The Sapient and Sentient Intelligence Value Argument (SSIVA) Ethical Model Theory for Artificial General Intelligence

By David J. Kelley
AGI Laboratory, Provo Utah
david@artificialgeneralintelligenceinc.com

Abstract This paper defines what the Sapient Sentient Value Argument Theory is and why it is vital to AGI research as the basis for a computable, human-compatible model of ethics that can be mathematically modeled and used as the basis for teaching AGI systems, allowing them to interact and live in society independent of humans. The structure and computability of SSIVA theory make it something we can test and be confident in for such ICOM based AGI systems. This paper compares and contrasts various SSIVA theory issues, including known edge cases and issues with SSIVA theory from legal considerations, to compare it to other ethical models or related thinking.

Keywords: *ethics, ai, agi, autonomous systems, artificial intelligence, artificial general intelligence, ssiva, iva, sapient, sentient intelligence value argument theory, ssiva theory, intelligence value argument*

Introduction

One problem with biasing and designing artificial general intelligence (AGI) cognitive architectures is, in part, how to compute things like ethics and behavior. Given the human model and the desire to replicate that at a high level, at least in so much that it uses the computation theory of mind. Moreover, wanting to bias all choices, decisions, and motivations on complex emotional states that are also computable, including computing qualia in the broad sense of objectively computing subjective experience, we needed an ethical model to frame processes of conditioning behavior in such a way as to make the system safe to be around. Given that most ethical models are not readily computable, we started from the ground up to design such a system. The Sapient/Sentient Intelligence Value Argument (SSIVA) 'theory' of ethics is our model we have designed that an AGI Independent Core Observer Model (ICOM) based system can experience "subjectively" while we measure that experience objectively. This paper focuses only on the ethical model itself.

SSIVA states that "ethically," a fully sapient and sentient intelligence is of equal value regardless of the underlying substrate which it operates on, meaning a single fully sapient and sentient software system has the same moral agency [10] as an equally sapient and sentient human being, and nothing else has greater value.

Let us break down that into its parts and the reasoning thereof, including contextual definitions, to be exact, as this is required for the research program, the theory is designed to enable.

According to dictionary.com, we first define **'ethical'** as "pertaining to or dealing with morals or the principle*s of morality; about right and wrong in conduct."*

According to Wikipedia; **Moral Agency** is *"an individual's ability to make moral judgments based on some notion of right and wrong and to be held accountable for these actions. A moral agent is "a being who is capable of acting regarding right and wrong."*

Such value judgments in the moral or ethical sense based on SSIVA theory need to be based on the potential for sapient and sentient intelligence to prevent ethical problems across various edge cases. For example, can the system kill a newborn child if the child is a drain on resources? Under SSIVA, the answer is a clear and resounding 'no' because the child, even though it might not meet some of the parameters of being full sapient and sentient now, can and has the potential to and therefore has the right to moral agency, and therefore there is not any case in which this is a moral or ethical action, and thus a clear and present boundary condition that cannot be crossed.

This line of reasoning places the value of any individual human and their potential for intelligence above virtually all things save the one wherein a single machine intelligence capable of extending its own Sapient and Sentient Intelligence is of equal value as any other human-based on a function of their potential for Sapient and Sentient Intelligence, thus being assigned the right of moral agency. It is not that human or machine intelligence is more valuable than the other inherently, but that value is a function of the potential for Sapient and Sentient Intelligence, and SSIVA argues that at a certain threshold, all such intelligences should be treated equally as having moral equivalence. Given this equality, we can, in effect, apply the same ethical rules and apply them to such software systems that exhibit the same levels of Sapient and Sentient Intelligence. Let us start from the beginning and define the critical elements of the SSIVA theory.

While the same 'value' between intelligence is implied, it is the treatment as equals in making their mind through their moral agency from which we derive all 'value' from the SSIVA ethical model as applied to humans and AGI systems fully sapient and sentient. Any more 'value' than this becomes abstract and is therefore subjective (not consistently computable between independent minds). The moral agency is the right we assign to those Sapient and Sentient intelligences based on the value of the potential of such entities being the same.

To break that down further, let asking;

What is the most important thing in existence?

On the surface, this seems a very existential question, but, in truth, there is a simple and elegant answer; Sapient and Sentient Intelligence is the most important thing in existence. One might ask why? Why is Sapient and Sentient Intelligence so important as to be the most important thing in existence, especially when 'value' is subjective?

