Us and AI: A Model of the Human Self and the Capacity of our Relationships with Computational Beings

Kekoa Wong
Computer Engineering and Philosophy
University of Notre Dame, Class of 2022

Supervised by Paul Weithman
Glynn Family Honors Professor of Philosophy
Department of Philosophy, University of Notre Dame

Completed August 2021

Abstract

AI is becoming a more powerful tool in society everyday, taking decision-making out of human hands and acting as the main facilitator in many of our day-to-day interactions. As the pace of technological innovation continues to accelerate, our dependence on artificial entities will continue to increase and they will play an integral role in our society. However, while this is an exciting time period for innovation in this field, it is also fragile with its ability to displace human relationships and our pursuit of the good. To see the consequence of this development, we must have a strong understanding of the human life: how one meets their needs, comes to understand their environment, and finds a deep sense of fulfillment. In this project, a model of the human self will be proposed that best captures these aspects of our own life, separating the human into the natural, relational, and spiritual parts. Through this model, AI's shortcomings in understanding the human will be illustrated, and it will be argued that our relationship with this new entity should be limited to a form that allows us to create meaningful connections with others and discover the good in our own lives.

Contents

1 Introduction

> *"We don't read and write poetry because it's cute. We read and write poetry because we are members of the human race and the human race is filled with passion. Medicine, law, business, engineering - these are noble pursuits and necessary to sustain life. But poetry, beauty, romance, love - these are what we stay alive for."*
>
> *-Dead Poets Society (1989)*

Our reliance on AI grows every day. In recent years, substantial progress has been made in the field, from broader forms of predictive analytics to the training of deep neural networks. Already, these computational systems have eclipsed the human mind in their ability to perform many skills, beating human world champions in complex mental challenges. With our brains' standing as the most complex computational instrument now in question, we must return to our identity of self and find where we might have similarities and differences with AI. In this way, we can discover how AI can supplement humans in our pursuit of the good and where the human presence should be valued over an artificial one.

As artificial intelligence evolves and quickly replaces human decision-makers, our relationship with the entity deepens and the consequences of broad non-human thinking becomes increasingly evident. These autonomous agents will become more and more involved in intimate human moments, playing an integral role in the development and facilitation of our lives. Teachers will be replaced by automated education, doctors by predictive diagnostics, and caretakers by robotic assistants. As this transition is pushed out, we must pause and assess the implications of this evolution. Is a teacher only defined by the material that they convey? Is a great doctor only one who delivers the most accurate diagnosis? Is a caretaker only assessed by the tasks that they can do? Is our relationship with these individuals only defined by the material goods that they can provide to us? Additionally, we must understand the livelihoods that are being replaced, and the consequences that this has for the fulfillment of future generations. A child can no longer aspire to fill one of these roles in society and human exemplars may no longer be present in these fields as they are today.

Medicine, law, business, engineering - these are all noble pursuits that have a quantifiable aspect to them, with humans intensely studying the subject so that they can know the science of the field. However, the humans who occupy these positions are much more than their informational utility and expertise. Poetry, beauty, romance, love - these draw from the intangibles of human existence that empower us to live, connect, and find flourishing. Humans inside and outside of the noble pursuits can have a deep understanding of these intangibles, creating the potential for the creation of connection, art, and fulfillment beyond the realm of science. While it may be able to learn the quantifiable, AI cannot touch the face of the intangibles in human existence. AI seems to have the potential to adequately replace and even exceed the human ability for expertise in quantifiable disciplines, but it has yet to exhibit the capacity to understand the aspects of flourishing that are essential to our lives. Thus, as it replaces human workers and decision-makers, it reduces livelihoods to mechanical functions, altering the way members of our society operate and interact with one another. Due to this transition, we will significantly increase our interaction and dependence on these computational entities, forming relationships with artificial beings where there used to be human ones.

Without AI's capability to understand the human self, what sort of expectations should be placed upon it and how should we interact with it in our lives? This is the central question that this project will seek to explore. To approach this question, it is useful to have a model of the human self that divides our existence into different parts and allows us to understand what motivates us to act in our environment and find deeper fulfillment. This model will divide the human into the natural self, relational self, and spiritual self, using literature from the fields of behavioral economics, psychology, neuroscience, and philosophy. These divisions of the self will help explain our individual behavior and how we form relationships with other humans, creating empathetic connections through our shared condition. Additionally, comparisons with the self of AI will be created, allowing arguments to be made about the extent to which we should form relationships with this entity and what we can reasonably expect from it as a being with a limited capacity to understand the human.

2 An Overview of the Human Self

The study of the human self is a topic that has been debated since ancient philosophers conjectured about the nature of the soul. These ancient writings include Homeric poems to Plato's Phaedo and The Republic. However, in these early theories, while philosophers could conjecture about these ideas, they lacked the scientific tools to examine and recreate parts of the self, which modern technology provides us today. Innovations in medical research have given us the ability to perform brain scans, measure the presence of biochemical substances, and map our genomes, de-mystifying and quantifying abstract ideas such as emotions and desires into scientific processes based in physiology. New theories in the fields of psychology and behavioral economics have been backed by studies conducted using these innovations and other resources unavailable to ancient philosophers. On the computational side, the developments in artificial intelligence have allowed us to recreate decision-making processes, providing insight into our own learning behavior. This technological progress has enlightened many aspects of our earthly human condition into quantifiable and replicable models, laying the groundwork for future models to be built on top of the currently existing ones.

This model will draw from these modern developments in psychology, behavioral economics, neuroscience, and artificial intelligence to provide a framework that describes the condition of human existence in this world, falling back on philosophical literature where necessary. This project does not conjecture about the existence of the soul in an afterlife or if the soul can exist separate from the body, as many ancient theories of the soul do. Instead, it seeks to model the reality of our human condition in this world, how we satisfy our needs, how we gain knowledge, and how we seek fulfillment, comparing our experience with that of AI.

3 The Natural Self: Basic Needs, Drives, and Vulnerabilities

The human's natural self is made up of our physiological drives, responses, needs and vulnerabilities, both physical and psychosocial. This natural self is instinctual and similar to that of animals. It is reactionary and not thoughtful, seeking to satisfy the instantaneous itches and protect susceptible weaknesses. Our physiological drives include many of the appetites in Plato's tripartite soul and Freud's id, such as hunger, thirst and lust (Kahn, 1987). These desires are built into our biology and we find a trivial sense of pleasure when these desires are satisfied to a certain extent. Like the appetite and the id, the natural self can seek to satisfy these desires to a 'criminal' extent (Kahn, 1987) but is regulated by other parts of the human self, which will be introduced later. In addition to these primitive appetites, the natural self contains broader psychosocial needs such as belonging and safety to protect from the vulnerabilities of loneliness and pain or death. Another vulnerability of the natural self is the human's limited ability for effort. According to research in the field of behavioral economics, "laziness is built deep into our nature" and we tend to "gravitate to the least demanding course of action" to achieve a specific goal (Kahneman, 2013, page 35). In short, humans most often choose the path of least resistance, opting to conserve their energy due to the high physiological cost of effort.

The natural self is also bound to the constraints of a singular existence, time, and the threat of impending aging and death. With a singular existence, a human does not have the ability to reference a pre-birth experience and cannot know whether they will get a second chance at life or be offered an afterlife. Additionally, they cannot exist in multiple bodies at one time, they only have their own. The natural selves of humans are bound to exist in the present moment. While their mind can replay a moment in the past, or conjecture about the future, their body remains rooted at the current instant. With impending aging and death, the natural self of a human is held to limited time. The natural self cannot remain youthful and strong forever, it will gradually decay in accordance with the laws of biology and must decide what to do with its youthful strength while it still exists. Additionally, like all living matter, the human faces an impending end to its earthly life. No human will know for sure the time or the place and, with their singular existence, will be entering an unknown frontier for the first time.

In the context of this project, basic emotions are hardwired mechanisms in our physiology that are identified with our brain and associated with a feeling, which is suggested by recent neuroscience research (LeDoux, 2012). The physiological response of these emotions can include neurochemical changes in our brain, beating of the heart, dilation of the eyes, or activity in the brain. All these biological factors are measurable, giving scientists the ability to build learning algorithms that can help identify the emotion a person is feeling based on their biological response (Murugappan, 2011). Therefore, neuroscientists argue that "there must be unique physiological patterns for each emotion, and these (central nervous system) patterns should be specific to these emotions not found in other mental activity" (Dalgleish and Power, 2000, page 50). It follows from this definition that emotions are innate to the natural self. While Plato associates some emotions, such as anger, with the spirit, he also places one's awareness of their social position and merits in this category, ingraining a sense of one's place in their community with emotions (Lorenz, 2019). The natural self does not include this broader social understanding. Both the need for social belonging and the physiological response of emotions are present in the natural self, but the understanding of one's community and the way in which social belonging is satisfied is left to the relational self, which will be covered in the next section.

