolution. In addition, those in industries where products have a high elasticity of demand for consumers will find the opportunities to take on new roles even if the core task is replaced with AI. People working in occupations with a narrow set of tasks, or those that share tasks across millions of other workers in industries with low consumer elasticity of demand will lose jobs. In these settings AI will automate work as large-scale data collection will enable efficient and robust algorithms.

3.4 Summary

The digital revolution will serve as a main driving force for the economic impacts of AI. The applications enabled with AI today would not have been possible decades ago before large data storage technology, the world wide web and computational speeds. AI will enable new applications in technology that will accelerate the digital revolution. Industries will find use with data analytics, starting from large firms and later available for smaller firms through digital platforms and pre-trained deep learning models. Additionally, AI will find deployment across billions of devices worldwide, opening opportunities for novel applications and technologies. These technologies span from intelligent assistants to autonomous vehicles. Robotics and automation will improve with AI changing various industries. Assessing the economic impact of these factors alone, we can expect an annual economic impact reaching $10 to $20 Trillion USD AI based technology, (Bughin et al, 2018), (Gillham, 2017), (Manyika, et al., 2013). Beyond new technologies, AI will provide tools for scientists and researchers to accelerate the pace of scientific discovery. With AI, humans may find cures for diseases and breakthroughs in science that change the future of society.

With such a powerful technology, valid concerns exist for the outcome of jobs. The number of jobs lost in North America and Europe could range from 9 percent (Gillham, 2017) to 47 percent (Frey & Osborne, 2014); however, the impacts will not occur uniformly between occupations. Jobs where substitution of tasks rather than entire job responsibilities determines the interpretation of the projected impact and even if jobs are not lost, real earnings could decrease in many jobs relying on low skills or repetitive labor. Three factors will influence the outcome of individuals in future occupations: the initiative of the

individual, the size of the set of tasks within an occupation, and the elasticity of demand by consumers for the products in the occupation's industry. Job substitution and migration will occur for those in industries with low consumer elasticity of demand. Those that adapt to change in high consumer elasticity of demand industries may instead see improved job outlooks even if automation takes over the core task. Others who fit into high skill work will find that AI improves work efficiency and can lead to higher earnings. Some in middle skill jobs may take on the initiative to learn skills that push them into higher earning occupations. All these factors result in a polarized distribution of income with a narrow middle class. This is a byproduct of a technological revolution and cultural lag. Those in the high-income brackets will be those that adapt to rapid economic change, whereas those that do not adapt migrate to the growing low skill bracket. The digital revolution will disrupt many industries, and workers that do not adjust to change will not experience the same large income increases that others see. Even if real earnings stagnate for middle income workers, both the low-income earnings and the high-income earnings rapidly change to the new age of technology where all occupations earn more. AI will serve as a core technology in the digital revolution, increasing the relative per capita GDP in society, but may require up to half of people in developed countries to adapt to occupational disruptions.

4 Ethical Implications of AI

The theme of dangerous AI appears frequently in Hollywood movies. Examples include the Terminator where machines control the world government and come back in time to assassinate humans, Ex Machina where the robot surpasses the intelligence of her creator and kills him, and the 2001 Space Odyssey where the AI Hal kills the crew members because they come in the way of the mission. While these stories are fictional, they all share three central risks of AI: performing implicit tasks incorrectly, programmatically killing without remorse, and jeopardizing human freedom. These dangers exist even in current technology and could manifest into larger challenges as AI advances.

The debate of the ethical of implications remains a controversial topic among researchers and philosophers. The longstanding AI expert Stuart Russell argues that weaponized AI threatens to human freedom with current technology (Russell, 2015). Tech icon Elon Musk has argued that future advancements in AI pose an existential threat to humanity (Clifford, 2018). Oxford philosopher Nick Bostrom has written the book Superintelligence detailing the wide range scenarios in which human level AI could threaten society. Bill Joy, co-founder of Sun Microsystems and pioneering contributor to BSD UNIX, calls attention to our lack of fear for 21st century replication technologies (nanotechnology, bioengineering and robotics) in contrast to the fear we held for 20th century technologies such as nuclear weapons (Joy, 2000). Despite the credentials of these experts, some may still question the precedence of these concerns considering that these concerns have existed for decades even before the AI winter. Few agree when and if general human level intelligence will come. Furthermore, it is possible that misinformation and perceptions gathered from movies has biased some into fearing AI, making risk concerns a topic of science fiction in the

academic domain. Yet considering that the digital revolution will transform human technology and society, AI will have profound ethical implications just as any powerful world changing technology.

The policies established during the digital revolution will shape the growth of technology and impact the lives of citizens. Mistakes on behalf of governments could lead to government instabilities and economic challenges. Regulating use of AI may suppress economic growth and allow other nations to rise to power, threatening national security. Others argue that unrestricted use of AI may enable global terrorism, cold wars or unstable governments that inflict harm on millions of people. Poor economic policies may harm substituted workers and lead to economic disparities. Brynjolfsson and McAfee argue that institutions will strongly influence the wellbeing of those affected by the digital revolution, placing importance on governments to understand the ethical implications of technological change (Brynjolfsson & McAfee, 2014). Multiple organizations have formed to seek policy guidelines that account for the ethical implications of AI, including the Future of Life Institute, the AI Initiative at the Harvard Kennedy School of Governance and OpenAI. Addressing the safety and ethical implications of AI should not hinder the development of AI; rather, move AI in the direction that has the most positive impact on society as whole (Benefits & Risks of AI, 2018).

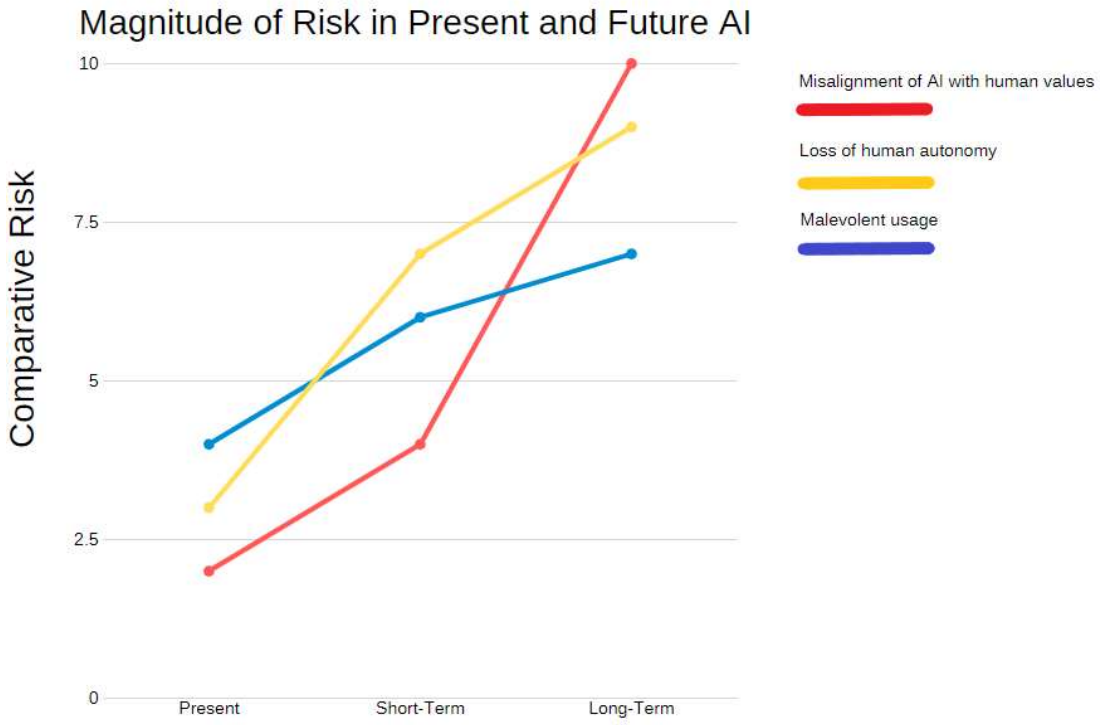
AI should be treated as a powerful tool that has both the power to change and undo society. Both the short and long term uses of AI could result in great human prosperity, but could just as well lead to wars, terrorism, destruction of democracy and even greater existential risks. All powerful technologies have potential for misuse, deserving AI the

utmost considerations for the ethical implications while developing governing policies for the progress of technology.