First, let us acquire some context by defining what intelligence is in this context, which will then act as our base frame of reference for the rest of this argument. There are, in fact, many definitions for intelligence as can be seen by its definition on Evolutionary Computer Vision [1]

> *"Intelligence … defined in many different ways including, but not limited to, abstract thought, understanding, self-awareness, communication, reasoning, learning, having emotional knowledge, retaining, planning, and problem-solving."*

As one can see, these are some of the ways the term can be understood; but in this paper, 'intelligence' is defined as the measured ability to understand, use, and generate knowledge or information independently. This definition allows us to use the term 'Intelligence' in place of sapience and sentience where we would otherwise need to state both in this context where we have chosen to do that, in any case, to make the argument more easily understood.

It is important to note that this definition is more expansive than the meaning we assign to sapience, which many people mean when they use the often-misunderstood term sentience.

Sapience [11]:

> *"Wisdom [Sapience] is the judicious application of knowledge. It is a deep understanding and realization of people, things, events, or situations, resulting in applying perceptions, judgments, and actions in keeping with this understanding. It often requires control of one's emotional reactions (the "passions"), so that universal principles, reason, and knowledge prevail to determine one's actions. Wisdom is also the comprehension of what is right, coupled with optimum judgment as to action."*

As opposed to Sentience [15], which is:

> *"Sentience is the ability to feel, perceive, or be conscious, or to have subjective experiences. Eighteenth-century philosophers used the concept to distinguish the ability to think ("reason") from the ability to feel ("sentience"). In modern western philosophy, sentience is the ability to have sensations or experiences (described by some thinkers as "qualia")."*

Based on these definitions, we see the difference with the term sapience vs. sentience, where sapience is more closely aligned with what we are driving at in the SSIVA theory. That notwithstanding, it is Sapience and Sentience together that we will consider by using the term Sapience and Sentient Intelligence.

In the case of the SSIVA, we will apply sapience to refer specifically to the ability to understand oneself in every aspect; through the application of knowledge, information, and independent analysis and to have subjective experiences. Although sapience depends on intelligence or the degree of sapience depends on the degree of intelligence, they are different. The premise that intelligence is important, and the most important thing in existence, is better stated as Sapient Intelligence is of primary importance, but intelligence (less than truly Sentient Intelligence) is relatively unimportant in comparison.

Binging us back to the point about "Why?" Why is Sapient and Sentient Intelligence, as defined earlier, so important? The reason is: without Sapient and Sentient Intelligence, there would be no witness to reality, no appreciation for anything of beauty, no love, no kindness, and for all intents and purposes, no deliberate creation of any kind. This is important from moral and ethical standpoints, in that only through the use of applied 'Intelligence' can we determine "value" at all, even though once intelligence is established as the basis for assigning value, the rest becomes highly subjective.

It is fair to point out that even with this assessment, there would be no subjective things of relative value such as love or kindness without Sapient and Sentient Intelligence to appreciate those things' subjective nature. Even in that argument about subjectivity, it is only through one's intelligence that one can make such an assessment. Therefore the foundation of any subjective experience that we can discuss always gets back to having the prerequisite Intelligence to be able to make the argument.

Without a Sapient and Sentient "Intelligence," there would be no point to anything; therefore, sapient and sentient Intelligence is the essential quality, or there is no value or way to assign value and no one or nothing to hold to any value of any kind.

That is to say that Sapient and Sentient "intelligence," as defined earlier, is the foundation of assigning value objectively, and thus needed before anything else can be assigned subjective value. Even the subjective experience of a given sapient and sentient intelligence has no value without an Intelligence to assign that value.

Through this line of thought, we also conclude that intelligence being important is not connected with being human, nor is it related to biology, but the main point is that Sapient and Sentient Intelligence, regardless of form, is the single most important 'thing.'

Therefore, it is our moral and ethical imperative to maintain our own or any other fully Sentient and Sapient Intelligence (as defined later with the idea of the SSIVA threshold) forever as a function of the preservation of the basis for 'value.'