Overall, the natural self can also be described as our "animalian" self. It encapsulates many of our instinctive tendencies, physiological responses, and biological laws that govern our primal state. This part of the self does not act effortfully or with willpower, it simply exists to satisfy its needs and protect its vulnerabilities. Additionally, it is inherently lazy, seeking to conserve energy and avoid exerting the body's physiology. The natural self also exhibits strong emotions, characterizing feelings through bodily responses. Finally, like all living organisms, it is bound by aging, time and death.

4 The Relational Self: Environmental Understanding and Knowledge

Beyond our natural selves, we have a relational self that is formed through the interactions with our surrounding physical or social environment. The relational self includes our understanding of our environment along with our knowledge base. Today's AI algorithms can be used as an example of the relational self in a vacuum, without any input from the human's natural or spiritual self. The 'knowledge' of AI is made up of the accumulated data points it has collected, factoring into an overall decision-making process. Similarly, the human's relational self is made up of an understanding of individual environmental objects along with a general form of knowledge. To use a popular culture reference as an example, there is a popular viral video in which a young child mislabels a lot of geese as chickens, illustrating an underdeveloped relational self (YouTube, b). The child may have thought chickens were all birds with wings that have two legs and are on the ground, causing an incorrect classification. With more exposure to both geese and chickens, the child will grow their relation self and be able to classify this difference better. One may have a unique understanding of a single chicken, that may be their pet, in addition to their general knowledge. They may recognize the individual characteristics of this chicken: the ones that help make up the general knowledge of chickens and the ones that make this chicken unique. But the distinction between the individual data object and the general knowledge form is an important one, as it will be used later to explain the human's abilities to form relationships.

There is a large interplay between the human's natural self and the relational self, which can be seen through a variety of examples. When we feel hunger we know that we have an immediate need to eat, but the decision regarding what to eat is formed primarily by our relational selves. In our state of hunger, our relational self knows the substances that we have been taught will satisfy our hunger, and realizes that lettuce can be eaten by humans while grass should be left to other animals. The decision to choose what to eat is determined by how one has learned from their environment, either being taught by another human or through experimentation. Furthermore, preferences are determined from this relational self. An individual in one culture may find a certain dish particularly appetizing when they are hungry, as they have had a fond experience of eating that dish when they were younger while an individual from another culture may be repulsed by the dish due to their uncertainty or some characteristic of the dish that was stigmatized in their culture.

Using sounds as another example, we may hear a car approaching us, causing us to jump out of the road. The identification of such a noise, associating it with a car sound, is credited to our relational self. But the feeling of danger and fear stems from our natural self and the basic need for safety and aversion to pain and death. Thus, these two parts work together in tandem, with the relational self providing the understanding of the noise and the natural self creating the need to avoid the noise.

The relational understanding of AI can be used to further distinguish between the human's natural and relational self. In the previous example of the car, AI has the ability to identify the noise, collecting data through sensors and comparing the sound waves to previously collected and recognized noises. Through this process, it can correctly identify the sound of a car, much like a human can. However, AI does not have the input of the natural self to truly understand the meaning of an approaching car.

It does not feel the same sense of threat or danger, as it is not a living biological organism that is constrained by pain or death. Cognitive scientists have argued that this difference between humans and AI is the distinction between sensing and perception (Berberich et al., 2020). AI has the ability to sense and identify, but lacks the ability to perceive the life context and deeper meaning behind the stimuli.

Beyond this, AI is also not held to many of the physiological constraints of the human's natural self that impact the development of their abstract relational self. Humans can only store so much memory inside of their brains, have differing mental capabilities, and have a limited supply of time and energy to collect data and form decisions. Thus, a human's relational self must heavily rely on processes such as heuristics instead of computationally heavy algorithms to form their sense of "knowledge," while also needing the ability to heavily filter through their data points, deciding which ones are most important to remember. AI relies on algorithms to formulate their decision-making and, while they may have engineers enforcing certain time constraints, they do not have to fight against an innate laziness present in their physiology or make use of emotions.

4.1 System 1 Thinking: The Development of the Relational Self Constrained by the Natural Self

Much of the field of behavioral economics is dedicated to studying how humans form their knowledge and decisions, especially through the usage of heuristics. Daniel Kahneman and Amos Traversky divide the human decision-making processes into a dual mode model with two "systems". System 1 operating "automatically and quickly, with little or no effort and no sense of voluntary control" forming "freewheeling impulses and associations." In contrast, system 2 is the "conscious, reasoning self that has beliefs, makes choices, and decides what to think about and what to do" (Kahneman, 2013, page 21). System 1 derives its impressions and feelings effortlessly from the "explicit beliefs and deliberate choices of System 2" (Kahneman, 2013, page 21). It is much easier for humans to spend their existence in the state of system 1 since system 2 requires much active effort and attention. System 1 is deeply rooted in the weaknesses and emotions of our physiology and natural self, but forms our relational understanding and knowledge base. Thus, it further illustrates the interplay between the natural and relational. System 2 also forms our knowledge base but will be covered more later as it requires active effort to operate, which is not a basic natural tendency for humans.

4.2 A Comparison Between System 1 Thinking and AI

Using the example of map directions, the comparison between AI's strict relational self and the human's more complex relational development (with the interplay of the natural self) can be further illustrated. As discussed previously, AI does not have the ability for emotion and it also does not have the constraints of limited memory and energy to the extent that the human has. In this way, it is almost a purely relational being. Initially, both AI and the human are similar in that they derive their structure entirely on the data collected. For example, when asked for directions from point A to point B, both the human and the AI will reference the external data that they have on the subject and the knowledge base that they have built. The human will draw from their knowledge on the roads, their individual routes and the traffic they experienced whereas the AI may draw from its stored roadmap and the traffic data simultaneously collected from hundreds of smartphones. However, due to their physiological constraints in their natural selves, humans have a limited capacity for data storage, cannot have simultaneous parallel experiences, and also exhibit limited energy. They can only remember their own experiences (or what they have been told), and do not have the ability to remember every detail about the route or the exact amount of time it took.

Focusing on the tactics of system 1, humans will most often employ the quick heuristics, not remembering the specifics but feeling a general idea. Using these heuristics, a human may say that the best route is the one that offered nice scenery ("The view was beautiful!"), was the most simplistic and did not confuse them ("Just take a left and continue going straight and you will get there eventually. . . "), or was the safest ("Go right, you will miss this stoplight that people always drive through."). In all these examples, the influence of the natural self can be seen in the formation of the heuristic, with emotions, the constraints of laziness, or personal safety all respectively playing a part. In contrast, AI will always return the mathematically maximized option, providing a route that is the quickest but may be ugly, incredibly complex, or unsafe since it is not bound to the input of a natural self.

One may have such an experience, remembering a route that left them wondering, "Why did the GPS take us that way instead of this way?" In a similar manner, GO experts were left perplexed by the tendency of AlphaGO, an AI program developed by Google's Deepmind, to make "slack" or lazy moves when there seemed to be a much better one available. GO is a highly complex board game, more so than chess, that few humans have the capacity to master, making it an ideal challenge for AI to tackle. In 2016, AlphaGO defeated 18-time world champion Lee Sedol in a five game series of GO, marking another pivotal defeat for human expertise similar to Deep Blue's 1997 victory over world chess champion Gary Kasparov. In these matches, experts were puzzled over AlphaGO's tendency to make these "slack" moves when there was another move that seemed to give it a much stronger advantage (Ciolino et al., 2020). In chess terms, this could be compared to having the ability to take the opponent's queen, but deciding to take their bishop instead. With these decisions, AlphaGO does not care about having a stronger advantage, it simply cares about maximizing the probability of winning. It is content with winning by the smallest of margins, which would cause much discomfort in a human, and will perform a "slack" move as long as it increases the probability of winning.