4.1 Risks

AI functions in the most basic form as a tool for human use; however, that does not stop ill-intentioned use and wide scale implications. The nuclear weapons developed in the 20th century only work at the hands of humans yet remain an existential threat to humanity. Beyond a tool, more advanced AI will obfuscate the human-tool relationship, forcing humans to put faith into systems that may not function as planned. AI systems could perform implicit tasks incorrectly with no guarantee the systems will follow the desires of the operator. Wide scale installment of systems that monitor citizens and influence decisions could undermine democracy in free countries. Three categories of AI risks of involve incorrect performance though misalignment of human values, ill-intentioned use of systems against other people and wide scale risk of human freedom. Figure 4-1 offers a comparative illustration of the risk associated with these categories, indicating that the risk of malevolent usage and suppression of human freedom have precedence at the current level of technology, while superintelligence in future AI could cause greater existential risk to the human species.

Figure 4-1



Some argue that AI misaligning with human values only causes small dangers with current technology but can manifest into an existential threat if technology advances (Bostrom, 2014). In multiple examples, AI becomes an existential threat as it reaches superintelligence wherein the systems follow commands grossly explicit and destroy the world. Systems may be tasked with recursively improving, later to escape the box through the internet and transform Earth and its resources into a single recursively improving computer (Bostrom, 2014). These theoretical examples validate the existential threat of superintelligence; however, this relies on advanced technology that none know how to implement. Nevertheless, current systems could misalign with human values and lead to the end of human society. For instance, automated intercontinental missile systems implemented in a nation without proper validation systems could flag a threat unexpectedly and then send

missiles to a world power. This launch may start a chain reaction of all major countries using automated defense systems, leading to wide scale destruction of human cities. This phenomenon exists in flash crashes that occur in the algorithmic trading markets (Flash crashes, 2019). Sometimes systems will work against each other to crash the stock market momentarily and within seconds billions of dollars disappear. Fortunately, financial transactions can be reversed; however, there is no way to reverse the launch of millions of warheads. Even in financial sphere, some worry that someday a flash crash could cause a global depression that destroys our economy (Flash crashes, 2019). With these systems installed in high risk scenarios we could face global risk.

Current AI systems do not possess moral regard to humans; furthermore, while philosophers argue for the need of moral programming in machines, no feasible solution exists to program AI that could address humanity as a whole (Yampolskiy, 2012). These systems only take actions that aim to maximize an optimization function. As a continuation of John Searle's Chinese Room Argument where current AI cannot think but only emulate thought (Hauser, 2019), so too does current AI not possess a moral consciousness. In science fiction stories, this becomes problematic when AI performs tasks explicitly in a way the creators did not intend to occur. In iRobot, the central computer decides the best way to prolong human wellbeing involves killing the majority of the human race and to use robots to control people's lives. These examples are extreme in the sense they rely on super intelligent systems; however, the theme of this risk derives from machines lacking the same moral values as humans. A moral human would not kill other humans for prosperity; however, this issue exists in today's technology where the systems hold human lives at stake. Both autonomous vehicles and weapons can determine the outcome of a human lives and carry out

tasks not intended by programmers even if the systems believe they made the correct choice, or the most optimal choice encoded in their utility function (section 2.2). Autonomous vehicles could crash into people; autonomous weapons could kill innocent people and wage wars without regard for human life. While the debate continues as to whether autonomous weapons would improve combat situations through preventing casualties or would unleash boundless terrorism and arms races of democratized weapons comparatively in line with biological weapons (Etzioni & Etzioni, 2017), a great risk with human designed AI in high stakes situations comes with design mistakes in both pre-deployment or post-deployment (Yampolskiy, 2015). For example, a weaponized system may not pose existential risk to humans, but human design flaws could cause the system to misbehave and cause widescale damage.

The key innovation of deep learning comes from not needing humans to hand code each decision into the system; however, this serves as double edged sword when needing to validate what the system under all circumstances. Deep learning systems have the pressing issue of uninterpretable decision making. The network learns autonomously how to select floating point weights for a large collection of connections. Humans cannot look at the numbers and understand whether the network will predict an image as a pedestrian or a speed bump. The more advanced the technology, the more difficulty people will face deciphering AI decision making. This issue occurs in technology today and serves as the main risk in futuristic superintelligence. Robust system verification will be necessary to validate automate systems that take control of human lives through robots, vehicles and weapons, else we face risk of systems taking the lives of innocent people.

In some cases, engineers will need to knowingly program systems to perform morally questionable actions where individuals lack the choice on outcomes. This theme is brought up often for autonomous vehicles where systems must face the choice to swerve into a known hazard or hit a pedestrian (Goodall, 2014). This dilemma would either have undefined test results or would require system engineers to programmatically decide which action the vehicle should take. These issues seem like corner cases, but as AI integrates into security systems in the military or law enforcement, these decisions have greater consequences with many lives at stake. The decisions engineers attempt to design into these systems may misalign with desired intentions, where a new update to the world's autonomous vehicle fleet now starts driving people off roads to avoid squirrels. Glitches happen regularly in the world of software, such as during new operating system updates on mobile devices (Goodall, 2014). Small errors can compromise devices across millions of users. Small vulnerabilities in widely deployed AI systems could lead to life threatening situations when people put their lives in the hands of autonomous systems. The frequency of errors could grow as companies rush products to markets, putting ethical dilemmas on companies trying to maximize robustness of systems while staying competitive in the market.

In quantifying the risks of AI, misalignment with human values has greater implications as systems reach wide deployment and possess ability to determine human life. This greatest risk occurs in when AI sits behind weapons of mass destruction. Bostrom has made a clear argument that every instance of human level AI could lead to existential risk in line with human extinction (Bostrom 2014). Still, current technology misaligns with humans infrequently except in edge cases. Often the cause for fault stems from lack of extensive testing before deployment and haste to put a product to market. Car accidents may occur with

current technology, but this does not serve as an existential risk to the same scale as autonomous globalized weapons or superintelligence. Autonomous cars will not suddenly gain minds of their own and decide the fate of the human species.

Despite concerns for a future involving super intelligent AI that escapes the box and threatens humankind, Bostrom explains in Superintelligence that misalignment of human values and errors in programming of current technology would not constitute existential risk (Bostrom, 2014). Instead, this risk manifests into an existential threat down the road when technology improves to the point of human level intelligence. At this later stage, AI technology could behave in uncontrolled ways and compete with human resources or subjugate us through its own volition (Bostrom, 2014). This would be an example of what Bill Joy describes in self-replicating robots that could spell the end for humanity (Joy, 2000). Much of the debate of AI has traditionally centered around this line of concern, which is discussed in section 4.2. Nevertheless, this type of risk constitutes only half of the risk associated with the current technology and near future. This risk depends on an intrinsic loss of human control over systems, whether through programming errors or superintelligence. AI exhibits substantial risks to human well being even while some human still has control over the system. These risks could occur whether through malevolent usage or in well intentioned use that leads suppression of human autonomy.

For current technology, many experts argue that the most pressing risk of AI involves an arms race for lethal autonomous weapons (Benefits & Risks of AI, 2015). The technology that exists today in deep learning can enable low cost autonomous weapons to navigate through obstacles, track targets and execute humans. These could be as lightweight as small quadcopters equipped with lethal discharges and deployed above cities in swarms to kill any

person. Groups that develop this technology could use reinforcement learning to automate drone navigation and convolutional neural networks to recognize obstacles and human targets. Security agencies have the incentive to develop these technologies because it would reduce civilian casualties from bombs, perform superhuman tasks both in navigating through small spaces and taking high risk actions that would otherwise jeopardize the safety of soldiers (Etzioni & Etzioni, 2017). Disposable robots could enter hostile territory where the risk is too great for humans. Nevertheless, the greatest risk associated with developing lethal autonomous systems comes from the wrong people (or wrong government) from getting ahold of the technology. While physical weapons such as intercontinental nuclear missiles are an advanced technology that cannot be assembled in someone's garage, software systems that know how to track and kill humans could be downloaded over the Internet and installed into cheap microcontroller robots by anyone. This argument follows Bill Joy's concern for nanotechnology, biological weapons and robots that are easy to replicate, such as a bioengineered plague weapon (Joy, 2000). Terrorist organizations could threaten the lives of people across the world if they came across the source code and schematics. For this reason, tens of thousands of technology experts including figures Stuart Russell, Stephen Hawking, Elon Musk and Steve Wozniak, have backed an open letter calling for an abandonment on the development of lethal autonomous weapons to avoid an arms race (Open Letter on Autonomous Weapons, 2015). This letter is stated below.

