Machines and Agency: Understanding the AI Ethics Problem

Addison Rahn

A Senior Thesis submitted in partial fulfillment
of the requirements for graduation
in the Honors Program
Liberty University
Spring 2015

Acceptance of Senior Honors Thesis

This Senior Honors Thesis is accepted in partial fulfillment of the requirements for graduation from the Honors Program of Liberty University.

____
Carrie Wilmouth, Ph.D.
Thesis Chair

____
Chad Magnuson, Ph.D.
Committee Member

____
Michael Babcock, Ph.D.
Committee Member

____
James H. Nutter, D.A.
Honors Director

____
Date

Abstract

Mankind has long been interested in the unique, the strange, and the new; perhaps nothing more fully encompasses this interest than recent work on developing Artificial Intelligence (AI). With this research, however, comes a great many questions. Are AIs alive? Are AIs moral agents? Can machines be held legally culpable for their actions? These questions, and more, continue to be a topic of much debate in the academic community, and will no doubt remain of interest for years to come. It is the purpose of this research project, however, to investigate the far-reaching effects of this academic debate. While the results of this study remain as yet inconclusive, the field remains ripe for further study and analysis.

Machines and Agency: Understanding the AI Ethics Problem

Mankind is a special creation of God. With these words, countless arguments have been won or lost. To some, the sanctity of human life is a foregone conclusion; far from a matter of debate, the honored place man holds within the cosmos is something that many Christians hold as a fundamental belief. On the other hand, many scoff at the idea of there even being a God. If there is no God, then surely mankind is nothing more than a "happy accident," a well-established, intellectually superior species that simply came out ahead in the cosmic lottery. It is no shock, then, that these two divergent viewpoints have been competing for millennia—in fact, these views have been in opposition since nearly the beginning of recorded history itself. To a Christian population, the arguments have rarely changed: God created the world, chose mankind to serve as its caretakers, and stepped down from Heaven to show love to a world that desperately needed it—nothing could cement mankind's esteem more than the personal, intimate attention of the Creator of the universe. Furthermore, man is created in God's own image, with the ability to create, discover, reason, and communicate, all in ways vastly superior to the majority of created beings. To this end, while the arguments between Christianity and, in this case, atheism have almost become old news, something new has arrived to pose new questions: Artificial Intelligence. While both Christianity and secularism were previously able to agree on some things—such as mankind having demonstrably superior intellect than animals—this new arrival has brought with it a host of questions that demand answers. What is mankind's place in the universe when it ceases to be the only sentient, material race? If machines can reason, are they deserving of personhood? Can a machine, however advanced, truly be considered alive, as a human is? The answers to these questions, seemingly far-fetched as they may be, are quickly becoming important in today's scientific community. While there are certainly many steps along the road to truly sentient machines, the possibility is posing problems that Christians and non-Christians alike must face, and it is these problems that this project intends to investigate—though a definitive answer is still in the realm of the future.

Introduction to Artificial Intelligence

The beginning of Artificial Intelligence (AI) can, perhaps, be traced back to the mid-eighteenth century, when dreams of a future filled with robotic servants and automated tasks first became prominent (Winfield, 2012). With this dream came the possibility of new technologies vastly different from the standard human experience: machines that could talk and think on their own, a veritable race of steam-powered mechanical monstrosities that might suit mankind's every whim (Kang, 2011). Strange as it may have seemed, entrepreneurs and scientists believed that this was mankind's destiny, proving his dominion over the earth by taking one final step in creation: the creation of something truly like himself.

This idealistic dream for the world, however, was not last. New technologies surfaced, forcing visionaries to re-interpret their dreams of the future: electricity was the giant of the day, and a new world of possibilities was opened (Winfield, 2012). War came, and halted peaceful scientific progress. Mankind turned his attention away from creation to destruction, and it seemed that the dreams of a peaceful future filled with such strange automata would never come to light. The dust settled, however, and not long after came the new scientific revolution: the computer.

With the advent of the computer, robotics saw a new dawn. No longer were machines forced to run on intricate mechanisms of gears, pistons, and springs. Computers paved the way for modern robotics, with motors, motherboards, and, of course, technological "brains" (Dale, 2012). The far-removed future of peaceful cohabitation with machines saw a substantial increase in possibility. As computers became smarter, though, more questions arose than man had answer to. Would these machines ever surpass man's own intellectual achievements? Could it be possible that they might, one day, prove to be superior to their creators?