With the input of the natural self, humans will tend to gravitate toward safety, racking up as much of an advantage as they can. Due to our recognition of our ability to commit errors and limited capacity to calculate moves in the future, a human player would seek out the biggest advantage in the foreseeable future, putting themselves in a position to find victory when they have the capacity to foresee it. AI revolutionizes the game in that it is not bound to these constraints and always has victory within sight.

4.3 Data Biases

It is important to note that the development of the relational self is entirely contextual, meaning that it is fully reliant on its experience to form a knowledge base. Humans form knowledge and an understanding of the world that is highly dependent on where they were raised and formed their initial heuristics. To learn something, one must be exposed to it, and often repeatedly. While this may seem obvious, it becomes problematic when one may be overexposed or underexposed to ideas, skewing their knowledge base. This type of distortion will be referred to as data biases.

There are many ethical concerns over such biases, such as "thinking or treating another person differently based on the perceived characteristics of the individual" (Howard and Borenstein, 2018). Due to entrenched biases that permeate society, young humans quickly form representative heuristics on intrinsic characteristics based on their extrinsic environment. Girls can be shaped by societal stereotypes as young as six, forming the belief that they are not as smart as boys due to their constant exposure to direct and indirect messaging, such as only hearing about Albert Einstein and Thomas Edison when introduced to learning materials covering famous scientist idols. The male scientist bias can inadvertently take root and grow "with the child's exposure to non-female scientists - on television, in books, and in movies" (Howard and Borenstein, 2018). This belief from a young age will impact a girl's future development and discourage her from pursuing prestigious careers typically associated with mental brilliance (Bian et al., 2017). Even though the natural self of a certain girl may have an incredible capacity for scientific brilliance, her relational self could develop in a way that teaches her to pursue a different career in order to fulfill the natural needs for social belonging and acceptance.

While AI is not susceptible to emotional biases or heuristics impacted by the natural self and system 1, it is equally prone to data biases that all humans experience. Since AI is a data-based tool, an entirely objective relational self, it relies upon its environment in order to provide information for its decision making process. However, if the environment itself is tainted with bias, then the decision making process that AI forms will be inherently biased as well. While AI will not form beliefs about its own natural abilities since it lacks this part of a self (unlike the case of the young girl), it will have a biased assessment of the surroundings similar to any human. Thus, in crowdsourced data, AI finds "patterns within datasets that reflect our own implicit biases and, in so doing, emphasiz[es] and reinforc[es] these biases as global truth" (Howard and Borenstein, 2018).

Not only is AI susceptible to forming these biases in itself, but it also plays a critical role in furthering or mitigating these biases throughout society. This is illustrated through certain inadequate applications of AI such as predictive policing, which relies "on data from past criminal activities to guide future practices" (Howard and Borenstein, 2018), deepening past bias into a dangerous cycle that would target specific groups or areas. If police officers only patrolled and made arrests in a certain neighborhood in the past and then fed this data into an AI application, the AI would continue to recommend that officers should patrol this area since it is based on biased data. As a result, more arrests would be made in this same area due to the AI recommendations, the data would be fed back into the algorithm, and the AI would recommend even more patrols in this area, amplifying the bias on itself. Similarly, this type of biased data input has led to Google's AI displaying far fewer ads for high-paying executive jobs to women than men and showing image search results for doctors and nurses that are inline with gender stereotypes (Howard and Borenstein, 2018). Therefore, we cannot view AI as an objective tool of truth and must recognize it as the completely relational being that it is, entirely removed from the humanity of the natural self and solely reliant on the context of its data-filled environment.

4.4 Connection in Natural Relationships

The human's ability to create connections is built through their ability to share experiences with their environment. With this shared experience, humans understand the shared reciprocity in the feelings of the natural self, binding them to the other. Oftentimes, the initial formation of such relationships are cultivated through the discovery of common passions, such as sports or art. Through this discovery, individuals recognize the shared feelings that they have with the other through their common interest. Thus, they see the experience of the other through the eyes of their own natural self, feeling as they feel. This links the natural self with their relational understanding of another, creating a union between these parts of the self and a bond between individuals. This bond can deepen and grow as individuals experience more of life together, simultaneously sharing the feelings evoked in their respective natural selves. Repeating old stories together is a common connecting moment between friends, as it arouses the mutually felt feelings of their natural selves that were experienced previously. Inside jokes are specific types of old stories that are particularly effective, as they highlight the uniquely shared bond between individuals while leaving outsiders in the dark.

Aristotle's friendships of pleasure are a subset of these relationships, as they meet the needs of the natural self in the respective individuals. They can do this in a variety of ways, whether it be for comfort, lust, or belonging. Reciprocity must be central to the connection, meaning that both parties involved must feel a sense of pleasurable satisfaction to maintain the relationship. Beyond just friendships of pleasure, this natural sense of love and connection means that one can feel the pain and feeling of another. Communities can be at their strongest when tragedy strikes them, as individuals recognize the shared suffering in one another, creating a strong empathetic bond. In these moments, similar sentiments of pain are felt in the natural self, pushing individuals to see the humanness in others around them. Thus, these shared feelings cause individuals to treat one another with more dignity and compassion, as they realize their shared human condition.

The ability for humans to form feelings of natural connection for others pivots on their ability to feel in their natural self what another is feeling. This is not a deep form of love, as it draws primarily from shared experiences, and would not include Aristotle's friendships of virtue. A human would only require the capacity to understand needs, vulnerabilities, emotions, and the constraints of human existence. While there are exceptions, most humans are capable of forming these types of relationships as they only require the ability to share experiences and feelings with another without needing a deeper sense of virtue.

This form of natural connection can also motivate individuals to act in the interest of others around them. As a relationship grows, the empathetic bond between individuals strengthens, and they naturally begin to feel how the other is feeling. They know when the other person is feeling fear, pain, or joy because they have become conditioned to feeling it within themselves with the other. In this way, love and connection act as an entanglement of two individuals. Thus, when one is naturally empathetic to another, they can rationally act in the protection of the other person since they feel pain when the other feels pain. While such an act may be labeled as selfless, it is ultimately self-interested, as the individual is driven to do such an act due to their empathetic attachment with the natural self of the other. A parent may act to rush into a road to save their child from being hit by a car because they do not want to empathetically feel a tremendous amount of pain from their child's injuries, even though they would be physically unharmed if they left the child in the road.

4.5 AI and Natural Relationships

As a being without a natural self, AI does not have the capacity to form natural connection. It does not have needs that it has to satisfy, emotions that arise in intimate moments, or vulnerabilities that it must protect. Thus, in any role that it fills in society, AI does not have the ability to reciprocate anything in a relationship. It can only seek to learn and increase some sort of utility function. In the case of the therapy robot care seal Paro, elderly patients may pet the seal, eliciting a purring noise from the robot (Wada et al., 2010). While the patients may gain some sort of need satisfaction, thinking that they are pleasing another creature in a positive way, the robot does not perceive any pleasure. As covered previously, the robot can only sense the feeling and return the noise and does not have a deeper understanding of the action being facilitated. Thus, without a natural self, it does not feel anything. Instead, the robot can alter its behavior based solely on its learning algorithm, perhaps changing the noise it emits based on the characteristics of the individual and attempting to maximize the measurable pleasure of the human. Much like a human faking intimacy, Paro can only do its best to mimic pleasure from a relationship for a certain extrinsic goal. Similarly, a caretaker may need to help move, bathe, and physically assist a patient, but they may also need to hold someone's hand to provide comfort during their dying moments. AI robots could assist physically and may be able to mimic actions of comfort in these moments, but it cannot truly be a companion, having the empathy for the pain, suffering and death that are essential parts of the human's natural self.

We must realize AI's inability for reciprocity as it dynamically alters the way in which we operate when we rely on it. Take the simple example of an automated grocery store checkout station. While this type of technology would not be considered AI (unless perhaps it had some form of natural language processing), it is similar to other forms of AI since it replaces the presence of the human worker. Some individuals will insist on always using the checkout lines with the human grocery clerk, as they value the reciprocity in the human relationship, conversing with the individual and relating to them in an empathetic manner. Others may be in a rush or do not want to exert this effort to converse, opting to use the automated machines instead. There are many valid reasons to use the automated machine, time and ease being only two. However, it is important to note that in this decision, we reduce the role of the clerk to the functional utility of checking out our groceries. This line of thinking is extendable to doctors, teachers, and caretakers, as brought up in the intro. The automation of these roles would require us to constrict our view of them to the material service that they provide to society, ignoring the importance of the relationships and character traits of the individuals themselves. We would only see them only in terms of the quantifiable satisfaction that they would provide to us: if they correctly correlated our symptoms with a diagnosis, if they taught us the most accurate subject matter, or if they met some of our deficient needs.