On Artificial General Intelligence

Whatever entity achieves full Sapient and Sentient Intelligence, as defined above, it is, therefore, a moral agent. Artificial Intelligence refers to soft AI or even the programmed behavior of an ant colony, which is not essential in being compared to fully Sapient and Sentient Intelligence. However, the idea of "Strong AI," which is genuinely sapient and sentient Intelligence, would be of the most value and would be classified as any other human or similar Sapient Intelligence.

From an ethical standpoint, 'value' is a function of the 'potential' for fully Sapient and Sentient Intelligence independent of other factors. Therefore, if an AGI that is 'intelligent' by the above definition and is capable of self-modification (in terms of cognitive architecture and sapient and sentient Intelligence) and increasing its 'Intelligence' to any easily defined limits, then its 'value' is at least as much as any human. Given that 'value' tends to be subjective, SSIVA argues that any 'species' or system that can meet this limit is said to hit the SSIVA threshold and has moral agency and is equal ethically amongst themselves. This draws a line in terms of moral agency, in which case we have a basis for assigning AGI that meets these criteria as having 'human' rights in the more traditional sense or, in other words, 'personhood.'

This, of course, also places the value of any individual fully Sapient and Sentient Intelligence, human or otherwise, and their potential for Sapient and Sentient Intelligence above virtually all other considerations. All moral agents are equal ethically based on the SSIVA theory.

Removing 'Artificial' from Intelligence

The SSIVA line of thought really would take the term 'Artificial' out of the term(s) "Artificial Intelligence." If such a software system is fully Sapient and Sentient, it is not an 'artificial' Intelligence. However, it would be better to call it a human-made Intelligence, or a machine Intelligence, as the term 'artificial,' besides implying being human-made, implies that it is a fake Intelligence, not a real Intelligence. A real 'AGI' would not be fake based on the SSIVA line of thinking and would have as much moral value as any other human being.

The SSIVA Threshold

One problem with the SSIVA threshold is determining the line for sapient and sentient intelligence, assigning moral agency. The SSIVA threshold is at the point of full Sapience and Sentience in terms of understanding and reflecting on one's self and one's technical operation while also reflecting on that

same process emotionally and subjectively or having the capacity to do the same. We draw the line not just at that threshold but also at the potential of meeting that threshold, which allows us better to address edge cases with a stable clear-cut line. Using SSIVA thinking, a post-threshold Sapient and Sentient Intelligence, meaning one that has met the SSIVA threshold, cannot be prevented from creating new Sapient and Sentient Intelligence ethically, and must also, therefore, consider that any being whose potential for being a fully Sapient and Sentient intelligence, without direct manipulation or reengineering external to that agent at the lowest mechanical (chemical/biological or physical level), is considered post-threshold as well from an ethical standpoint, being assigned moral agency or the right to such. Therefore, any such 'Intelligence,' regardless of form, has the same rights as any other Sapient or Sentient being whose creators are then ethically bound to exercise the rights of that entity, to protect it until it is developed enough to take on its self as the fully Sapient and Sentient being that it is or will become.

This also implies that an AGI will not meet the threshold until the first AGI does meet the threshold. A baby AGI does not meet the threshold until it is proven that the system can develop on its own without additional engineering, either by human or other Sapient and Sentient Intelligent agents. Along those lines, then SSIVA theory would argue that any action that would kill or prevent an entity that meets the bar from being fully Sapient or Sentient would be unethical unless there is a dire need to save the lives of other entities at or above the SSIVA thresh hold. Meaning, two SSIVA entities are of more value than one, and in the case of a choice between saving one AGI or two humans, one must pick two humans to be strictly speaking ethically compliant with SSIVA theory-based ethics. Additionally, if the choice is one human vs. two AGI agents that hit the threshold, then the lives of those two AGI systems are more important than the one human.

Defining the Bar for the SSIVA threshold

Having a discreet method of measuring sapient and sentient intelligence is essential not just for the SSIVA threshold model in this application but for research into AI systems generally. While the above definition of sapient and sentient intelligence in the abstract allows us to discuss the matter from a common point of reference, for additional work to be built on this, it is essential to be more precise in defining Sapient and Sentient Intelligence as referenced in SSIVA theory.

There are several systems like the "Intelligence Quotient" [17] tests, but that is not as specific to sapient and sentient intelligence as we would want to be given the key differences between intelligence as customarily defined and "sapience and sentience" as used here. The best model for this comes from a paper [16] by Dr. Porter from Portland State University, wherein the 2016 BICA proceedings, he has a paper articulating an indexed system for measuring consciousness. While individual elements of Dr. Porter's system for assessing consciousness might be subjective, the overall system is the best quantifiable method currently available.