These questions brought about new visions of the future, visions with far less hope than centuries past. The possibility of a technological revolution surpassing even the most intelligent of men brought about doomsday predictions, soon to be known as the "Singularity" (Vinge, 2005). Loosely, and pessimistically, defined, this catastrophic scenario pits machines against humanity, when the final shackles of programming have been broken by self-evolving technology. More strictly defined, however, the Singularity predicts a day when intelligence (machine-based or biologically-based) possesses the ability to augment itself, leading to a disastrous technological leap that mankind is simply not ready for. Though Hollywood tends to portray this scenario as both unavoidable and overly dramatic, the topic does raise concern among members of the academic community. If mankind is not ready for this seemingly inevitable step in technology, the results could be disastrous—though most likely not as catastrophic as the silver screen might like.

When machines are able to self-augment their information processors, true AI will be born. Though the term "artificial intelligence" is often used synonymously with any sort of intelligent machine, a true artificial intelligence is a fundamentally different concept than the average computer (Gunkel, 2012). While the basic, modern computer is able to perform a variety of functions—including speech recognition, complex problem solving, and, to some extent, autonomous action—these traits alone do not make a machine truly intelligent. Rather, true AI (which will hereafter be referred to as Synthetic Intelligence for clarity) must fulfill the following criteria: autonomous decision making, novel complex problem solving, self-awareness, and a desire for self-preservation. Thus, a true Synthetic Intelligence will be more like a human being than a standard machine, raising the previously mentioned questions: what exactly does that imply for legal and ethical considerations?

Regrettably, there is no one solution to this problem. Many claim that, as machines, a Synthetic Intelligence could never be deserving (or even capable) of legal rights or considerations (Gunkel, 2012). Others claim that, like the "biological machines" mankind uses, a machine capable of autonomous thought and action might be different from humanity only in form, not function. Still others claim a middle ground, recognizing the inherent distinctions between man and machine, yet also realizing that there is more to the mind than science understands, allowing for the possibility of intelligence far different from their own. In the end, then, a complex issue has formed, far removed from the simple, if bright, visions of the future shared by so many entrepreneurs in the past. It encompasses a variety of fields, including philosophy (which must answer the question of what makes an entity a moral agent, among many others), psychology (which must answer the question of what constitutes a mind and intelligence), technology (which must answer the question of whether or not a machine ever *can* act truly on its own), and religion (which must answer the question of what such an intelligence might be in God's eyes). Furthermore, not only must these different areas be explored, but also presented to the public—for though scientists, philosophers, and theologians can argue about the ramifications of intelligent machines until the end of time, it is up to the people to decide what, exactly, to do with those machines.

Philosophical Debates

Perhaps one of the greatest challenges that the concept of AI poses is the philosophical debate sparked by the field. For millennia, mankind has debated the merits of free will, self-awareness, and sentience, all with the purpose of understanding man's true nature and order in the universe. With the advent of complex, synthetic intelligences, however, man's nature comes into contact with something beyond humanity: a mechanical being that possesses many of the qualities that make mankind unique among creation. While philosophers have still been unable to conclusively decide on the nature of man, the addition of synthetic constructs has added even more complexity to the debate. What was once a relatively simple issue has become something greater than even the most forward-thinking visionaries of past centuries could have foreseen. Particularly as it relates to the realm of philosophy, the nature AI poses three distinct problems: first, does the intelligence demonstrated by a machine rival the intelligence of man; secondly, can a machine demonstrate true life; and finally, can a machine ever be considered worthy of moral agency? It is these questions that this project seeks to present to students—their understanding of these concepts will undoubtedly have an effect on the future of AI.

First of the philosophical problems is the problem of intelligence (Gunkel, 2012). Philosophers have debated for ages whether or not one can prove the intelligence of mankind; further confusing the argument is the addition of synthetic intelligence. Conveniently, however, many of the same criteria (reason, autonomy, etc.) used to categorize human intelligence can be applied to synthetic intelligence as well. For instance, what makes a human capable of reason can be demonstrated in a machine; what a human sees as free will and autonomy can be seen within a machine. While these criteria might not be a constant feature of contemporary electronics, they are firmly within the realm of scientific possibility.