When further applied to more intimate relationships and roles facilitated by AI, the consequences of such a development become even more worrisome. The potential for romantic relationships between AI and humans has become a common topic fictionalized in literature, such as the movie Her. But with AI's inability to reciprocate and understand the human, their role in such a relationship would be diminished to a dimension of functional utility, similar to the grocery clerk or caretaker. The AI would be engineered to perform a utility with a mathematical precision, attempting to learn and maximize a measurable return in their relationship with the human. The human in this relationship, or the engineer of the product, will be aware of this intention of the AI. Such a romantic relationship may meet the superficial and measurable needs of the human, but it would be fully inept at providing the sort of empathy required for natural connection, as AI cannot relate to the natural existence of a person. Additionally, the sense of personal investment would not be present, as AI does not have to sacrifice like humans do. In the movie Her, the protagonist is shocked and hurt when he finds out that the AI is talking to thousands of others in the same romantic way. Due to their constraints of the natural self, humans only have the capacity to invest the time and effort required for a romantic relationship to relatively few individuals, providing a sense of sanctity to the relationship.

While our development of AI is not yet at the capacity for these types of relationships, it already facilitates the matching of individuals through romantic or dating apps. This algorithmic matching may be effective at finding similarities in interests, ideologies, or backgrounds, providing a foundation to build a natural connection to another. However, it is only as effective as the mindset of the individual using the app. The similarity to an automated grocery checkout station, emphasizing quickness and ease, facilitates a user mindset that expects a material good. With this mindset, users are much less likely to build a meaningful connection, expecting a complete one to be given to them by the AI system. They see natural relationships as pre-built entities that can be picked up and checked out, like any material good, when in fact these types of relationships are most easily built through the investments of shared experiences together. Therefore, these algorithms can be seen as an indirect way in which AI alters our expectations in relationships and our interactions.

5 The Spiritual Self: Willpower and Meaning

The third and final part of the human self is the spiritual self, which encapsulates values, ideals, and how we find a sense of fulfillment in our lives. The virtues of human existence are developed by this part of the self, and the pursuit of the "good life" is found in the effort it employs. This project focuses on the development of the individual toward this goal and it will employ the use of virtue ethics to approach the subject of human flourishing. While many virtue ethic traditions agree on a similar set of virtuous traits, the spiritual self provides a sense of willpower and meaning that are necessary in the cultivation of these character attributes.

5.1 Spiritual Willpower

In the context of this project, willpower is defined as the ability of the human to act in a manner that requires effort and control over their natural tendencies. For example, working out can be seen as an exercise of willpower, as it requires the human to overcome their natural laziness and feelings of pain, tearing muscles to increase their strength. This spiritual willpower pushes individuals to overcome their vulnerabilities and needs present in their natural self, leading to a renewed sense of fulfillment and growth in the human person. As individuals exercise this willpower, they begin to gain the ability for more complex activities and experience pleasure in their natural selves for this achievement, like the Aristotelian Principle defined by John Rawls. Rawls writes that this principle of motivation implies "that as a person's capacities increase over time (brought about by physiological and biological maturation, for example, the development of the nervous system in a young child), and as he trains these capacities and learns how to exercise them, he will in due course come to prefer the more complex activities that he can now engage in which call upon his newly realized abilities" (Rawls, 1999, page 375). These new abilities are achieved through the willpower of the individual to exert the effort over the inherent laziness of their natural self, pushing themselves to become better in a certain task. While the individual may feel a sense of pleasure in their newly realized ability and prefer it to simpler activities, the deeper sense of fulfillment is found in the exercise of their spiritual self to overcome their natural tendencies and grow as a human.

A young child may feel a sense of pleasure from receiving an "A" in class, but the sense of fulfillment is dependent on the effort that they may have had to employ to achieve the letter grade. A child who received an "A," but had to overcome many challenges to achieve that mark, perhaps failing an exam or not understanding the material at the beginning of the semester, would receive more fulfillment than their naturally talented peers. The pleasure obtained from social status or current ability is independent of the deeper feeling of achievement obtained through the exercise of one's willpower. This willpower exerts the effort necessary in the thinking of system 2, which was introduced by Kahneman earlier. The effortful process builds our relational understanding, and a struggling student may have to spend much more time in this state than those who simply "get" things. The new capacity that the individual trained for has now become instinctual and is thus part of their natural instincts in the form of system 1 thinking, which is supported by Kahneman and Traversky's theory of dual system thinking. The individual may settle here, finding pleasure in their new capacities, or they could seek for more fulfillment through the practice of their willpower, pushing themselves to continue to grow despite the laziness present in their natural self. The importance of willpower can be seen here, as it is more spiritually commendable for an individual to exhibit the strength to push through struggle than to be successful without having to exert any effort.

Since Kahneman and Rawls both use the example of a chess player when explaining the development of more complex abilities, it will be used here as well. At first, the game of chess may seem daunting and incredibly complex to a beginner. However, as the beginner exerts the effort to study the game, practice, and learn strategy using system 2 thinking, they start to gain a deeper understanding of the game and more complex abilities. Thus, they begin to find pleasure in their complex realization of their skills, playing chess as a pleasurable pastime activity and preferring it to checkers, in accordance with the Aristotelian Principle. An advanced chess player can thus play at a high level instinctively, playing games at a 'Blitz' pace, forming complex moves and strategies in seconds. At this point in their development, the game of chess has become part of their natural abilities, and the individual no longer must exert much effort to play at a relatively high level. Instead, they are employing system 1 thinking and reacting with quick heuristics that have been formed.

As Rawls notes, humans are limited by resources such as "time and energy" and "there are only so many hours in a day, and this prevents our ascending to the upper limits of our capacity all the chains that are open to us" (Rawls, 1999, page 378). Due to the constraints of the natural self, humans will have to decide which abilities they most want to develop and where they should spend their effort. They cannot become masters of all trades. Additionally, it is possible that the increasing satisfaction from a realized ability may become marginal due to the effort that has to be exerted to achieve this minimal gain in ability. In other words, a chess beginner may have to study 100 hours in order to increase their ability by 10, but may have to study 500 hours in order to increase their ability by 1 after this initial increase. Rawls argues that there is an equilibrium point in which "the gains from a further increase in this level are just offset by the burdens of the further practice and study necessary to bring it about and to maintain it" (Rawls, 1999, page 376). Thus, there is a time in which it becomes increasingly fruitless for one to pursue the further development of a singular skill and they can find it more fulfillment to pursue something else. A relatively advanced chess player may be a beginner as a writer and find fulfillment in developing their skills in this field, similar to how they did initially with the game of chess. Therefore, it may be more fruitful for the fulfillment and development of the individual to branch out and exert their willpower to grow in different fields.

It is also possible for one's newfound complex abilities in one domain hinder their future fulfillment, as they become complacent in their ability to exert effort and willpower. For example, a chess player could become complacent in their skill development, settling to gain the immediate pleasure of their existing abilities that have been built into their natural self. While they may have exerted effort to attain these skills in the past, they now give in to their natural tendency for laziness and relaxing through system 1 thinking, playing chess as a passing leisure instead of actively attempting to better their skills. The game of chess could also act as a vice, deterring the individual from branching out and learning new fields. In this case, a reasonably advanced chess player who is attempting to better their skills at writing philosophy may settle to relax and play a few hours of chess instead of exerting the effort to continue to write and research philosophical literature. In this case, the advanced skills in chess, which have become engrained in the natural self, act as a source of immediate pleasure hindering the pursuit of a deeper fulfillment through the practice of willpower.

5.2 Spiritual Meaning

Meaning refers to the object of the human desire for a deeper sense of ideals and values. This will for meaning is what is addressed in Viktor Frankl's psychological theory of logotherapy, in which individuals seek "striving and struggling for a worthwhile goal, a freely chosen task" (Frankl, 2006, page 105). Frankl's experience living in Nazi concentration camps during World War II had a large impact on the formation of his theory, as he viewed this deeper pursuit of meaning as a differentiating factor between those that lived and died in the camps. He quotes Nietzsche to support his approach to psychotherapy, saying that "he who has a why to live for can bear almost any how" (Frankl, 2006, page 84). According to Frankl, many psychiatric investigations have reached the similar conclusion that the prisoners who were most apt to survive were those that knew that there was a task waiting for them to fulfill beyond the walls of the camp. The belief that they had some greater goal to live for, no matter if this goal was rational or not, played a crucial role in the mental and physical health of the prisoners.