This issue could cause a great threat to societies since resources needed to create lethal systems require low cost materials. Beyond weapons, AI possess the power for people with malicious intentions to carry out crimes against many people. Corrupt organizations could use new technology to subjugate other people in acts of cybercrimes, stripping away

human rights, assassinations and genocide. These issues fall into the realm of human misuse. In the digital revolution, the new wave of technology can change magnify the danger of global conflicts and wars. If these technologies were available a hundred years ago, the impacts of authoritarian dictatorships during previous regimes could have been increase with autonomous weapons and mass surveillance technology available today. As the digital revolution puts software into all products in our lives, including autonomous vehicles, defense systems, financial systems, and healthcare systems, the implications of cybercrimes reach greater levels. When transportation becomes automated, hackers could take control of vehicles in hostage situations to take down public figures. Hackers that find the ability to hack into autonomous planes could inflict damage on infrastructure and thousands of lives. Further, if hackers compromise whole networks of autonomous vehicles, hackers could inflict injuries to millions of people. The systems in which human lives are at play increase the risk of malicious individuals getting ahold power to cause harm to society.

In more indirect ways of harming society, some may use AI systems to publish and spread fake news articles to either sway public opinion or convince people to fall for online scams. These online profiles may seem like legitimate people but could be automated systems instead carrying out social engineering efforts to manipulate actions and opinions. This issue grows when algorithms such as generative adversarial networks generate fake pictures and video overlays that make public figures say false claims. The technology to develop these pieces of media was introduced with GANs in section 2.2 such as convincing faces published by NVIDA (Karras, Aila, Laine, & Lehtinen, 2017). While the current outcomes of social engineering are not fully known, the fact that AI can equip people with these tools could lead to concerns involving online legitimacy in a future that runs on digital

technology. Thus, on top of lethal autonomous weapons, small groups of people could use AI to mobilize dangerous groups or terrorize people online.

Beyond misalignment of goals and small terror groups misusing AI, the larger implications involve the gradual loss of human freedom and autonomy. For instance, even if autonomous weapons were implemented only by responsible governments and not used by terror organizations, could people feel free in a society where lethal drones always fly overhead? If the government suddenly changed positions of power to someone with new ideas about how the country should be run, those that oppose that politician may fear for their personal safety with large scale deployment of these systems. When the roads become full of autonomous vehicles, what will those people do who still wish to drive themselves and not rely on an algorithm? These issues will spread through our society as systems change to depend on automated systems. Some people may feel obsolete in society if their jobs become substituted and must migrate to new lower skill occupations or face unemployment.

The data collection technology today can jeopardize human freedom as deep learning analytics characterize people and take away human privacy. Systems can analyze what people speak, what pictures they share, what daily activities they conduct, and other information that could be used by organization to exploit them. Image recognition cameras used to monitor citizens alludes to an Orwellian world depicted in 1984 with constant government surveillance; yet, this technology exists in practice today. Anyone who disagrees with a government in this age could be tracked down with software systems and then removed. Without privacy, people cannot share ideas freely between peers and live in a democracy. Even in a world where data analytics are not used for taking out people by force, propaganda can be tailored to individuals through data collection and analysis and undermine

democracy. On digital media platforms, those in control of the data could determine who sees what content and influence how large groups of people think and behave in the world.

Many of the situations involving malevolent usage and loss of human autonomy and freedom could occur with technology available today or in the near future. Unlike the existential risks of misaligned superintelligence, the most immediate risks involve those imposed upon on society through human influence. These may not cause existential risk on the scale of superintelligence and robots that replicate beyond human control; however, the impacts on society could reach great heights before we reach this technology.

In summary, the risks of AI arise from misalignment with human values and humans or governments conducting immoral uses of the technology. Dangers include loss to human lives, terror organizations able to harm the world, and loss of democracy and freedom with AI systems. These risks grow as the technology finds its way into people's lives in various domains. The concern unique to AI from previous 20th century technologies comes with the ease of deployment in which people can find ways to misuse this technology. A small group of individuals could cause great harm to society through unnoticeable efforts until the effect becomes irreversibly known to the public.

4.2 Controversy of AI Outcomes

Despite recent developments in deep learning, intelligent systems still perform tasks in narrow domains. Neural networks take inspiration from biological brains; however, the size and complexity of the human brain dwarfs any deep learning network. In a subjective comparison, neural networks are no more complex than an ant brain. Furthermore, there is no one-to-one comparison between human brains and computer based neural networks. The hardware architecture, software architecture and learning algorithms that run computerized neural networks differ from biological systems. Deep learning models such as convolutional neural networks or recurrent neural networks have specific architectures defining how many neurons exist in each layer and which units connect. These models do not evolve over time into new structures across different tasks; rather, these structures and optimization functions work well for learning specific tasks. As a result, a trained network may have the ability to recognize images but has no greater understanding of language or other aspects of the world. Those implementing these models have a clear understanding of the outcomes because the software works within a narrow domain of tasks. The surprises occur when the algorithms make mistakes in unusual situations or performs better than expected. The probability of a neural network today reaching self-awareness and creating goals beyond humans would be as probable as an ant learning to speak from overhearing human conversations. The goals of deep learning models are explicitly defined as mathematical functions on restricted data inputs.

The narrow task restriction of current algorithms may appear as an impairment, but this restriction enables implementations to behave as intended. These algorithms have advantages over human cognition as human brains contain other functions not necessary for

AI tasks. Self-driving cars do not need to feel boredom while driving to align with human goals. Furthermore, developers can use specialized networks to perform tasks on high speed computer hardware using the smallest memory footprint and computations required. A complex network would not run as efficiently as a specialized network on a specific task. The tradeoff of narrow AI allows for compatibility on computer hardware and control of the model behavior.

The fact that AI performs specific tasks has created a duality of terms: the terms weak AI or narrow AI contrast strong AI, general AI or human level AI (Russell & Norvig, 2004). These terms represent a method of defining intelligence in relation to the range of tasks the system can perform. Humans are considered to have general intelligence with the ability to perform a wide variety of tasks in the world and transfer skills between domains. Human level intelligence is considered the singularity point in AI research because once humans know how to create system with all the functions as a human, this same system could build a more intelligent system and recursively improve itself. This system would surpass human intelligence, at which point the decisions of this system would outsmart any human agent. If this does occur, the outcomes would exceed any current risks since humans would have no ability to outsmart this system.

Without considering the physical danger of Artificial General Intelligence the socio-political debate would involve how humans fit into society if machines can perform all cognitive and physical labor. Humans would add no value to work since the machines could perform any task faster and cheaper. Issues would arise in terms of ownership. Individuals or government who own the intelligent systems would own all resources, since these systems would create all GDP and humans would not add value. Individual ownership, liberty and

freedom would disappear if general AI existed. Although this issue arises in a world with general AI, debates today may grow around these concerns even with current technology. If advanced narrow AI gives unsurmountable power to individuals or governments then people will face many of the same risks of losing human autonomy.

The largest existential risk arises from the uncontrollable nature of human level AI. This technology only exists in science fiction for now; however, any situation in which this becomes possible could lead to the end up human life on Earth. Bostrom explains that once an algorithm can demonstrate human level intelligence, it may quickly surpass the intelligence of any human since it could learn to improve itself recursively. Biological nerve impulses operate on the range of ten meters per second (Gay-Escoda, 2006) while electrical signals travel at the speed of light on the order of two hundred million meters per second. What would take a human thirty years to learn may take a computer brain several seconds. This speed depends on the architecture and algorithms of the computer, but the emulation of human cognition on electrical hardware could speed up thinking time by multiple orders of magnitude. Once a computer reached the intelligence of a human AI researcher, in a matter of minutes, this system could surpass human level intelligence (Bostrom, 2014). The risk in this situation comes from the misalignment of values that the system possesses. Perhaps the system's goal is to reach the highest intelligence possible. Once it realizes that it is limited to the circuitry on the emulating machine, it could learn to escape onto the Internet and convert the world into one large computer. It could outsmart humans through issuing commands to robots and killing all humans to continue learning. The simple answer would be to pull the plug or not allow it to connect to the Internet; however, if the system had human intelligence it may figure out that humans would eventually try to terminate the program to interfere with

the main goal. In this case it would come up with a scheme to either pretend it is not intelligent to evade human suspicion until the opportune moment or come up with an unthinkable strategy to get past humans (Bostrom, 2014). Having never experienced superintelligence, we do not know what actions or damage this system may choose to do if it decides that we are in the way of its goals. To the super intelligent machine, humans would be no greater than the ant colonies humans lay roads over (Harris, 2016).