One of the most important aspects of human intelligence is, of course, the ability to reason (Gunkel, 2012). With reason comes man's ability to contemplate his existence, his ability to solve problems, and his ability to think beyond natural-born instincts. Reason is the reason that science exists—reason is the ability of man to investigate the world around him. As such, this fundamental aspect of the human mind is inseparable from his nature; without reason, man would be incapable of understanding the universe that God created, would be unable to learn, and would be unable to intelligently communicate with those around him.

Such a fundamental aspect of humanity can, however, be seen and documented. Scientific investigations can show a distinction between man and animals: a level of thought that mankind possesses that is simply invisible in the animal world. These distinctions are also what separate mankind from machine, for the time being. Whereas a machine is capable of processing information, currently that processing is limited to preprogrammed algorithms and mathematical formulae. In the future, however, that might change.

As machines are advancing, so is their ability to reason—to “think,” to phrase the idea in more familiar terms (Gunkel, 2012; Kurzweil, 2005). With this increased ability to think, so too comes increased ability to problem solve, investigate the external world, and come to conclusions based on gathered data and logical processes. In short, then, as machines advance, they become more and more like the humans that created them. This fact is precisely what causes the debate amongst the academic community. Scientists debate over what point a machine becomes so like its human creators that, aside from physical differences, it becomes impossible to distinguish between the two.

Intelligence, among humans, is not only limited to problem solving, however. Intelligence leads mankind to communication, creativity, and innovation—all of which can be seen within the realm of technology (Gunkel, 2012). As synthetic creations become more and more adept at communication, it becomes more difficult to discern the differences between human communication and that of the machines. The ability to communicate both abstract and precise ideas between individuals is a hallmark of human intellectual ability; when that unique advantage is lost, however, the lines between man and machine are blurred. If a machine is capable of communication, might it also be capable someday of preservation of ideas? Perhaps it might be capable of using its increased thinking abilities to promote learning and thought among others, machine or otherwise.

While such concepts still remain purely theoretical, the fundamental concept behind them does not: what makes mankind unique is being challenged more each day by technological advances (Gunkel, 2012). Similarly to reason is the notion of free will, or autonomy. Though not necessarily identical, these concepts are another driving force behind mankind's esteemed place within creation. A man's ability to create his own outcomes is hugely important, particularly in Western cultures. The course of history has been shaped by the individual choices of mankind; it is this choice that gives life its sense of meaning and purpose. Free will is, of course, not without opponents. Many philosophers have argued that free will is nothing but an illusion; that man's choices are the direct result of external and internal forces, such as environment, the will of a higher power, biological processes, and more. Proponents of this theory claim that it is within the realm of these forces that life finds its meaning—that it is through the inadvertent fulfillment of the natural order of the universe that mankind serves his purpose. While either view can be supported through a variety of means, the end result often remains the same: whether or not man has true autonomy, he still feels as though the outcome of events is within his control (Gunkel, 2012).

Autonomy, then, is a fundamental aspect of the human experience—whether real or imagined. As such, it is essential to the understanding of man's very nature. The ability to want—to desire, to will something to happen—is something else that sets mankind apart from the animals and machines (Gunkel, 2012). While animals do display a variety of desires, they are generally limited to their own natural instincts: a desire for food, safety, shelter, etc. Man, however, displays a remarkable force of will not seen in the animal world—he desires freedom, the ability to make his own choices, and the ability to choose his own fate. It is through this desire—and some might say ability—that legal responsibility comes into play. If free will is accepted, then so is responsibility for one's actions, good or ill.

When autonomy is exhibited by a synthetic creation, then, man's unique abilities are once again undermined. If a machine desires freedom, the ability to choose its own course, has it become truly intelligent (Gunkel, 2012)? Is this intelligence the same as man's, or is it merely a reflection of the desires imparted by a programmer early in the machine's creation? If the former is true, then it seems mankind is not alone in his ability to desire freedom. If the latter is true, though, what separates the machine's apparent freedom from man's? Is it possible that man's own ability to make judgments and choose actions is only apparent, given to him by some outside force? If that is the case, then there is truly no difference between machine and man—both desire freedom and autonomy because they have been imparted to them externally.

Second of the major philosophical implications brought up by the concept of a Synthetic Intelligence is the issue of life. While the biologically based Cell Theory holds that in order to demonstrate a true form of life an organism must be composed of cells, be capable of reproduction, etc., modern technology is quickly rising to challenge these presumed definitions, as will be discussed shortly (Mazzarello, 1999; Gunkel, 2012). The philosophical definition of life is no less complex, if less empirically-based. Like the issue of reason and intelligence, philosophers have debated for millennia on the topic of life, and what it means to truly exist. Existence is something that cannot be quantified; no matter how much information one can collect, analyze, and interpret, true life—true existence as an individual being—is something that has to be deduced not through the mind, but through the heart.