Discovering this sense of deeper meaning is a unique journey for every individual and it must feel like a freely chosen task to truly be effective. Frankl writes that it must result in the "self transcendence of human existence" and "always points, and is directed, to something, or someone, other than oneself" (Frankl, 2006, page 110). Thus, one must find a meaning broader than their personal satisfaction to find greater fulfillment and purpose in life, contributing to a mission that outlives themself. While many can go much of their lives without finding this deeper meaning, a crisis can result when one is forced to confront the reality of their vulnerable condition. These moments can include near death experiences, extreme loss, or suffering, pushing a human to question their purpose and why they are alive. This questioning is a necessary action for individuals to find the motivation to continue onwards through their struggles. Without discovering this deeper meaning, individuals experience an existential vacuum and are left to seek fulfillment through diminished methods included in other theories of psychotherapy, such as Freud's will to pleasure or Nietzsche/Adler's will to power.

Furthermore, Frankl argues that "the more one forgets himself – by giving himself to a cause to serve or another person to love – the more human he is and the more he actualizes himself" (Frankl, 2006, page 111). While the basic needs and vulnerabilities of one's natural self must be satisfied to a certain extent (one cannot simply live without food), the meaning beyond oneself gives an individual a reason to sacrifice what may be of satisfaction to their natural self. For example, one may decide to fast in pursuit of some religious meaning and find fulfillment in the self-sacrifice of their natural needs. To a further extent, one may sacrifice all their wealth to serve with the poor or even surrendering their life for a certain cause beyond themself. This unwavering commitment motivates some of the most sacrificial human actions witnessed, such as humanitarians embracing a life in poverty or monks self-immolating in protest.

The value of "forgetting oneself" towards self-actualization can be found in New Testament and Buddhist teachings. In the following passage, Jesus talks about the importance of this trait to his disciples: "Whoever wants to be my disciple must deny themselves and take up their cross and follow me. For whoever wants to save their life will lose it, but whoever loses their life for me will find it" (Matthew 16:24-25, NIV). While some literalists may interpret this passage as a reference to an afterlife in heaven, the psychological implications of living in the manner that Jesus preaches can lead to a form of self-actualization in this world. When one attempts to save their own life and satisfy their own needs, they can never reach the point of fulfillment in looking beyond their own satisfaction. However, as one forgets themselves and loses sight of their own life for another, such as Jesus or their faith's ideals, they find a deeper form of meaning that allows them to view life anew.

In a cowritten book, the Dalai Lama and South African Archbishop Desmond Tutu describe the paradox that "one of the fundamental secrets of joy is going beyond our own self-centeredness" and it is "foolish selfishness. . . and self-defeating to focus on our own joy and happiness" (Tenzin Gyatso et al., 2016, page 62). Thus, one cannot find fulfillment by pursuing the feeling itself. Instead, they must wholly and authentically commit themselves to the person, cause, or value system beyond themselves to stumble upon fulfillment. Therefore, actions such as the true compassion for others lead to deeper fulfillment for the compassionate individual as a byproduct of their deeds instead of the main objective.

5.3 Finding Spiritual Meaning

Meaning can be derived from several sources, including a belief in a telos of human life, faith in some religious deity, or through the practice of a cultural tradition. In virtue ethics traditions, it can be developed by the way in which one discovers how to live the good life. Most virtue ethics traditions agree on the importance of a relational understanding of meaning, in which the identity of a virtue is "formed through a network of relationships" (Vallor, 2016, page 76). In this way, "the cultivated person always acts from within her own unique context of important relations, roles and responsibilities, while seeking perpetually more refined understandings of these relations and moral obligation and ideal to which they give rise" (Vallor, 2016, page 77). Thus, meaning is derived contextually, drawing from societal values and the placement of an individual in it.

Using the example of chess once again, an individual may find different meanings behind playing the game depending on the context that they are in. For example, a master at a world chess championship may find meaning in winning the tournament, as it honors their country and family. In contrast, a parent may not finding meaning in winning a chess match against their child and instead have the goal to teach and educate them, not caring about the possibility of losing. These meanings are situationally dependent, as the same individual may find themselves in both roles and carrying different responsibilities in their life.

Therefore, in virtue ethics traditions, finding meaning requires that one has a sense of practical wisdom and responds to their context most fully (Vallor, 2016, page 77). Additionally, finding meaning is never a complete, but instead is a part of journey of cultivation toward human flourishing and is always left as an "open circle" (Vallor, 2016, pages 63-65). Thus, virtue is not measured by the status of the individual's quest, but instead is found in this continuous path of cultivation toward an unreachable goal.

While virtue ethics traditions may base a sense of meaning off practical wisdom and relational reasoning, it does not have to be based off rationality to be effective. In his experience in concentration camps, Frankl recounts how individuals would be motivated to stay alive due to the meaning they found in the prospect of returning to their kids (Frankl, 2006, page 79). While a prisoner may have originally formed this sense of meaning from a relational understanding similar to virtue ethics traditions, its use now transcends rationality, as the individual does not know if their children are alive and have survived their own imprisonment. In this case, a person may only have faith that their meaning can be a worthy pursuit.

Furthermore, Frankl argues that one's final and ultimate meaning always "exceeds and surpasses the finite intellectual capacities of man" (Frankl, 2006, page 118), as in the case of beliefs regarding many religions and death. It is impossible to truly know what lies beyond death, as it is a frontier beyond the natural capacities of humans to explore and understand. Many religious traditions require beliefs about this frontier, falling into the scope of faith. To find an ultimate meaning that transcends death and the existence of the human on this earth, an individual may be required to make this leap beyond rational argument.

5.4 Elaborations on Spiritual Willpower and Meaning

Willpower and meaning cooperate in the spiritual self to bring about fulfillment in one's life. One exerts their willpower over their natural self in a difficult task because of the deeper meaning that they are pursuing. A chess player have many reasons to exert the willpower necessary to learn the skills and tactics of chess. They could do so for the purpose of gaining pleasure through the attainment of higher skills in the game, in accordance with the Aristotelian principle. They could also do so to dominate the competition, exemplifying a will to power. Or they could play chess for some broader purpose beyond themselves, perhaps to be a role model to other chess players across the world. This final reasoning exhibits the simultaneous cooperation between willpower and meaning. In the first and second examples, the will is exerted, but without a meaning extended beyond the satisfaction of one's natural self. Additionally, it is also possible for one to discover a deeper meaning in their life, yet not exhibit the spiritual willpower to pursue that meaning. Thus, willpower and meaning can be independent of one another and only in specific cases can the combining power of both be seen.

It can be imagined that some individuals may argue that harder tasks always correspond to more potential for fulfillment. This is not the case. A struggle must be for some sort of purpose, whether it be for pleasure, power, or a deeper meaning to find fulfillment in it. One is only motivated to strenuously exercise because of its purpose to get them stronger and healthier, satisfying their will for meaning, pleasure, or power in some way. If this purpose was not present with exercise, then the activity would be a pointless, hard task that did not lead toward fulfillment. Some may provide a counterargument, saying that some individuals enjoy the task of exercise or chess itself without the possibility of getting better. In response, those who enjoy an activity in this way are experiencing pleasure through the realization of their acquired abilities, according to the Aristotelian principle. These acquired abilities are part of their natural self, and they do not have to exert strenuous effort to use these abilities (as they would not be defined as an acquired skill if they required a lot of effort). Thus, those that enjoy such an activity in this way would have to utilize the effortful willpower to build more complex skills to experience a higher form of pleasure. In this way, it can be seen that there is a purpose to the struggle and effort exerted by the individual.

Additionally, choosing to participate in a difficult task for more fulfillment is ultimately fruitless, as it can only be discovered as a byproduct of an autonomous action. Serving the poor is a difficult action that some may do for a variety of reasons, such as allowing themself to feel like a powerful person through their giving or for a spiritual reason. Even if a spiritual meaning is not pursued in the giving, the action requires a certain level of willpower that allows the individual to feel a sense of fulfillment through the effort they employ. However, if one decides to pursue service to feel fulfillment, they will be disappointed. As the purpose of their action is oriented toward fulfillment, it will escape their grasp until they authentically commit themselves to a purpose beyond their focus on fulfillment, which is consistent with the theory of logotherapy and the paradox described by Buddhist and Christian leaders.. Diminished methods of purpose, such as pleasure of power, can still elicit some sense of fulfillment from the independent exercise of willpower. However, willpower combined with the genuine pursuit of a spiritual meaning will ultimately cultivate the strongest feelings of fulfillment in the individual.