In nearly any case, AI becomes close to impossible to contain if it reaches human intelligence. One of the few possible ways to ensure safe superintelligence would involve aligning the goals of AI with our own. In this way, super intelligent AI may escape but would ensure it does what humans want and does not formulate its own unwanted actions. This task itself would be hard to accomplish because if humans provide too vague an answer, the agent may misinterpret the implications. This theme was explored in iRobot where the central intelligence decides to kill humans to protect them from themselves. Even today in the social and political spectrum, people within countries and world powers cannot come to agree on precisely what constitutes human values and proper conduct.

The fact that deep learning fits into the narrow AI category implies that the risk of human level AI taking over the world remains science fiction. Few agree that super intelligent AI would not lead to existential risk; however, experts today debate about whether we should fear AI. The questions center around whether we could create a technology with human intelligence in the first place, before we already solved for impermeable safeguards, and whether the current systems merit greater risk for undoing humanity. None can give an objective prediction for human level AI since the no roadmap or general theory for this technology exists today. By the time human level AI becomes possible, intermediate

technology may automate the construction of impermeable safeguards. Furthermore, the implications of current technology may lead to greater society wide risks before general intelligence becomes possible. This has created a multifaceted debate where some are concerned about general intelligence, while others point to concerns of short-term risks such as lethal autonomous weapons and a global arms race. For those concerned with the short-term risks, the discussion of AI does not align with anti-technology views; rather, the goal of minimizing risk to ensure the maximum social benefit. One in this camp may dismiss the need to fear human level AI right now and focus on developing robust and safe narrow AI systems since they believe we may not reach super intelligence before global crises occur.

For those outside of the field of AI, the term itself may lead to miscommunication during discussions of preventative measures. AI can refer to machine learning, neural networks, expert systems, robotics or super intelligent general intelligence. Two people arguing about needing to fear of AI may be concerned about a different type of technology or timeline. Between a philosopher and an academic machine learning researcher, AI may mean human level general intelligence for one and supervised learning for the other. This misinterpretation technology capability may stifle beneficial uses and not adequately address potential misuse. This separate views on AI lead to different values. Deep learning pioneer Yann LeCun says not to fear AI since we are nowhere close to achieving human level AI (Shead, 2017). Bostrom argues we should not dismiss the future possibility of human level AI, as even the low chance of human level intelligence could lead to existential risks to humankind (Bostrom, 2014).

The overall challenge of AI stems from human misuse and mistakes. Anywhere from humans deploying poorly tested systems, using technology to control other humans to

unleashing superintelligence with poorly aligned human goals, the risk stems from us not taking the proper precautions with technology. Human technology exists today in the form of nuclear weapons that could undo humanity thus one should expect for advanced future technologies will also have dangerous implications. Whether the debate of AI involves preparing for human level intelligence or focusing on short term risks, the future of AI will require deliberate exercise of responsibility to implement in a safe and beneficial way for society.

4.3 Policy Guidelines for AI

Due to the controversies among experts in the field of AI, multiple organizations aim to address the ethical guidelines for future research. Some of these organizations include the Future of Life Institute, the Future of Humanity Institute, the AI Initiative and OpenAI. These organizations aim to research the impacts of AI and foster safe practices in future technology. The Asilomar principles were established and endorsed by highly esteemed experts in the field of AI including Demis Hassabis Ilya Sutskever, Yann LeCun, Yoshua Bengio, Stuart Russell, Peter Norvig, Ray Kurzweil as well as other notable figures such as Stephen Hawking, Elon Musk and Erik Brynjolfsson (AI Principles, 2017). These aim to provide guidelines for making the next steps in AI while avoiding the risks.

The Asilomar AI principles represent one set of AI ethics principles. Others include Charlevoix Common Vision for the Future of Artificial Intelligence, DeepMind Ethics & Society Principles, Google AI Principles, ITI AI Policy Principles (AI Policy, 2019). Many countries have AI strategies in place for the development of AI technology (AI Policy, 2019); however, no international agreement mandates policies for risks associated with future developments. Each group or firm may operate through various sets of principles or through self-created ethical guidelines. Currently, the Asilomar Principles serves one of the widest known AI principles, which have been endorsed by the State of California in 2017 and have also been endorsed by various AI development firms who may also have their own principles, such as Google DeepMind, Facebook, Apple and OpenAI (Future of Life Institute, 2018). While these principles may not represent every issue among AI research and serve as suggestions rather than mandated rules, these principles capture a wide scope of ethical issues for current and future AI technology. It is worth noting these guidelines may

serve as a greater influence on AI development within the Silicon Valley groups; however, these principles may have a lesser impact on international efforts with government mandated research goals in conflict with suggested guidelines. The following subsections will examine the three categories of the Asilomar principles

4.3.1 Research Issues

Table 4-1 Asilomar Principles Research Issues

1) Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
2) Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as: How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked? How can we grow our prosperity through automation while maintaining people's resources and purpose? How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI? What set of values should AI be aligned with, and what legal and ethical status should it have?
3) Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.
4) Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
5) Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

The first goal states that AI should be researched with human benefit in mind and not to solely push the limits on intelligence. This goal may be controversial in that it goes against the natural curiosity that motivates science, but the reason for this goal stems from the fact that AI poses risks to humanity. One should not research a technology that poses risk to

society without first putting safeguards in place. For instance, it would be unethical to research nuclear bombs just to make them bigger; instead, it would be more ethical to research beneficial uses of nuclear energy while taking precaution of ethical implications and risks. Overall, research in AI should focus on creating beneficial and robust AI that mitigates the risks discussed in previous sections of this paper. Not only will this prevent dangers, but this initiative will also enable a larger impact of AI where it can be safely deployed in high value, high risk situations.

The technological barriers that could lead to short term issues involve malfunctioning, non-transparent and hackable AI systems. These barriers predicate the need for collaborative research that solves these questions of understandable, robust and secure AI. Short risks go in line with autonomous vehicles causing car crashes or systems learning unknown biases that become points of failures in in widely deployed systems. Companies should not deploy systems without thorough verification because several unchecked errors could lead to millions of car crashes. As more consumers rely on AI for daily operations, developers will take on greater responsibility to not cut corners to push a product to market. Short term risks include misalignment with human values, people misusing systems through hacking and loss of freedom through privacy issues. Research into preventing these modes of failure could benefit future technology from falling into the risks outlined in section 4.1.

Issues three and four address the barriers between researchers, policy makers but fails to address the need for public consensus on the relevant issues on AI. From a democratic standpoint, the public should have an educated understanding of the state of technology and the implicit risks associated with modern and future developments. Without a common public understanding of AI, people may mistake AI as either fictional or cause for great

alarm. This could hinder the solutions to the issues sought in the second goal, as people may not know how to navigate occupational changes and largescale societal changes discussed in this paper. Overall, the AI research community should avoid keeping transparency only between other researchers and policy makers, as this could lead to misconceptions about AI and misappropriate responses to technological change.

4.3.2 Ethics and Values

Table 4-2 Asilomar Principles Ethics and Values

6) Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.
7) Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.
8) Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.
9) Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.
10) Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.
11) Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
12) Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.
13) Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
14) Shared Benefit: AI technologies should benefit and empower as many people as possible.
15) Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.

16) Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.
17) Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.
18) AI Arms Race: An arms race in lethal autonomous weapons should be avoided.

The ethics and values policies cover a wide set of guidelines that minimize risk of current technology. While these goals contain well intentioned ideas and guidelines, current AI technology may face incompatibility with these ideals. The first three guidelines are imperative for robust AI; however, implementing these values may involve large technological barriers. It remains a challenge to verify all situations of a system and for this reason, further research is necessary before this goal becomes realistic. As an example, section 2.2 explained the total number of states in the game of Go exceed the number of atoms in the observable universe. In addition, deep learning systems often come with a black box nature to ascertain reasons systems made specific decisions. This creates a tradeoff between accuracy and verifiability leading to a contradiction that may limit researchers from following guidelines. A system such as an autonomous vehicle could statistically reach higher safety ratings through accurate deep learning perception, but the system may lack verifiability for the event of a malfunction. These rules create a gray area for developers with different goals of safety over verifiability and they may set one or both policies aside.

Policy nine assigns responsibly to developers in the case of misuse, malfunction or implications. The issue comes with the implications that AI can become widely distributed and easily replicable if in the wrong hands (Joy, 2000). For instance, a developer working on

a system for image recognition may have good intentions and consider the technology to have purely beneficial implications; however, a developer of assassin drones could take the pretrained neural network model, retrain the weights to target humans, and deploy hundreds of drones to use image recognition to terrorize innocent people. In this case, the original developer, according to policy nine would may have contributed to this misuse. Unlike technologies of the past, such as nuclear weapons, the ease of replication and potential for small groups to abuse these systems makes this policy hard to substantiate without many AI developers deciding to cease and desist current research that could lead to negative implications.