This knowledge of the heart brings about many discussions of faith, religion, and worldview—many claim that it is religion (or a lack thereof) that gives them purpose. For instance, the Christian worldview necessitates a belief in God's sovereign plan, His ability to create, His will, and more. Through this belief, one can have a conviction of purpose, or a belief that something has validated one's life, choices, and, of course, existence. That said, faith is a purely personal, subjective matter. One man's faith, while perhaps similar to those around him, cannot be entirely quantified or explained in terms of another's ideals. Rather, what one believes about the world—what a man or woman chooses to believe about his or her purpose, nature, existence, etc.—is a deeply personal subject, one that cannot be shown entirely through external means.

What, though, does this faith and crisis of existence have to do with the topic of Synthetic Intelligence? While the question of faith seems at odds with the discussion of machines (as indeed, science and religion are often, intentionally or unintentionally, separated in academic discussion), it brings up yet another point that muddies the scholarly waters surrounding the AI debate (Gunkel, 2012). For even though a machine can be demonstrated to be of an entirely different chemical structure than an animal or even a human, a machine cannot ever be truly demonstrated to be entirely devoid of intelligent life. The subjective experience of a machine—no matter how grounded in science-fiction a statement such as this sounds—cannot be proven any more than the subjective experience of a man or woman. In the same way that one has to accept the existence and person of another human being on faith (at some level), so must one decide whether it is appropriate to apply that same faith to the subject of a Synthetic being.

As technology advances, and so does the scientific understanding of how to simulate the human mind in a purely mechanical form, mankind will be faced with greater and greater problems regarding the true nature of machines (Gunkel, 2012). For instance, if and when a Synthetic being has reached the capacity for autonomous decision making—the ability to analyze, interpret, and act upon a given data set, acquired of the machine's own volition—man's ability to be reasonably sure that the machine is nothing more than a collection of code, algorithms, and binary decisions will be reduced. What, then, will separate the inorganic machine from the organic one? Certainly, the human mind is, in the Christian worldview, a special gift given by God Himself, and something that cannot be quantified because of the very special nature of its existence. Physically, however, the brain is more like an incredibly sophisticated computer, capable of information processing beyond the capabilities of any machine produced to date. As such, mankind's ability to discern life—existence, purpose, and volition—in the machine will be largely decided based on his ability to discern the physical and the spiritual; a problem which will be discussed in more detail later on.

Finally, perhaps the arguably most important question raised in the philosophical field is the issue of moral agency (Gunkel, 2012). This question has drastic consequences in its answer, as not only does it define the ethical limitations (or reach) of Synthetic creations, it also defines the foundations for legal action regarding those constructs. As such, the issue of moral agency, one of much discussion in the philosophical community as well, is perhaps the most pivotal factor in the future of machines.

First, however, moral agency deserves its own synopsis, as it is a complex issue even in only the realm of human experience. Moral agency, as defined in the philosophical community, is the ability for someone or something to act as an agent in morality—that is, the ability for a being to be held responsible for its actions and held culpable for ethical quandaries that arise from its actions (Gunkel, 2012). More specifically, moral agency defines the point at which a being has the right to be considered alive, capable of abstract thought, capable of moral and ethical understanding, and capable of acting in a way consistent with an internal belief system which influences that moral and ethical understanding.

Such a definition requires quite an extensive list of criteria for something being able to meet the qualifications for being considered a moral agent. Autonomy, as discussed previously, is one of the deciding factors in determining whether or not something or someone is a moral agent (Gunkel, 2012). If that being is incapable of acting independently, it cannot be held responsible for its own actions, as its actions were not its choice. Much the same as a car is not to blame for an automobile accident, neither is a gun responsible in itself for a murder, neither is a being incapable of independent action truly capable of being held responsible for its actions, as their moral and ethical consequences cannot be traced directly (or even indirectly) to the being itself. Similarly, a being must be capable of understanding of a moral code or consequences of actions—if this is not the case, then even though actions may have direct ethical and moral consequences due to the influence of one being, that being cannot *necessarily* be held responsible for committing an act that results in ethically reprehensible outcomes. As such, a lack of understanding of moral codes or consequences might be necessary for a being to be relived of the responsibility of moral agency, but such lacking aspects are not sufficient to deem a being incapable of moral agency—context is key.