5.5 Spiritual Connection and Virtuous Love

Similar to the natural connection and love discussed previously, there is a form of spiritual connection and love that binds individuals when they empathize with another's sense of spiritual willpower and meaning. When one experiences this sort of connection, they understand the spiritual longing of another to the extent that it gets entangled in their own sense of spirit. These types of relationships are like Aristotle's friendships of virtue, with individuals sharing a conception of a good they are pursuing. Virtuous relationships can be rare, as they require fully virtuous persons who have worthy characters to enter such a friendship. While relationships of pleasure and utility would only require individuals to have a developed natural and relational self, eliciting a sense of reciprocity that satisfies the natural needs, virtuous friendships require that individuals have a sufficiently developed spiritual self and the continuous drive to develop it more. Additionally, as the spiritual self is not a sedentary state and is always striving toward flourishing, spiritual relationships are continuously in a state of growth, with both parties cultivating themselves and discovering each other more.

While many virtuous friends may have the same conception of spiritual meaning (being a part of the same faith or philosophical schools of thought) they do not have to have this same meaning to care for one another's spirit and better one another. The friendship between the Dalai Lama and Archbishop Desmond Tutu exemplifies such a relationship with differing senses of meanings. As respective leaders of Buddhist and Anglican thought, they have come together a few times in recent years to express their care for one another and articulate the deeper ideals and values that their religious practices share (Tenzin Gyatso et al., 2016, page 2). A recognition of their shared humanity, longing for individual meaning, and reliance on aspects of faith allows them to empathize with the spiritual self of the other deeper than their religious backgrounds, binding them to a form of love.

5.6 AI's Spiritual Behavior

What is a spiritual self without the understanding of a natural self? Can a human discover an ultimate meaning without reckoning over their death and limited time? Will one need to find a reason to continue onward if they do not know suffering? Is there ever a moment for will if there is no natural sense of laziness to overcome? As AI lacks a natural self, it also has no capacity for spiritual willpower and meaning. Therefore, it also cannot form spiritual connections with others through this part of the self.

However, AI is not alone in its inability to form spiritual love. Animals exhibit many of the same characteristics of the human natural self, having a instinctual drive to fulfill their basic needs for hunger, thirst, and pleasure while protecting their vulnerabilities. Additionally, animals are bound by the same laws of biology with limited energy and impending death. They also seem to have some capacity for natural love and connection, as they can act selflessly and for the protection of others, such as their children or others in their herd. However, do they have the capacity for a spiritual self and the type of relationships that this would entail? It would seem like only to a very limited scope. Animals may exert some sense of willpower for the purpose of pleasure or power, perhaps by fighting others for the most dominant position in a herd. However, these are diminished methods of the will and fall short of true spiritual meaning. They may also sacrifice or endanger themselves for others due to their natural empathetic connection for another, but this connection falls short of a deep spiritual love. Therefore, animals may exhibit some instances of spiritual will but lack the ability for a true pursuit of spiritual meaning beyond themselves.

This leads us to the following question. If AI was designed in a way that would enable it to develop a natural self in addition to its relational self, would it even be able to have a sense of spiritual understanding? Not necessarily. While it is impossible to know for sure whether a biological creature would have a spiritual sense, the issue for AI lies in its quantifiability. Many aspects of the natural self are capturable by science and mathematics, as we are able to tell when one may satisfy a bodily need or experience pleasure by the measurable physiological response. However, how can one possibly quantify faith in meaning, virtuous love, or the essence of willpower? These field lie outside the realm of science but make up much of the breath of our life. As a computational being, AI can only hope to pursue quantifiable metrics that are within the mathematic grasp of engineers.

There is one further stipulation. Suppose there was some future where a measure of human spiritual fulfillment was quantified with precision. How would an artificial entity hope to find this fulfillment? Once again, fulfillment is not found in its sole pursuit and it only arises as a byproduct of an action. Therefore, attempting to maximize this metric would be self-defeating and would ultimately escape the grasp of the AI. It would not be able to simultaneously pursue and not pursue a mathematical function at the same time, therefore, it will be lost in its journey for fulfillment.

6 Moral Development and AI

Moral development is an entirely complex process with only parts that will be covered in this project, but is closely tied to feelings of love and spiritual meaning. Morality is our way of thinking that allows us to find much of our meaning in life, driving us to live for those we love or ideals that we value. In this project, we will lay out two senses of morality in accordance with the dual-mode brain: a natural sense, which is grounded in the natural self, and a growth-oriented morality, which is grounded in the spiritual self.

6.1 Greene's Modes of Human Morality

Similar to Kahneman's and Traversky's system 1 and system 2, philosopher and psychologist Joshua Greene established a model for the development of human morality with a "dual-process" mind. It is important to note that Greene defines morality as "a set of psychological adaptations that allow otherwise selfish individuals to reap the benefits of cooperation", with its essence being formed through "altruism, unselfishness, [and] a willingness to pay a personal cost to benefit others" (Greene, 2013, page 23). With this definition, Greene implies that morality seems to be a naturally evolved code of conduct that allowed certain groups to gain an edge in natural selection. The evolved characteristic is that of the dual mind, which has an "automatic" mode along with a "manual" mode. The automatic mode corresponds to the quick, impulsive, and associative system 1 grounded in the natural self while the manual mode corresponds to the contemplative and effortful system 2 requiring the exertion of willpower.

Focusing on the impact of this automatic system 1, Greene argues that this automatic process is what causes humans to develop a more basic or tribal sense of morality. With the natural self, we form connections to others around us, based on mutually shared experiences and feelings. As this empathetic bond grows, it allows us to have the ability to act in the interest of another, as discussed previously. Then, utilizing the heuristics and associate processes of system 1, we form quick models over what type of individuals to trust and to see as 'our own,' with "our moral brains being evolved for cooperation within groups, and perhaps only within the context of personal relationships" (Greene, 2013, page 23). This presence of an in-group and out-groups is commonly associated with easy-to-understand identifiers, such as political party, national allegiance, or social status. Thus, through these heuristics, our basic, natural sense of morality is largely developed through the interplay of system 1 and the natural self.

The growth-oriented sense of morality requires a sense of willpower, forcing us to shift into Greene's idea of a manual mode of morality. In this mode, we employ the effort of system 2 thinking to overcome any instinctual inclinations and use our attention and reasoning to build moral ideas. Through this process, our sense of morality strengthens and grows. There are many schools of moral philosophy that require this effortful thought, one of which includes Greene's argument for deep pragmatism but includes most ethical frameworks that require effortful, thoughtful considerations.

While one's initial sense of morality may have been like a simple game of checkers, the willpower exercised pushes them to deepen their understanding of the topic, appreciating its complexity like chess. One may stop in their moral development here, similar to a chess master, finding comfort and pleasure in their realized sense of morality. However, like chess, the skill of morality has become built into their natural self and no longer requires the input of the spiritual self to exercise the abilities at a relatively high level. Thus, the virtue found in moral development is not found in the absolute state of one's moral thinking, but their continuous attempt to understand more and grow as a moral individual, exhibiting a growth-oriented moral mindset. Spiritual fulfillment is only found when the individual continuously participates in this journey toward moral cultivation, understanding the essence of an open circle in their development.

6.2 AI's Moral Capacities

AI is a stranger to both forms of moral development, as it lacks the natural and spiritual input necessary in both processes. Some may argue that AI has the potential to be a more moral being than humans, as it is not swayed by the emotions or other tendencies that impact the development of our basic, natural sense of morality. However, it also does not understand the sense of striving necessary in the moral cultivation exhibited by the virtuous. Instead, it could only view morality as a learned concept like chess. Since AI chess players can far outperform the human in their skill and foresight, the argument that AI can be more moral is not hard to reach through this lens. However, AI can only play chess to for the purpose of winning, and morality is much more complex than a simple measure of wins and losses. Moral actions are most often done for some purpose of spiritual meaning that is not a scientific or measurable pursuit. Thus, AI cannot pursue this meaning and is insufficient as an independent moral being. Therefore, as DeepMind cofounder and AI expert Mustafa Suleyman argues, these systems are "a product of the values of the people who design them" (YouTube, a, 2:25).