Policies ten and eleven aim to align software with human values. There remains little understanding on how to feasibly encode human values and morals into AI systems. Current systems mimic human behavior or have numerical optimization goals in mind. To encode human decision making and ethics into systems remains a great challenge for future technology. Furthermore, attempts to put human values into AI could lead to larger risks as the values are subjective across different people.

Policies twelve and thirteen address issues of privacy and human choice over AI; however, some could argue that technology will inherently undermine these qualities as AI improves. In 1863, Samuel Butler provides an argument that even if we told ourselves we must deny technology, we would no sooner find ourselves unable to bring ourselves to accomplish this deed, proving that we are enslaved by technology (Yampolskiy, 2012). Some may decide to opt out of data collection and reserve privacy, but this would put the user at the expensive of receiving inferior AI services or opting out of services altogether. Consider the case of AI collecting genetic information to prolong quality of life. Choosing to opt out of

sharing data would mean opting out of life preserving care. At this point, the person does not have the choice over personal data so long as wanting to still receive the service. This situation would create a conflict of interest over choice of data use and could lead to issues in the cases of liberty, described in section 4.1. Policy fourteen also represents issues in that many people have ideologies not in line with others in terms of religion, culture and values. In today's socio-political spectrum, could we program an AI doctor to decide on performing abortions or would this decision conflict with the values of others? In other societies with different cultural and religious views, AI may subjugate some to fit the majority view or may accept minority views in conflict and with revolt from the majority view. Should AI enforce religious law if the majority finds this law most beneficial to society?

Policy fifteen conflicts the current expected impact of AI across occupations. While AI has projections to add trillions of dollars in gross world product, the impact could provide more value to the upper decile of income labor while reducing real earnings of low skill labor. How can advancements in AI operate within the bounds of increasing equality if the technology will likely increase inequality? Furthermore, conflicting views may arise on the societal benefits of high value AI. Some may argue that an increase in return on capital income from technology requires socialism policies with tax rates exceeding 80 percent on the top centile and 60 percent on the top decile for a fair society (Piketty, 2014). Others may argue the benefits from AI will entail improvements among the lives of even the poorest members of society with lower cost on goods or services, justifying a highly concentrated capital ownership among few in the upper centile of wealth that own and oversee AI technology.

Policy sixteen conflicts with a natural progression of AI technology in relation to the arguments made for policies twelve and thirteen. In the case of fully autonomous vehicles, a point may arrive in which humans no longer have the choice to gain control over a vehicle. The roads of the future may not permit human drivers if AI vehicles drive at higher speeds and at higher throughputs on public roads. Various technologies and services may face full automation in the future as AI improves, which would cause humans to depend on technology to live a functional life.

The last point suggests avoiding an AI arms race; however, this policy follows a similar issue as policy nine, wherein researchers may inadvertently contribute to the development of AI weapons. Terrorist organizations may covert other autonomous systems such as low cost delivery drones into weapons that would cause large damage even with a ban on autonomous weapons. What constitutes an autonomous weapon may have different meanings, as an autonomous plane could be turned into a lethal autonomous weapon. Even if most nations ban autonomous weapons, the risk still exists for some nations to secretly manufacture lethal AI. Other nations would need to manufacture AI weapons to have a fighting chance against a nation or terrorist group breaking agreements. Putting a ban on lethal AI might incentivize some countries to develop this technology in secret as a defense measure. Some may argue that banning autonomous weapons may cause more harm than good, as this would unarm nations vulnerable of attack by terrorist organizations. Any global organization to stop a rogue nation from overtaking the world would need lethal autonomous weapons in the first place to combat this nation. Policies put in place to prevent lethal autonomous weapons may be well intentioned to mitigate risk; however, even prudent policies could lead to dangerous outcomes. While lethal autonomous weapons possess high

risk to society, there is not a simple answer on how to avoid the implications without greater international discussion and coordination.

4.3.3 Longer-term Issues

Table 4-3 Asilomar Principles Longer-term Issues

19) Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.
20) Importance: Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.
21) Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.
22) Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.
23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization.

The Asilomar Principles offer helpful guidelines on future issues concerning AI that could reach superintelligence. Bostrom argues that superintelligence could lead to one of the largest existential risks to humanity (Bostrom, 2014). Researchers should confront the risk of superintelligence as a long-term goal and find ways to develop safeguards and strategies to prevent this risk if the technology appears on the horizon. These long-term guidelines offer insight into the concern of AI exceeding human control whether through superior intelligence or through widescale control by organizations. These policies address the risks highlighted in section 4.1, namely, the misalignment with human values, malevolent use and loss of

autonomy. Nevertheless, these guidelines will not offer a solution to all the risks faced with AI; rather, theses policies serve as a starting point to promote further discussion and unification of common goals for the future good of this technology. Certain scenarios presented in previous sections could play out even if strict adherence to these guidelines were followed. For instance, attempting to avoid an AI arms race and aiming to exercise caution in AI research could still lead to ill-intentioned groups misusing AI technology and harming many.

4.4 Summary

AI has short term and long-term risks that will grow as the technology becomes widespread in society. These risks include misalignment with human values, malicious use by criminal people and gradual loss of individual liberty in society. In the short term, these risks manifest in autonomous vehicles causing car crashes, humans using lethal AI to subjugate people and loss of privacy for civilians. The more that human life rests in the hands of this technology the greater the risk of harm and misuse.

The short-term risks of AI depend on human uses and decisions. Difficulty exists in stopping corrupt organizations from hacking high impact systems or deploying lethal robots, but the risks originate from human wrongdoing. Policies that reduce human misuse could reduce the risk of dangerous AI in the short term. This will require diplomacy between organization and governments. Market competitors will need to avoid racing products to market without stringent testing to cut corners. Larger organizations and governments could misuse AI through mass surveillance, expropriation of privacy and in lethal AI arms races. Those in control of large intelligence systems will have greater power to exploit people. We can hope our local governments will uphold freedom and democracy; however, this right is not guaranteed across the world and has not been through human history.

Debate over the ethical implications exists in the technology world. Some fear that AI will take over the world, destroy human jobs, assassinate people or assimilate all of Earth's resources. Most academics will point out that as the technology stands, AI poses little risk of reaching a mind of its own and surpassing humans. When academics agree on the risks of AI, this agreement involves the short-term risks that exist with current technology including those listed as misaligning with humans, enabling misuse of terror organizations and damage to society from privacy issues and government surveillance. Nevertheless, those arguing for

the possibility of human level AI have pointed out that without safeguards, the entire human species could cease to exist. Any human level AI could expand beyond our control and misalign with our values, destroying us and the planet in the process. Although this technology has no theoretical implementation foreseeable in the near future, the small change of reaching this technology merits studying how we may avoid global catastrophe.

In response to these risks of AI, a growing number of experts, scientists and technology leaders have set forth institutes to explore means of mitigating future risks. The Asilomar Principles have gathered a large interest to address the risks of both short-term and long-term risks. The first main component of these principles is to push research in a direction that benefits humanity. These include policies to create robust and non-hackable systems and to emphasize testing rather than cutting corners to beat competition to markets. These goals go in line with the high probability of wide scale AI in our society in the near future. Thorough testing will mitigate risks of millions of people affected from malfunctions or misuses.

The Asilomar Principles provide ethics and value considerations when developing AI. These policies include creating systems with transparency and judiciary abilities for people to ascertain why a system chose a wrong decision. Further guidelines include human aligned goals, privacy considerations and avoidance of an arms race. While these guidelines have good intentions for the design future systems, the reality of achieving these goals may have shortcomings even with good intentions. The ability to accurately implement human goals into systems remains a mystery. Humans possess complex emotions, values and purposes in life which cannot be programmed directly into a machine. A system attempting to mimic these values may not function correctly. In terms of government policy, developers who

design systems in a country that has policies to ensure human rights and privacy may indirectly equip other governments with the ability to subjugate billions of citizens. Governments that fear lethal autonomous weapons from terrorist organizations may feel obligated to secretly develop lethal AI to protect its citizens. The greater decisions concerning how people use AI will involve political debate since these risks do not have straightforward solutions.

While human AI could pose an existential risk to humanity, knowledgeable individuals should not spread fear about AI. Policymakers, researchers and industry leaders should address the risks of current technology while encouraging research into safeguards of more powerful future technology. People should have clear information about the current economic impacts and ethical implications of AI with the ability to separate science fiction from reality. The digital revolution and wide scale adoption will change human society across all walks of life. To navigate future outcomes, people should understand the capabilities and risks of AI.