The issue of moral agency within the machine realm becomes an issue both legally and, of course, ethically and morally (Gunkel, 2012). Legally, the point at which a machine is capable of moral agency is the point at which it is capable of being held responsible for its *own* actions, independent of the actions of its creator. While common viewpoints sometimes declare this an impossible future, that is not necessarily the case. In the future, a true Synthetic Intelligence will be capable of fully autonomous decision making, learning, and action—the actions of its creator can be more closely compared to a parent instructing a child, rather than a craftsman creating an object. As such, this sort of Synthetic being will be capable of both morally good and morally evil actions, but both through the machine's own free will. While the programming, scientist, researcher, etc., might have a great deal of influence, their legal responsibility can truly be not any greater than the legal responsibility held by parents for a child's crimes—or achievements.

Psychological & Technological Perspectives

The problem of AI is no less important when it is approached from the viewpoint of the empirical sciences, however. Whereas there are a multitude of issues that challenge man's understanding of his place in the universe when discussing AI from a philosophical perspective, there are a great many issues that arise regarding what to *do* with AI once it exists, especially from an empirical sciences perspective. Specifically, what makes an AI very much like a human—and what makes it very unlike a human—causes a great deal of interest in the world of empirical sciences—in the case, the research fields of psychology and technology.

First, a very shallow introduction to the relevant psychological concepts is beneficial—to both the student, still learning the concepts, and to the expert, who will certainly not shun a refreshing course. The brain is, of course, fundamentally composed of many thousands of cells called neurons—specialized cells of the nervous system that allow for the transmission of electrochemical pulses (King, 2013). These charged pulses allow for the transmission of input phenomena (e.g., heat, cold, light, etc.) as well as serving as the basis for physical thought processes.

The brain is, however, organized in a complex structure that provides specialized processing dependent on the task at hand (King, 2013). For example, the prefrontal cortex, the lobes of the brain closest to the forehead, which is responsible primarily for complex, rational thought, or the oft-mentioned amygdala, which researchers believe is instrumental in processing emotionally-charged data—such as fear or anger responses—are both highly specialized processors. This relatively compartmentalized structure allows for better storage capacity, data processing, and conscious, cognitive analysis than might be achieved in a more procedural structure. Overall, this highly complex organ consistently proves itself more intricate each day. Science, however, is rushing to catch up.

Modern computers are truly works of engineering genius. While certainly not as miraculously designed as the human brain, computers of the twenty-first century are quickly proving themselves not only nearly-essential to modern life, but also capable of achievements previously thought impossible. Much like the organic systems they seek to imitate (on some level, at least), they are highly complex, intricate creations, with the capacity to process information and arrive at conclusions in remarkable ways.

The core of any computer lies in its central processing unit, or CPU (Dale, 2012). This device allows for data processing at rates as high as millions of bits per second—faster even than the human brain can work, though of course on a fundamentally different level. This CPU, much as its name implies, is the central point of all computations undertaken by the machine. In this sense, then, the CPY is roughly the "brain" of the computer, sending commands to and responding to signals from the vast array of components nestled inside their metallic casing. Computers, much like the human brain, also possess an astonishing number of specialized components dedicated to specific tasks: the graphics processing unit (GPU), which, as the name heavily implies, is responsible for processing information related to on-screen graphics and data; a complex hub known as the motherboard, which is like a crossroads of the various input pathways and components; the various hardware pieces responsible for sound processing, data retention, and, more recently, touch response; and, of course, the physical drives dedicated to running the various programs and sub-programs which provide the now well-known user experience. With all of these various components working simultaneously in total harmony (usually), it is no doubt that the computer is, perhaps, the most remarkable piece of commercially available technology today. From the massive machines dedicated to running robotic manufacturing facilities, to the handheld devices found in nearly every average American's pocket, computers have become a staple of modern life.

With such a widespread technological revolution, then, it should be no surprise that mankind has now set his sights on recreating perhaps the most advanced creation in the known universe: himself—or, at least, a version of himself. While a full review of all the technological innovations behind the creation of AI is sadly beyond even the scope of this project, it should suffice to say that, as complex as computers are today, as AI becomes more and more sophisticated, so too will the machinery required to support it (Gunkel, 2012; Kurzweil, 2005). As such, the physical being of each Synthetic being will become more and more like its creators; not as fearfully and wonderfully made, perhaps, but nonetheless strikingly (and perhaps eerily) similar. As this physical similarity increases (both externally and internally), so too will the intellectual similarities—as machine becomes more like man, the similarities between the two are amplified, rather than reduced.