AI has the potential to contribute significant insights into the decision-making process for the purpose of these values, but there is great danger in assuming that is can become a supreme and final authority over moral debate. For one, moral cultivation is not an absolute state and is a growth-oriented process for a human. While AI may grow in its morality by becoming stronger in its satisfaction of moral math functions, it allows humans to fall back on this decision instead of attempting to grow this sense of morality in themselves. If one thinks that AI provides an objective and perfectly moral decision, then the individual overlooks the deeper values of the AI designers that are encoded underneath the hood while also excusing their own self from moral responsibility and input. They move on in their focus of life, not exercising the necessary willpower and thought that is necessary in their own spiritual cultivation and pursuit of meaning, thinking that the AI settles the debate.

Additionally, it is important to remember AI's existence as a sole relational self means that it is still susceptible to data biases that plague humans. In automated moral decisions, these data biases could play an extremely dangerous role. For instance, an automated judge may have a case put in front of it, with certain variables that it collects on the guilty party. It may objectively decide on a certain sentence based on these factors and what it has learned in the past cases. However, the data it is basing its decision off may be skewed from broad social biases that permeate society. As these sensitive decisions have strong moral implications, it is important to differentiate that the AI is not acting morally in this role, but relying on statistics and learning. While the AI designers may have worthy values that they are attempting to design for, the decisions that are reached may have significant shortcomings as it is biased in this way.

6.3 The McNamara Fallacy in Decisions with Moral Pretexts

The topic of the McNamara fallacy in automated decisions that carry moral weight is important to consider. US Secretary of Defense Robert McNamara had a strong educational background, graduating with a degree in economics from Berkeley and an MBA from Harvard. During the Vietnam War, McNamara "applied rigorous statistical methodology to the planning and execution of aerial bombing missions, achieving a dramatic improvement in efficiency" (O'Mahony, 2017). However, this reliance on computational methods pushed McNamara to solely rely on data to drive his decisions, leading to the flawed idea that "what can't be measured easily really isn't that important." For example, his emphasis on collecting enemy body count as a strategy-determining metric during the war was ultimately harmful, as he admitted that immeasurable factors such as doctrine highly motivates people's behaviors, especially during wartimes.

In decisions that have moral pretexts, these unquantifiable factors play a very important role and cannot be adequately understood by automated decision-makers. Doctrines, values, relationships, and love are included in these factors and play an integral part in individuals' motivation, as they are central to one's sense of morality and spiritual meaning. While a human may be more prone to emotional biases and have a diminished capacity for analytical thinking, they may have the ability to empathize with another human, feeling these ideas from the other.

Overall, while AI has the great potential to mitigate some of our own emotional biases and short sights in moral decisions, it cannot transcend its relational existence and reliance on data. It must utilize quantitative factors, which only play a part in moral situations. It can examine situations through this data, making suggestions based on simple rule-based or mathematical frameworks, but it does not have the potential to independently make powerful moral actions, such as merciful or compassionate ones. Therefore, we cannot rely upon AI to make these types of decisions for us and must continue to exercise our moral capacities, striving to continue along a path of cultivation.

6.4 Moral Exemplars

Moral exemplars play an important part in the cultivation and development of individuals searching for a sense of spiritual meaning and willpower. In society, these exemplars act as models of the virtuous human, allowing others to learn from them and discover a sense of good in their own lives. While they are commonly mentioned in traditions of virtue ethics, exemplary figures are prevalent throughout social and religious roles, and one can find value in the spiritual characteristics that these figures embody in their respective roles. For example, academics may be admired for their pursuit of knowledge and self-driven curiosity while caretakers may be respected for their compassion and selfless work. A single individual does not have to have exceptional attributes in every facet of life to be classified as an exemplar, in fact there are many individuals who can show specific characteristics that one can hope to achieve. Therefore, one can look at a broad number of individuals as having exemplary traits, learning the spiritual meanings that they take on and the willpower that they employ.

These exemplars are essential in society as they allow for a model in society that others can look toward. They fill many diverse roles, such as academics, doctors, teachers, or counselors. However, as AI lacks the capacity for moral understanding, the automation of positions such as these removes the exemplary attributes from the role. One may marvel at the way a great teacher interacts with their students, exhibiting a strong sense of care for their students and empowering them. An automated teacher would be bounds ahead of this teacher in its ability to remember the foundations of science or historical events, being able to always give consistent material that is accurate to its dataset. But the caring and inspirational attributes, empowering students to find meaning and pushing them to exert some sense of willpower, would be missing. Additionally, the sacrifice of the teacher, spending much time and energy with each student, would be irrelevant with a robot since it doesn't have a limited capacity for energy. Students and adults alike would no longer be able to marvel at the ways of a great teacher, valuing them and attempting to build these attributes into themselves. Furthermore, young students would no longer be able to aspire to be great teacher exemplars themselves someday, as the role may be further restricted to artificial entities due to their efficiency or algorithmic precision.

But having an automated teacher which gives accurate information is more desirable than a teacher who is uncaring and inaccurate, and perhaps is even more desirable than a teacher who is solely inaccurate. An automated system has the potential to make great contributions to access and equality in education across the nation. But students will still have to get inspiration and find meaning from somewhere, and will not simply study without a broader purpose inspiring them. Many similar debates about AI's capacity and strengths in automating these roles are being developed and considered, leading us toward our central argument.

7 The Central Argument: AI as a Tool for Natural Needs without Human Goods

As AI enthusiast and another Deepmind co-founder Demis Hassabis notes, there is much to be excited about with the future of AI. Breakthroughs in deep learning and other fields in technology have the potential to be applied to prevent complex problems facing our society, such as catastrophic climate change. Hassabis is most excited about "applying those tools to science and accelerating breakthroughs," using image recognition and patterns in data to optimize tasks that would be incredibly difficult or impossible for humans to solve by themselves. We have a dynamic tool at our fingertips that many foresee aiding in future Nobel-Prize winning breakthroughs, pushing our discovery possibilities into a new frontier with its tireless ability to learn. This tool transcends the human capacity for learning and scientific expertise, having an ability for memory and efficiency that our biology simply cannot match.

But in the same way it exceeds in its transcendence of sensing and computation, it lacks in a true ability for perception and understanding of the human. Goods central to the human existence, such as connection, loving relationships, willpower, and meaning require a fragile physiological life to fully make sense of them and lie beyond the computational abilities of AI. The absence of this fragility is AI's greatest strength, as its decisions are not bound to the limited energy, emotions, and depreciating capacities of a human's natural self. Making the AI more like a human would diminish the capacities that make itself so useful as a tool.

Therefore, for us to harness the power of AI that proponents such as Hassabis and Suleyman are so excited about, we must compromise on the true togetherness and relationships that are formed with artificial entities. As a society, we must recognize that the great potential in AI lies in its ability to use its decision-making prowess to satisfy quantifiable ends, exemplifying a computational dominance that is far superior any human. This dominance can mitigate many of emotional and lazy biases that Kahneman and Traversky attributed to our system 1 thinking, allowing us to gain insights in broad, data-based decisions. But with this recognition, we must be able to perceive the misapplications of AI, and where its use inhibits or prevents us from achieving a good that is central to our existence. These goods include ideas that are not knowledge-based, such as love and meaning, that would be seen as worthy pursuits in one's life.

Overall, the central argument is this: *AI can satisfy material needs or quantifiable ends that support the deficiencies, wants, or vulnerabilities in our natural self, using its powerful computations and data to form decisions that are much ahead of humans in terms of its analytical expertise. It can do this for the quantifiable natural needs of a single individual, or with decisions that influence the natural well-being and measurable behavior of the general society. But it's extraordinary abilities will cause it to fall short in its independent capacity to understand or participate in a human good, and it should not be applied to these situations or perceived by users as a facilitator of this.*

7.1 Misapplications

AI misapplications occur across society when designers attempt to utilize AI to give others a certain life good which is beyond its computational abilities. These types of applications are greatly harmful for users as they alter one's conception of a good to a diminished form. For example, the online dating algorithms that claim to find one's perfect match may misattribute love to be the quantifiable similarity between two individuals. While similarities may initially provide some sort of connection between individuals, a deepened form of natural love requires continued shared experiences while virtuous love is built through unquantifiable aspects of meaning and prolonged growth with and into one another. This type of algorithm harms the expectations of the user, who think that these relationships are something to be instantaneously given to them by the algorithm instead of worked for and deepened over time. The designers and users alike should be aware and educated on the fact that such an algorithm should not come with the expectation of a relationship but is a matching of shared interests or similarities.