In Dr. Porter's method, we essentially have a scale of 0 to 133, where the standard human is around 100 on that scale. Given that the SSIVA threshold is about the potential for sapience and sentience, we can say that having a consciousness score potential of roughly 100 points on the Porter scale is high enough to say that 'that species' or 'Intelligence' meets the SSIVA threshold test. There is some differential as the Porter test does not differentiate between Sapient and Sentient, but it is inclusive enough to give us a basis of measurement for determining if a given system is approaching the SSIVA threshold. This

allows us to apply that standard to any machine intelligences we may create in the lab, to determine at what point they can meet that standard.

Comparing and Contrasting Related Thinking to SSIVA

In building out an argument to support the aforementioned ethical model based on the 'value' of Intelligence related to sapient and sentient entities such as artificial general Intelligence software systems and humanity, let us compare other related lines of thinking related to the following cases.

Utility Monster and Utilitarianism

The Utility Monster [1] was part of a thought experiment by Robert Nozick related to his critique of utilitarianism. Essentially this was a theoretical utility monster that got more 'utility' from X than it did from humanity, so the line of thinking was that the "Utility Monster" should get all of the X even at the cost of the death of all humanity.

One problem with the Utility Monster line of thinking is that it puts the wants and needs of a single entity based on its assigned values higher than that of other entities. This is a fundamental disagreement with SSIVA theory, where SSIVA theory would argue that one can never put any value of anything other than other sapient and sentient intelligences themselves above any other assigned value. This would mean that the utility monster scenario would be purely unethical from that standpoint.

Utilitarianism does not align with SSIVA thinking for an ethical framework, as utilitarianism asserts that 'utility' is the key measure in judging what we should or should not be ethical. In contrast, the SSIVA theory makes no such ascertain of value or utility except that sapient and sentient Intelligence is required to assign value in the first place, and past that "value" then becomes subjective to the Intelligence in question. The Utility Monster argument completely disregards the value of post-threshold sapient and sentient intelligences, and by SSIVA standards would be completely unethical.

Buchanan on Moral Status and Human Enhancement

In the paper 'Moral Status and Human Enhancement" [3], the paper argues against the creation of inequality regarding enhancement. In this case, the SSIVA is not directly related unless one gets into the definition of the SSIVA theory ethical basis of value and the fact that having moral agency under SSIVA theory means that only post-threshold moral agents can make a judgment as to any enhancement. It would be a violation of that entity's rights to put any restriction on enhancement.

Buchanan's paper argues that enhancement could produce inequality around moral status, which gets into areas that SSIVA theory does not address, or frankly disregards as irrelevant, except that in having full moral agency, we would not have the right to put any limits on another entity without violating their agency, therefore making those limits unethical.

Additional deviations with Buchanan include that sentience is the basis for moral status. In contrast, SSIVA theory makes a case for sentience and sapience together being the basis for 'value,' and we assume that definition or intent is similar to this idea of 'moral status' as articulated by Buchanan.

Intelligence and Moral Status

Other researchers such as Russell Powell further make a case that cognitive capabilities bear on moral status [4], whereas SSIVA does not directly address moral status other than the potential to meet the SSIVA theory threshold grants that moral status. Powell suggests that mental enhancement would change moral status; SSIVA theory would argue that once an entity can cross the SSIVA theory threshold, their moral status is uniform. The most considerable discrepancy between Powell and SSIVA theory is that Powell makes the case that we should not create persons where SSIVA would argue it is an ethical imperative.

Persons, Post-persons, and Thresholds

Dr. Wilson argues in a paper titled "Persons, Post-persons and Thresholds" [5] (which is related to the aforementioned paper by Buchanan) that 'post-persons' (being enhanced persons through whatever means) do not have the right to a higher moral status where he also argues the line should be Sentience to assign 'moral' status. In contrast, SSIVA theory would argue that the line for judgment of 'value' is that of the potential for Sapient and Sentient Intelligence together. While the bulk of this paper gets into material that is out of scope for SSIVA theory specific to this line for moral status, SSIVA theory does build on the line for 'value' or 'moral status' including both Sapient and Sentient Intelligence.