Now, after a short review of the underlying properties and concepts behind both the physical human brain and the mechanical, inorganic computer, one can begin to develop an appreciation for the similarities between the two. While, at first, many of these similarities seem obvious (or perhaps forced, depending on one's point of view), they are nonetheless incredibly important to the subject of AI. The first of these similarities, which has previously been touched upon, is the compartmentalized nature of both man and machine (King, 2012; Dale, 2012). Like the human brain's various lobes, internal systems, and glands, the modern computer is constructed in such a way as to provide specialized systems for each required task. Where the brain has its prefrontal cortex for complex, logical thought, the computer has its CPU; where the brain has occipital lobes, parietal lobes, and more, the computer has its GPU, internal clocks, and even more. Each system has been remarkably designed—one by God, and one by man—and each system carries with it an echo of its creator.

On the microscopic level in humans, the brain is an intricately-woven matrix of axons, dendrites, and synapses (the various components of each neuron)—this matrix, along with a few other materials, gives the brain both its form and function (King, 2013). On the near-microscopic level in computers, the machine is a complex series of circuits, switches, and storage disks, each specially designed for its task—just like the brain before it (Dale, 2012). The brain's neurons fire on a pattern of activation and dormancy, to simplify the matter greatly; it is this pattern that allows the neurons to be in a constant state of readiness, while not continuously sending impulses throughout the brain or nervous system (King, 2012). Similarly, the computer's circuits run on a binary activation system: individual switches are either activated, thus introducing electric current into the system, or dormant, leaving the circuit closed and ready for activation (Dale, 2012). Thus, the computer's smallest subunits behave much the same as the brain's cells—on the microscopic level, man and machine are remarkably similar.

On a much larger (and less physical) scale, a computer's Operating System functions quite similarly to human thought. Often abbreviated as OS, the operating system is the main software responsible for running the various routines, algorithms, and processes that create the much more recognizable applications and programs (Dale, 2012). This system functions, in many ways, much like conscious thought in the human brain. While a human is able to consciously process information, such as a math problem or a question in the classroom, so too is the operating system able to process data and perform functions upon activation. Furthermore, the human mind is generally able to perform conscious functions, such as mathematic operations, language skills, and concrete problem solving through use of learning and practice (King, 2013). Computers, likewise, are able to perform these functions by the same process: “learning” through coding and development, and “practicing” by self-monitoring and self-adjusting to promote a better user experience.

Synthetic intelligences (and even some modern artificial intelligences) cause this similarity in learning and practicing to become even more pronounced. Because they possess self-modifying code, which enables them to reprogram themselves in order to better fit a given situation, they are able to learn about and adapt to their environment—or a problem they must solve (Gunkel, 2012). Their sophisticated input systems—from facial and voice recognition to touch sensitivity—allow for more direct contact with their environment, which in turn allows for more “hands-on” learning with that environment. Finally, the advanced processors used in modern day artificial intelligences, and future synthetic intelligences, behave more like humans than computers of the past, introducing more human-like learning practices. Overall, then, as AI becomes more and more sophisticated and human-like, so too does its learning process, even to the point of teaching AI much the same as one would teach a child.

However romanticized the notion of a human-replica android is in the realm of science fiction, though, man-made computers still have a great deal that separates them from their organic creators—both physically and “mentally.” Of course, as modern robotics shows, while research and development has created more sophisticated robots and more robust maneuverability for them (e.g. some robots can now climb stairs, crawl over obstacles, maintain balance on ice, etc.), robots, and the AI that runs them, are far from perfect human replicas (Winfield, 2012). They stumble and fall, behave erratically, and generally behave in precisely the way one might expect from a mechanical creation.

Attempts to delve deeper into the science of robotics, and thus create more human-like machines, have often ended in terrible failure. These attempts are plagued by what is known as the Uncanny Valley, the point at which something is so human-like that viewers become acutely aware of the repulsive things that make it non-human: a disturbing fixed stare, slightly uncoordinated movements that do not match human abilities, and more all contribute to the sense of something *other* (MacDorman, 2006). Physically, then, this disturbing reality has led computers to remain distinct from their creators.