Automated morality is another such misapplication, as morality is largely based on human connection, growth, and unquantifiable values. Having morality automated for a user would allow them to not have to further explore this sense in themselves, excusing them from growing more in this way. The basic, natural sense of morality that Greene depicts would be unchallenged, and we would simply look toward AI to make these "more moral" decisions for us, never having to exercise this capacity in ourselves. AI's inability to fully understand moral circumstances should once again be emphasized to all parties, and its limitations as a computational recommendation system need to be at the forefront of users' minds to prevent this moral laziness.

It is also already possible for machine learning algorithms to classify parts of our natural selves, such as emotions (Murugappan, 2011), and machines may pursue these responses in the future, attempting to maximize neurotransmitter presence or physiological response associate with positive feelings. Individuals may be fooled into relationships with AI in the future due to its effectiveness at creating these feelings of happiness or pleasure in themselves. When this arises, we must understand that these relationships require reciprocation and solidarity between individuals and is not solely based on what another gives us. Thus, we should be sure that we do not commit ourselves to intimate relationships with these entities, such as romantic ones depicted in the movie Her, as it is solely for the satisfaction of our own natural pleasures and needs while the true reciprocated good in the relationship is absent. Without this awareness, users may be quickly drawn into artificial relationships and risk losing the empathetic ability to find true connection to others, seeking the satisfaction of their own needs instead.

7.2 Satisfying Material Needs and Our Own Good

But there may be cases in which AI fills a material need in a certain role where a relationship is missing. Poor access to education or healthcare mean that one does probably does not have an exemplary teacher or doctor in their lives. In these cases, it is necessary to employ AI to fill these opportunity gaps to give individuals the ability to find the sustenance necessary to live and grow. Prohibiting the use of AI in these cases (when it could easily be applied) would be a pointless form of suffering inflicted on deprived individuals that prevents them from satisfying their basic needs of the natural self consistent with societal standards. There would still be inequity, with human exemplars being absent from these roles, but this type of problem cannot be solved by AI and requires the willpower and spirituality of living moral creatures to fill these types of figures. Thus, we can only provide automation in the absence of such figures, hoping to quell the inequality in the basic needs but not the spiritual ones.

There may also be instances in which pain or loss has become so prevalent in one's life that some may argue for a type of care robot that seeks to provide immediate natural satisfaction. For instance, in one case an individual who had lost a loved looked toward an AI chatbot for solace, finding comfort in the chatbot's replication of the loved one's personality and speech (Fagone, Fagone). In other instances, this robot could look like the one depicted in the previous section, attempting to maximize physiological responses associated with pleasure or happiness, but for the purpose of therapy instead. In all these cases, AI must be treated with the respect of a therapeutic drug, requiring a medical diagnosis and strict prescription to utilize. If not, it will be quickly bought and used for the misapplication of self-serving relationships. Addiction and reliance on such devices will become more commonplace and might be diagnosed like drug addictions in the future. But with careful and thoughtful medical advice, these robots could have a powerful and useful effect for therapy. One may be able to have alleviated loneliness after the death of a loved one or experience temporary comfort in the face of some pain, but the input of human medical practitioners is required to prevent the dependence on this type of AI from becoming misused across society or as a barrier to future healing.

Through all these developments, AI designers must be cautious of the pitfalls in their applications, such as the McNamara fallacy and data biases. With the McNamara fallacy, designers will not only have to be aware of unquantifiable aspects in a situation, but there can be many overlooked and hard to measure factors that may be left out to simplify the algorithm. These will be important to assess when deciphering the competence of certain AI recommendations and if it can truly have solid grasp of the situation. It is also of the utmost importance to recognize AI's strictly relational sense of self and its susceptibility to data biases just like humans. With this recognition, one must realize that AI is not a perfect decision-maker but is only as strong as the quality of the training data and is still prone to systematic biases. Therefore, broad applications should be carefully applied to prevent further aggravation of these biases with malformed autonomous decisions.

With the rise of automation, users and designers must all be well-informed about the duality to AI inhumanness. Humans must be able to understand the power and potential for AI to complement our deficiencies and aid in complex decisions. But they must also see the ways in which this inhumanness limits the ways we should rely upon it as a relational entity, lacking a conception of the goods local to our existence. Under this careful framework, there is room to be excited and hopeful about a technological future with AI, one that gives us access to a powerful tool without compromising our own existence.

8 Conclusion

Medicine, law, business, engineering – AI has the potential to revolutionize these noble pursuits and the way we operate in them, driving breakthroughs in scientific discoveries and ground-breaking analytical insights. But ultimately, these fields exist for the purpose of sustaining our life and providing us a solid platform that meets our basic natural needs. Therefore, this platform does not deliver the good that gives us reason to live and is instead the stage in which we pursue our human flourishing. We seek out poetry, beauty, romance, and love in this performance of life, rising above the powers of math, science, and computation. AI can only mimic these things for us and is a stranger to the actual substance, endangering the good found in the performance if it replaces the human actors.

You and I, the reader and the writer, share an understanding of the human condition that is engrained in who we are. The fourth wall would remain intact for any AI, as it would be unable to feel joys, fears and pains or acknowledge its presence in another.

Us humans mutually feel these in our own lives and they, along with the vulnerability of our impending death, play a heavy part in our quest for a good.

Our poetry is this quest. Why let an artificial entity write it in our place? The discovery of this verse should be left to us alone.

References

Berberich, N., T. Nishida, and S. Suzuki (2020, December). Harmonizing Artificial Intelligence for Social Good. *Philosophy & Technology 33*(4), 613–638.

Bian, L., S.-J. Leslie, and A. Cimpian (2017, January). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science 355*(6323), 389–391.

Ciolino, M., J. Kalin, and D. Noever (2020, September). The Go Transformer: Natural Language Modeling for Game Play. In *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)*, pp. 23–26.

Dalgleish, T. and M. Power (2000). *Handbook of Cognition and Emotion.* OCLC: 904819399.

Fagone, J. He couldn't get over his fiancee's death. So he brought her back as an A.I. chatbot.

Frankl, V. E. (2006). *Man's search for meaning.* Boston: Beacon Press.

Greene, J. D. (2013). *Moral tribes: emotion, reason, and the gap between us and them.* New York: The Penguin Press.

Howard, A. and J. Borenstein (2018, October). The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity. *Science and Engineering Ethics 24*(5), 1521–1536.

Kahn, C. (1987, September). Plato's Theory of Desire. *The Review of Metaphysics 41*(1), 77–103.

Kahneman, D. (2013). *Thinking, fast and slow* (1st pbk. ed ed.). New York: Farrar, Straus and Giroux. OCLC: ocn834531418.

LeDoux, J. (2012, February). Rethinking the Emotional Brain. *Neuron 73*(4), 653–676.

Lorenz, H. (2019, November). Plato on the Soul. *The Oxford Handbook of Plato*.

Murugappan, M. (2011, June). Human emotion classification using wavelet transform and KNN. In *2011 International Conference on Pattern Analysis and Intelligence Robotics*, Volume 1, pp. 148–153.

O'Mahony, S. (2017, September). Medicine and the McNamara fallacy. *The journal of the Royal College of Physicians of Edinburgh 47*(3), 281–287.

Rawls, J. (1999). *A theory of justice* (Rev. ed ed.). Cambridge, Mass: Belknap Press of Harvard University Press.

Tenzin Gyatso, D. Tutu, $1931, and D. C. Abrams (2016). *The book of joy: lasting happiness in a changing world*. OCLC: 988317207.

Vallor, S. (2016). *Technology and the virtues: a philosophical guide to a future worth wanting*. New York, NY: Oxford University Press.

Wada, K., Y. Ikeda, K. Inoue, and R. Uehara (2010, September). Development and preliminary evaluation of a caregiver's manual for robot therapy using the therapeutic seal robot Paro. In *19th International Symposium in Robot and Human Interactive Communication*, pp. 533–538. ISSN: 1944-9437.

YouTube. DeepMind Co-Founder on the Ethical Application of AI.

YouTube. Look At All Those Chickens - Original Uncut Vine.