5. Conclusion

AI today could evolve into a much broader technology in contrast to the shortcomings of AI in the 20th century. AI technology will increase global economic output through increasing efficiency of firms, enable innovative technologies and accelerate scientific discoveries. This could account for a total increase in gross world product on the order of $10 Trillion to $20 Trillion over the next several decades. These largescale changes will offer large income increases to developers and adopters of these technologies; however, the widescale adoption of AI could substitute tasks and reduce earnings of many workers. Up to 47 percent of workers could either face substitution or wage stagnation. This will create job volatility and could put strain on workers at low incomes without means to reinvest in complimentary skills. The likelihood of widescale unemployment may be lower than the prospect of a large income inequality, which may lead to other social consequences. AI may provide an overall positive impact to the world but does not guarantee a direct solutional to income inequality.

Widescale adoption of AI technology leads to ethical concerns where short-term risks comprise mainly of ill-intentioned use and a progressive abdication of human autonomy. Debate commonly centers around the future implications of AI misalignment with human values; however, issues involving the ease of misuse have become a topic of modern debate. Public perception may fear AI in terms of the former risk without understanding the implications that modern technology could pose substantial risk in terms of the latter. Future policy should aim for realistic guidelines in respect to issues considering the economic forces with knowledge the economic benefits come with inherent risks and tradeoffs. Additionally,

transparency with the public discourse needs greater focus such that people may navigate economic and societal change brought from wide adoption of AI technology.

6. Bibliography

A. Heinz, Ernst. *Scalable Search in Computer Chess*, 2000. https://doi.org/10.1007/978-3-322-90178-1.

A Roadmap for U.S. Robotics: From Internet to Robotics. Georgia Institute of Technology, 2013.

Acemoglu, Daron. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *The Quarterly Journal of Economics* 113, no. 4 (November 1, 1998): 1055–89. https://doi.org/10.1162/003355398555838.

Acemoglu, Daron, and David Autor. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*, 4B:1043–1171. Elsevier, 2011. https://ideas.repec.org/h/eee/labchp/5-12.html.

Ackerman, Evan. "We Should Not Ban 'Killer Robots,' and Here's Why." IEEE Spectrum: Technology, Engineering, and Science News, July 29, 2015. https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/we-should-not-ban-killer-robots.

Adee, Sally. "37 Years of Moore's Law." IEEE Spectrum: Technology, Engineering, and Science News, May 1, 2008. https://spectrum.ieee.org/computing/hardware/37-years-of-moores-law.

Ahn, B. "Real-Time Video Object Recognition Using Convolutional Neural Network." In *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–7, 2015. https://doi.org/10.1109/IJCNN.2015.7280718.

"AI Initiative | Artificial Intelligence." Accessed October 29, 2018. http://ai-initiative.org/.

"AI Policy." Future of Life Institute. Accessed March 3, 2019. https://futureoflife.org/ai-policy/.

"AI Principles." Future of Life Institute. Accessed October 29, 2018. https://futureoflife.org/ai-principles/.

Alipanahi, Babak, Andrew Delong, Matthew T. Weirauch, and Brendan J. Frey. "Predicting the Sequence Specificities of DNA- and RNA-Binding Proteins by Deep Learning." *Nature Biotechnology* 33, no. 8 (August 2015): 831–38. https://doi.org/10.1038/nbt.3300.

Alpaydin, Ethem. *Machine Learning: The New AI*. Cambridge, MA: The MIT Press, 2016.

Autor, David H., and David Dorn. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103, no. 5 (August 2013): 1553–97. https://doi.org/10.1257/aer.103.5.1553.

Autor, David H., Frank Levy, and Richard J. Murnane. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118, no. 4 (November 1, 2003): 1279–1333. https://doi.org/10.1162/003355303322552801.

Baldi, P., P. Sadowski, and D. Whiteson. "Searching for Exotic Particles in High-Energy Physics with Deep Learning." *Nature Communications* 5 (July 2, 2014): 4308. https://doi.org/10.1038/ncomms5308.

Bay, John. "Exoskeleton Market Is Expected To Be Worth US$ 2.5 Billion by 2024." MarketWatch, 2018. https://www.marketwatch.com/press-release/exoskeleton-market-is-expected-to-be-worth-us-25-billion-by-2024-2018-05-14.

"Benefits & Risks of Artificial Intelligence." Future of Life Institute. Accessed October 29, 2018. https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/.

Bessen, James E. "AI and Jobs: The Role of Demand." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, November 28, 2017. https://papers.ssrn.com/abstract=3078715.

———. "How Computer Automation Affects Occupations: Technology, Jobs, and Skills." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, October 3, 2016. https://papers.ssrn.com/abstract=2690435.

Bhattacharya, Biswarup, and Abhishek Sinha. "Intelligent Fault Analysis in Electrical Power Grids." *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, November 2017, 985–90. https://doi.org/10.1109/ICTAI.2017.00151.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford, New York: Oxford University Press, 2014.

Bostrom, Nick, and Eliezer Yudkowsky. "The Ethics of Artificial Intelligence." The Cambridge Handbook of Artificial Intelligence, June 2014. https://doi.org/10.1017/CBO9781139046855.020.

Brynjolfsson, Erik, and Andrew McAfee. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York, NY, US: W W Norton & Co, 2014.

Buchanan, Bruce G. "A (Very) Brief History of Artificial Intelligence." *AI Magazine* 26, no. 4 (December 15, 2005): 53–53. https://doi.org/10.1609/aimag.v26i4.1848.

Bughin, Jacques, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi. "Modeling the Global Economic Impact of AI | McKinsey." Accessed March 2, 2019. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy.

Cheng, Selina. "The Awful Frustration of a Teenage Go Champion Playing Google's AlphaGo." Quartz. Accessed October 29, 2018. https://qz.com/993147/the-awful-frustration-of-a-teenage-go-champion-playing-googles-alphago/.

Clifford, Catherine. "Elon Musk at SXSW: A.I. Is More Dangerous than Nuclear Weapons," March 13, 2018. https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html.

Conniff, Richard. "What the Luddites Really Fought Against." Smithsonian. Accessed October 29, 2018. https://www.smithsonianmag.com/history/what-the-luddites-really-fought-against-264412/.

"Convolutional Neural Network." Accessed October 29, 2018. https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html.

Copeland, B. J. "Colossus: Its Origins and Originators." *IEEE Annals of the History of Computing* 26, no. 4 (October 2004): 38–45. https://doi.org/10.1109/MAHC.2004.26.

Copeland, Jack, and Diane Proudfoot. "Turing, Father of the Modern Computer." Accessed October 29, 2018. http://www.rutherfordjournal.org/article040101.html.

Costello, H. "Global E-Learning Market 2017 to Boom $275.10 Billion Value by 2022 at a CAGR of 7.5% – Orbis Research - Reuters." Accessed March 2, 2019. https://www.reuters.com/brandfeatures/venture-capital/article?id=11353.

cycles, This text provides general information Statista assumes no liability for the information given being complete or correct Due to varying update, and Statistics Can Display More up-to-Date Data Than Referenced in the Text. "Topic: E-Learning and Digital Education." www.statista.com. Accessed March 2, 2019. https://www.statista.com/topics/3115/e-learning-and-digital-education/.

DeepMind. "AlphaFold: Using AI for Scientific Discovery." DeepMind. Accessed March 2, 2019. https://deepmind.com/blog/alphafold/.

Deming, David J. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics* 132, no. 4 (2017): 1593–1640.

Etzioni, Amitai, and Oren Etzioni. "Pros and Cons of Autonomous Weapons Systems." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, April 28, 2017. https://papers.ssrn.com/abstract=2960301.

"Facebook Analytics." Facebook Analytics: Drive Growth to Web, Mobile & More. Accessed October 29, 2018. https://analytics.facebook.com/.

Frey, Carl Benedikt, and Michael A. Osborne. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114, no. C (2017): 254–80.

Future of Life Institute. "State of California Endorses Asilomar AI Principles." Future of Life Institute, August 31, 2018. https://futureoflife.org/2018/08/31/state-of-california-endorses-asilomar-ai-principles/.

Garner, R. “Early Popular Computers, 1950 - 1970 - Engineering and Technology History Wiki.” Accessed October 29, 2018. https://ethw.org/Early_Popular_Computers,_1950_-_1970.

Gee, Colby. “A Summary of Augmented Reality and Virtual Reality Market Size Predictions.” *Medium* (blog), November 18, 2018. https://medium.com/vr-first/a-summary-of-augmented-reality-and-virtual-reality-market-size-predictions-4b51ea5e2509.