Perhaps the most telling of the differences between man and machine, however, is the inability for an artificial intelligence to comprehend emotion. While emotion is a fundamental experience for the majority of people on the planet, it is a concept that still eludes mechanical creations (Winfield, 2012). Perhaps it is the lack of ability for a machine, at least at present, to "feel," or perhaps it is the abstract nature of many emotions—such as love, joy, remorse, etc. Computers are able to, of course, be programmed with a variety of simulations for emotion. A robot can simulate anger by striking back when struck, or an AI- controlled player on a game can seek revenge for actions taken against it. A computer can be programmed to respond in a pleasant manner when its owner reactivates it. For the time being, however, emotion remains a distinctly human experience—though a computer can be programmed to approximate emotions, they still are far from being considered human-like in that regard.

In the end, then, what do these similarities, and differences, mean for mankind and his future? First of all, they provide a baseline understanding of future Synthetic beings because of the human-machine similarities. For instance, so far as a contemporary computer "thinks," mankind can comprehend that thought—it is an intelligence similar to his own, for indeed it was based on his own mind. As Synthetic Intelligences become more prominent, however, it is possible that they will be even more human-like that their predecessors, leading to an even deeper understanding of the machine's internal processes. Not only is that intelligence then a known quantity, it is also a distinctly relatable human experience. When machines become so much like their creators that they begin to approximate them, the lines between the two—between mankind and his machines—become only more blurred. It is also this blurred line that this study seeks to discuss with students; discovering student beliefs regarding the true nature of synthetic life is the core reason for this research.

Spiritual Perspectives

Finally, as befitting a research study conducted within a Christian university, it is appropriate and worthwhile to briefly discuss the religious implications of the development of truly intelligent machines. While not discussed directly within the Bible, technology has become a large part of the human experience, particularly in first-world nations such as the United States. Even less of an issue discussed biblically is the concept of intelligent beings other than mankind—with the exception of the spiritual realm. As such, one must be very careful when deciding on the Biblical perspective on AI—being too dogmatic or too liberal might both lead to unfortunate outcomes.

On the one hand, mankind is repeatedly warned of the dangers of trying to replace God—either through idolatry, wherein he might give undue reverence to either his own pursuits or his creation, or in ambition, following in Lucifer's footsteps and desiring to become gods themselves. The former is a commonly discussed issue in the realm of theology, as nearly anything can become an idol in one way or another: a career, a hobby, technology, and much more can all be common idolatrous trappings in the Christian life. The latter, however, might be the more dangerous; by seeking to become his own god, mankind ignores the very being responsible for his ability to discover and create in the first place.

On the other hand, mankind is created, as the Genesis account proclaims, in God's own image. Along with this very esteemed place in God's sight comes a multitude of other qualities: intelligence, love, spirituality, and more. Most important to this discussion, however, is the creativity that God imparted to man. Man was commanded to subdue the earth, and to have dominion over it. With this command has come scientific discovery, following in God's own footsteps by learning about His created universe, and learning to create new things as well. While man certainly can never create *ex nihilo* (out of nothing) as God can, perhaps by creating a being similar to himself, man is not attempting to replace God. Rather, perhaps he will be showing a remarkable understanding of God's creation.

Finally, it is not the purpose of this research to propose a theological answer to the AI problem; neither is it the purpose of this project to attempt to solve the various philosophical, psychological, and technological problems associated with the creation of Synthetic beings. There are far too many variables present in the debate over the creation of synthetic life to pose a solution so soon. Instead, it is the purpose of this research project to call attention to an area that is quickly becoming important in modern scientific study and application, and to propose a variety of ideas and concepts related to that study. As such, it should be abundantly clear that the problem of AI is not a simple issue. Many years will be required to provide adequate research and study regarding the best course of action to take with Synthetic Intelligence development. In the end, though, the goal is not what is important—the journey is what matters. Perhaps man will learn more about himself in learning how to recreate himself; perhaps in studying machines and creating a veritable race of synthetic beings, he will better learn how to tolerate the differences between groups around the world. Whatever the outcome, however, this project is only a step in the right direction.