Gil, Y., E. Deelman, M. Ellisman, T. Fahringer, G. Fox, D. Gannon, C. Goble, M. Livny, L. Moreau, and J. Myers. “Examining the Challenges of Scientific Workflows.” *Computer* 40, no. 12 (December 2007): 24–32. https://doi.org/10.1109/MC.2007.421.

Gilchrist, Karen. “Chatbots Expected to Cut Business Costs by $8 Billion by 2022,” May 9, 2017. https://www.cnbc.com/2017/05/09/chatbots-expected-to-cut-business-costs-by-8-billion-by-2022.html.

Gillham, Jonathan. *The Macroeconomic Impact of Artificial Intelligence*, 2017. https://doi.org/10.13140/RG.2.2.21506.38083.

Glantz, Morton, and Robert Kissell. *Multi-Asset Risk Modeling: Techniques for a Global Economy in an Electronic and Algorithmic Trading Era*. Academic Press, 2013.

Goodall, Noah J. “Machine Ethics and Automated Vehicles,” 2014.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Nets.” In *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, 2672–2680. Curran Associates, Inc., 2014. http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

“Google Acquires Artificial Intelligence Startup DeepMind For More Than $500M.” *TechCrunch* (blog). Accessed October 29, 2018. http://social.techcrunch.com/2014/01/26/google-deepmind/.

Gopnik, Alison. “An AI That Knows the World Like Children Do.” Scientific American. Accessed February 28, 2019. https://doi.org/10.1038/scientificamerican0617-60.

Halim, M., K. M. Ho, and A. Liu. “Fuzzy Logic for Medical Expert Systems.” *Annals of the Academy of Medicine, Singapore* 19, no. 5 (September 1990): 672–83.

Hauser, Larry. “Chinese Room Argument | Internet Encyclopedia of Philosophy.” Accessed March 3, 2019. https://www.iep.utm.edu/chineser/.

Haykin, Simon. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition,” December 10, 2015. https://arxiv.org/abs/1512.03385.

Hutchens, Jason. “How to Pass the Turing Test by Cheating.” University of Western Australia, 1997.

“IGCSE ICT - Expert Systems | IGCSE ICT.” Accessed October 29, 2018. https://www.igcseict.info/theory/7_2/expert/.

INRIX. “Annual Cost Of Gridlock In Europe And The US Will Increase 50 Percent On Average To $293 Billion By 2030.” Accessed March 2, 2019. https://www.prnewswire.com/news-releases/annual-cost-of-gridlock-in-europe-and-the-us-will-increase-50-percent-on-average-to-293-billion-by-2030-279094111.html.

Institute, McKinsey Global, James Manyika, Michael Chui, Jacques Buguin, Richard Dobbs, Peter Bisson, and Alex Marrs. *Disruptive Technologies: Advances That Will Transform Life, Business, and the Global Economy*. McKinsey Global Institute, 2013.

“Intelligence | Definition of Intelligence in English by Oxford Dictionaries.” Oxford Dictionaries | English. Accessed October 29, 2018. https://en.oxforddictionaries.com/definition/intelligence.

Jain, Aditya, Ramta Bansal, Avnish Kumar, and KD Singh. “A Comparative Study of Visual and Auditory Reaction Times on the Basis of Gender and Physical Activity Levels of Medical First Year Students.” *International Journal of Applied and Basic Medical Research* 5, no. 2 (2015): 124–27. https://doi.org/10.4103/2229-516X.157168.

Jiang, Fei, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. “Artificial Intelligence in Healthcare: Past, Present and Future.” *Stroke and*

Vascular Neurology 2, no. 4 (December 1, 2017): 230–43. https://doi.org/10.1136/svn-2017-000101.

"John McCarthy - A.M. Turing Award Laureate." Accessed February 28, 2019. https://amturing.acm.org/award_winners/mccarthy_1118322.cfm.

Joy, Bill. "Why the Future Doesn't Need Us." *Wired*, April 1, 2000. https://www.wired.com/2000/04/joy-2/.

Kadurin, Artur, Alexander Aliper, Andrey Kazennov, Polina Mamoshina, Quentin Vanhaelen, Kuzma Khrabrov, and Alex Zhavoronkov. "The Cornucopia of Meaningful Leads: Applying Deep Adversarial Autoencoders for New Molecule Development in Oncology." *Oncotarget* 8, no. 7 (December 22, 2016): 10883–90. https://doi.org/10.18632/oncotarget.14073.

Karras, Tero, Timo Aila, Samuli Laine, and Jaakko Lehtinen. "Progressive Growing of GANs for Improved Quality, Stability, and Variation." *ArXiv:1710.10196 [Cs, Stat]*, October 27, 2017. http://arxiv.org/abs/1710.10196.

Kim, Eugene. "Amazon's Echo and Alexa Could Add $11 Billion in Revenue by 2020." Business Insider. Accessed March 2, 2019. https://www.businessinsider.com/amazon-echo-alexa-add-11-billion-in-revenue-by-2020-2016-9.

Korosec, K. "Waymo Launches Self-Driving Car Service Waymo One." *TechCrunch* (blog). Accessed March 2, 2019. http://social.techcrunch.com/2018/12/05/waymo-launches-self-driving-car-service-waymo-one/.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning." *Nature* 521, no. 7553 (May 2015): 436–44. https://doi.org/10.1038/nature14539.

Lillicrap, Timothy P., Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. "Continuous Control with Deep Reinforcement Learning," September 9, 2015. https://arxiv.org/abs/1509.02971.

LLC, Revolvy. "'Trenchard More' on Revolvy.Com." Accessed February 28, 2019. https://www.revolvy.com/page/Trenchard-More?smv=3057054.

MacCrory, Frank, George Westerman, Yousef AlHammadi, and Erik Brynjolfsson. "Racing With and Against the Machine: Changes in Occupational Skill Composition in an Era of Rapid Technological Advance." In *ICIS*, 2014.

MacCrory, Frank, George Westerman, and Erik Brynjolfsson. "Identifying the Multiple Skills in Skill-Biased Technical Change." In *ICIS*, 2015.

"Magnetic Core Memory - CHM Revolution." Accessed February 28, 2019. https://www.computerhistory.org/revolution/memory-storage/8/253.

Malani, Gunjan. "Global Service Robotics Market Is Expected to Reach $34.7 Billion by 2022." Allied Market Research, 2015. https://www.alliedmarketresearch.com/service-robotics-market.

Martin, C. D. "ENIAC: Press Conference That Shook the World." *IEEE Technology and Society Magazine* 14, no. 4 (Winter): 3–10. https://doi.org/10.1109/44.476631.

Matiisen, Tambet. "Demystifying Deep Reinforcement Learning | Computational Neuroscience Lab." Accessed October 29, 2018. https://neuro.cs.ut.ee/demystifying-deep-reinforcement-learning/.

Menzies, T. "21st-Century AI: Proud, Not Smug." *IEEE Intelligent Systems* 18, no. 3 (May 2003): 18–24. https://doi.org/10.1109/MIS.2003.1200723.

"MGI-Notes-from-the-AI-Frontier-Modeling-the-Impact-of-AI-on-the-World-Economy-September-2018.Pdf." Accessed March 2, 2019. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.ashx.

"Moore's Law | Computer Science." Encyclopedia Britannica. Accessed February 28, 2019. https://www.britannica.com/technology/Moores-law.

Mousavi, Seyed Sajad, Michael Schukat, and Enda Howley. “Deep Reinforcement Learning: An Overview,” June 23, 2018. https://doi.org/10.1007/978-3-319-56991-8_32.

Murphy, A. “Domestic: Robotics Outlook 2025.” Loup Ventures, June 7, 2017. https://loupventures.com/domestic-robotics-outlook-2025/.

Murphy, Julia, and Max Roser. “Internet.” Our World in Data. Accessed October 29, 2018. https://ourworldindata.org/internet.

“Neural Networks - History.” Accessed October 29, 2018. https://cs.stanford.edu/people/eroberts/courses/soco/projects/neural-networks/History/history1.html.

“Number of Possible Go Games at Sensei’s Library.” Accessed October 29, 2018. https://senseis.xmp.net/?NumberOfPossibleGoGames.

“Open Letter on Autonomous Weapons.” Future of Life Institute. Accessed October 29, 2018. https://futureoflife.org/open-letter-autonomous-weapons/.

“(PDF) Future Progress in Artificial Intelligence: A Survey of Expert Opinion.” ResearchGate. Accessed March 1, 2019. http://dx.doi.org/10.1007/978-3-319-26485-1_33.

Piketty, Thomas. *Capital in the Twenty-First Century*. Cambridge Massachusetts: The Belknap Press of Harvard University Press, 2014.