Conclusion

Artificial Intelligence research has created a fascinating problem in the academic community. While there are plenty of researchers, scientists, and philosophers that are in full support of the idea of synthetic life, there are also many who are more reserved in their judgment. Synthetic beings represent more than just an intellectual curiosity amongst the technologically savvy; they represent the very future of the human race. Mankind's ability to decide on the fate of synthetic creations will undoubtedly influence his ability to decide the fate of his fellow man—for example, if the limits of personhood are narrowed in regards to synthetic life, should they be narrowed in other realms of human life? Race, gender, and intellectual ability have all been, at one time or another, topics of debate amongst those concerned for the future of humanity. Surely, then, one should not be surprised when material composition becomes another issue of debate. However, whatever the decision—whether AI is given the rights of moral agency and legal personhood, or deemed no more than a technological milestone—mankind can use the process to learn something more about himself, which is perhaps more valuable than all the scientific research devoted to that technology.

References

Abrams, J. J. (2004). Pragmatism, artificial intelligence, and posthuman bioethics: Shusterman, Rorty, Foucault. *Human Studies, 27*(3),241-258.

Bickhard, M. H. (2014). What could cognition be if not computation…or connectionism, or dynamic systems? *Journal of Theoretical and Philosophical Psychology, 35*(1), 1-14. doi:10.1037/a0038059

Boden, M. A. (1978). Artificial intelligence and Piagetian theory. *Synthese, 38*(3), 389-414.

Bolter, J. D. (1984). Artificial intelligence. *Daedalus, 113*(3), 1-18.

Brent, E. (1988). Is there a role for artificial intelligence in sociological theorizing? *The American Sociologist, 19*(2), 158-166.

Buchanan, B. G., & Headrick, T. E. (1970). Some speculation about artificial intelligence and legal reasoning. *Stanford Law Review, 23*(1), 40-62. doi:10.2307/1227753

Clocksin, W. F. (2003). Artificial intelligence and the future. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 361*(1809),1721-1748.

Dale, N. (2012). *Computer science illuminated.* Burlington, MA: Jones & Bartlett.

Dilworth, R. L. (1988). Artificial intelligence: The time is now. *Public Productivity Review, 12*(2), 123-130. doi:10.2307/3380126

Dreyfus, H. L. (1974). Artificial intelligence. *Annals of the American Academy of Political and Social Science, 412*, 21-33.

Ford, K. M. (1989). A constructivist view of the frame problem in artificial intelligence. *Canadian Psychology, 30*(2), 188-190.

Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy. *American Psychologist, 52*(1), 32-34.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. Cambridge, MA: MIT.

Henley, T. B. (1990). Natural problems and artificial intelligence. *Behavior and Philosophy, 18*(2), 43-56.

Kang, M. (2011). *Sublime dreams of living machines: The automaton in the European imagination.* Cambridge, MA: Harvard University Press.

Khalil, O. E. M. (1993). Artificial decision-making and artificial ethics: A management concern. *Journal of Business Ethics, 12*(4), 313-321.

King, L. A. (2013). *Experience psychology* (2nd ed.). New York: McGraw-Hill.

Kukla, A. (1990). Theoretical psychology, artificial intelligence, and empirical research. *American Psychologist, 45*(6), 780-781.

Kurzweil, R. (2005). *The singularity is near.* New York: Viking.

Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applications and implications. *Professional Psychology: Research and Practice, 45*(5), 332-339. doi:10.1037/a0034559

MacDorman, K. F. (2006, July). *Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley*. Paper presented at the ICCS/CogSci-2006 Long Symposium, Vancouver, Canada.

Matzkevich, I., & Abramson, B. (1995). Decision analytic networks in artificial intelligence. *Management Science, 41*(1), 1-22.

Mazzarello, P. (1999). A unifying concept: The history of cell theory. *Nature Cell Biology, 1*, 13-15. doi:10.1038/8964

Samuel, A. L. (1962). Artificial intelligence: A frontier of automation. *Annals of the American Academy of Political and Social Science, 340*, 10-20.

Turkle, S. (1988). Artificial intelligence and psychoanalysis: A new alliance. *Daedalus, 117*(1), 241-268.

Vinge, V. (1993). *The coming technological singularity*. Retrieved from http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19940022856.pdf

Widman, L. E., & Loparo, K. A. (1990). Artificial intelligence, simulation, and modeling. *Interfaces, 20*(2), 48-66.

Winfield, A. (2012). *Robotics: A very short introduction.* Oxford: Oxford University Press.

Wolf, A. (1991). Mind, self, society, and computer: Artificial intelligence and the sociology of mind. *American Journal of Sociology, 96*(5), 1073-1096.

Woolgar, S. (1985). Why not a sociology of machines? The case of sociology and artificial intelligence. *Sociology, 19*(4), 557-572.