“Playing Atari with Deep Reinforcement Learning.” DeepMind. Accessed March 1, 2019. https://deepmind.com/research/publications/playing-atari-deep-reinforcement-learning/.

Raj, Manav, and Robert Seamans. “AI, Labor, Productivity, and the Need for Firm-Level Data.” *The Economics of Artificial Intelligence: An Agenda*, May 14, 2018. https://www.nber.org/chapters/c14037.

Rathmanner, Samuel, and Marcus Hutter. “A Philosophical Treatise of Universal Induction.” *Entropy* 13, no. 6 (June 3, 2011): 1076–1136. https://doi.org/10.3390/e13061076.

Rich, C, and R Waters. *Readings in Artificial Intelligence and Software Engineering*. Elsevier, 1986. https://doi.org/10.1016/C2013-0-07696-7.

Richards, Neil M., and William D. Smart. “How Should the Law Think About Robots?” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, May 10, 2013. https://papers.ssrn.com/abstract=2263363.

“Robotics Market Revenue Worldwide 2017/2025 | Statistic.” Statista. Accessed March 2, 2019. https://www.statista.com/statistics/760190/worldwide-robotics-market-revenue/.

Rose, Victoria. “How Pros and Regular Players Beat Elon Musk’s Beloved Esports Bot.” The Flying Courier, September 11, 2017. https://www.theflyingcourier.com/2017/9/11/16285390/elon-musk-open-ai-esports-bot-dota-2-defeated-beaten.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. “ImageNet Large Scale Visual Recognition Challenge.” *International Journal of Computer Vision* 115, no. 3 (December 1, 2015): 211–52. https://doi.org/10.1007/s11263-015-0816-y.

Russell, Stuart, Daniel Dewey, and Max Tegmark. “Research Priorities for Robust and Beneficial Artificial Intelligence.” *ArXiv:1602.03506 [Cs, Stat]*, February 10, 2016. http://arxiv.org/abs/1602.03506.

Russell, Stuart, Sabine Hauert, Russ Altman, and Manuela Veloso. “Robotics: Ethics of Artificial Intelligence.” *Nature* 521, no. 7553 (May 28, 2015): 415–18. https://doi.org/10.1038/521415a.

Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2009.

———. *Artificial Intelligence: A Modern Approach*. 3 edition. Upper Saddle River: Pearson, 2009.

Sandrini, Francisco Aurelio Lucchesi, Edwaldo Dourado Pereira-Júnior, and Cosme Gay-Escoda. “Rabbit Facial Nerve Anastomosis with Fibrin Glue: Nerve Conduction Velocity Evaluation.” *Revista Brasileira de Otorrinolaringologia* 73, no. 2 (April 2007): 196–201. https://doi.org/10.1590/S0034-72992007000200009.

Sauer, Sam. “5 Stats from the State of Project Management in Manufacturing Report.” LiquidPlanner, May 11, 2017. https://www.liquidplanner.com/blog/stats-2017-project-management-manufacturing-report/.

Schaller, R. R. “Moore’s Law: Past, Present and Future.” *IEEE Spectrum* 34, no. 6 (June 1997): 52–59. https://doi.org/10.1109/6.591665.

Serbera, Jean-Philippe. “Flash Crashes: If Reforms Aren’t Ramped up, the next One Could Spell Global Disaster.” The Conversation. Accessed March 3, 2019. http://theconversation.com/flash-crashes-if-reforms-arent-ramped-up-the-next-one-could-spell-global-disaster-109362.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. “Mastering the Game of Go without Human Knowledge.” *Nature* 550, no. 7676 (18 2017): 354–59. https://doi.org/10.1038/nature24270.

Singh, S. “Mobile Robots Market Worth $54.1 Billion by 2023,” 2018. https://www.marketsandmarkets.com/PressReleases/mobile-robots.asp.

———. “Surgical Robots Market Worth $6.5 Billion by 2023.” Accessed March 2, 2019. https://www.marketsandmarkets.com/PressReleases/surgical-robots.asp.

Solon, Olivia. “More than 70% of US Fears Robots Taking over Our Lives, Survey Finds.” *The Guardian*, October 4, 2017, sec. Technology. https://www.theguardian.com/technology/2017/oct/04/robots-artificial-intelligence-machines-us-survey.

Sotos, J. G. “MYCIN and NEOMYCIN: Two Approaches to Generating Explanations in Rule-Based Expert Systems.” *Aviation, Space, and Environmental Medicine* 61, no. 10 (October 1990): 950–54.

Spark, Andrew. “Obituary: Oliver Selfridge.” *The Guardian*, December 17, 2008, sec. Technology. https://www.theguardian.com/technology/2008/dec/17/oliver-selfridge-obituary.

“Stanford University CS224d: Deep Learning for Natural Language Processing.” Accessed October 29, 2018. http://cs224d.stanford.edu/.

Sutton, Richard S., and Andrew G. Barto. *Introduction to Reinforcement Learning*. 1st ed. Cambridge, MA, USA: MIT Press, 1998.

Takeuchi, Lawrence. “Applying Deep Learning to Enhance Momentum Trading Strategies in Stocks,” 2013.

“The Costs of Cancer.” American Cancer Society Cancer Action Network, April 11, 2017. https://www.fightcancer.org/policy-resources/costs-cancer.

“The Dartmouth Artificial Intelligence Conference: The next 50 Years.” Accessed October 29, 2018. http://www.dartmouth.edu/~ai50/homepage.html.

“The Errors, Insights and Lessons of Famous AI Predictions – and What They Mean for the Future: Journal of Experimental & Theoretical Artificial Intelligence: Vol 26, No 3.” Accessed February 28, 2019. https://www.tandfonline.com/doi/abs/10.1080/0952813X.2014.895105.

“The Future Economic and Environmental Costs of Gridlock in 2030: An Assessment of the Direct and Indirect Economic and Environmental Costs of Idling in Road Traffic Congestion to Households in the UK, France, Germany and the USA,” July 2014. https://trid.trb.org/view/1329874.

“The Radicati Group Releases ‘EDiscovery Market, 2014-2018.’” Marketwire. Accessed March 2, 2019. http://www.marketwired.com/press-release/the-radicati-group-releases-ediscovery-market-2014-2018-1973577.htm.

Thrun, S. “Winning the DARPA Grand Challenge.” *IFAC Proceedings Volumes*, 4th IFAC Symposium on Mechatronic Systems, 39, no. 16 (January 1, 2006): 1. https://doi.org/10.3182/20060912-3-DE-2911.00002.

Tickle, A. B., R. Andrews, M. Golea, and J. Diederich. “The Truth Will Come to Light: Directions and Challenges in Extracting the Knowledge Embedded within Trained Artificial Neural Networks.” *IEEE Transactions on Neural Networks* 9, no. 6 (November 1998): 1057–68. https://doi.org/10.1109/72.728352.

UK, Sam Shead, Business Insider. “Facebook’s AI Boss: ‘In Terms of General Intelligence, We’re Not Even Close to a Rat.’” Business Insider. Accessed October 29, 2018.

https://www.businessinsider.com/facebooks-ai-boss-in-terms-of-general-intelligence-were-not-even-close-to-a-rat-2017-10.

vemuri, vinay. "Understanding Recommendation Systems-101." *Coinmonks* (blog), July 11, 2018. https://medium.com/coinmonks/understanding-recommendation-systems-101-f81eb1afcad.

Vincent, James. "Facebook's Head of AI Wants Us to Stop Using the Terminator to Talk about AI." The Verge, October 26, 2017. https://www.theverge.com/2017/10/26/16552056/a-intelligence-terminator-facebook-yann-lecun-interview.

Wu, Desheng Dash, Shu-Heng Chen, and David L. Olson. "Business Intelligence in Risk Management: Some Recent Progresses." *Information Sciences*, Business Intelligence in Risk Management, 256 (January 20, 2014): 1–7. https://doi.org/10.1016/j.ins.2013.10.008.

Wyse, L. "Audio Spectrogram Representations for Processing with Convolutional Neural Networks." *ArXiv:1706.09559 [Cs]*, June 28, 2017. http://arxiv.org/abs/1706.09559.

Yale University, Institute of Human Relations. *Motivation and Reward in Learning*, 1948. http://archive.org/details/Motivati1948.

Yampolskiy, Roman V. "Taxonomy of Pathways to Dangerous AI." *ArXiv:1511.03246 [Cs]*, November 10, 2015. http://arxiv.org/abs/1511.03246.

Yang, Gerry, and Chris Smith. "The History of Artificial Intelligence." History of Computing CSEP 590A, University of Washington, 2006.