Taking the "Human" Out of Human Rights [6] is a paper that supports the SSIVA theory support argument in no small degree, in terms of removing 'human' from the idea of human rights. Generally, SSIVA theory would assert that 'rights' is a function of the sapient and sentient Intelligence based on the potential for sapience and sentience, and anything below that threshold would be a resource. In contrast, Harris's paper asserts that "human rights" is a concept of beings of a particular sort and should not be tied to species, but still accepts that a threshold, or, as the paper asserts, that these properties are held by entities regardless of species which would also imply that such would extend to AI as well which would be in line with SSIVA theory-based thinking. Interestingly, Harris further asserts that there are dangers with not actively pursuing further research, making a case for not limiting research, which is a significant component of SSIVA theory-based thinking.

The Moral Status of Post-Persons [7]

This paper by Hauskeller is in part focused on Nicholas Agar's argument on the moral superiority of "post-persons," and while SSIVA theory would agree with Hauskeller that his conclusion in the original work is wrong; namely, he asserts that it would be morally wrong to allow cognitive enhancement, Hauskeller's argument seems to revolve around the ambiguity of assigning value. Where SSIVA theory and Hauskeller differ is that as a function of intelligence, SSIVA theory would place absolute value on the function of immediate self-realized Sapient and Sentient Intelligence, in which case a superior Intelligence would be of equal value from a moral standpoint. SSIVA theory disregards other value measures as being subjective due to their assignment requiring sapient and sentient intelligence. SSIVA theory asserts that moral agency is based on the SSIVA theory threshold.

Now, if we go back to Agar's original paper [8], his second argument is wildly out of alignment with SSIVA theory, namely that Agar argues it is 'bad' to create superior Intelligences. SSIVA theory would assert that we would be morally or ethically obligated to create greater intelligences because it creates the most 'value' in Sapient and Sentience Intelligence. It is not the 'moral' assignment but the base value of Sapient and Sentient Intelligence that assigns such value, as subjective as that may be. Agar's ambiguous argument that it would be 'bad' and the logic that "since we do not have a moral obligation

to create such beings we should not" is opposite of the SSIVA theory-based argument that we are morally obligated to create such beings if possible.

Rights of Artificial Intelligence

Eric Schwitzgebel and Mara Garza [9] make a case for the rights of Artificial Intelligence, which at a high-level SSIVA theory-based thinking would support the idea articulated in their paper, but there are issues as you drill into it. For example, Schwitzgebel and Garza conclude that developing an adequate theory of consciousness is a moral imperative. SSIVA theory ignores this altogether as unrelated to the core issue where SSIVA works because consciousness is solved. Further, their paper argues that we should avoid creating moral entities whose moral status is reasonably disputable. SSIVA theory does not deal with creating such systems but deals with the systems once created. The big issue with SSIVA theory around AGI is that value exists in all sapient and sentient intelligence, and the implication is to optimize for the most value to the most sapient and sentient intelligences.

Problems with SSIVA Theory

There are elements of SSVIA Theory that could be construed as problematic and have been brought up on review on several occasions, and to that end, let us address those now.

Non-Emotion Based Intelligences

SSIVA Theory does define systems that do not use emotions but could be considered an 'intelligence.' That being the case, either from super intelligent logical robots or space aliens that use some kind of purely logical model for their 'version' of consciousness, SSIVA does not address, these 'entities,' if proven to exist, from SSIVA Theory standpoint are not moral agents, given the lack of emotional subjective experience. SSIVA Theory would need to be modified and extended, or such 'entities' would be resources under the SSIVA theoretical model, which could be problematic, but for the research we are doing, this is outside the narrow scope of supporting that research; therefore, this is a known issue with SSIVA Theory, but a purely theoretical edge case.

Biased towards the Human Experience

SSIVA has been accused of being biased towards the human experience. This is true. SSIVA Theory is biased towards the subjective human experience as it needed to be like this to include AGI systems built on similar emotional models and nothing more. One of the key design parameters was building an ethical model that is computable and closely aligned with the human experience. We need our systems we build, teach, and train to love, cherish, and respect humanity, and we need to be able to document that experience mathematically to 'prove' safety in such systems.

Dealing with Sub SSIVA Threshold Individuals

From an SSIVA Threshold standpoint, this could very well produce a moral dilemma. Given that this is a grey area, SSIVA must fall on the side of not having ambiguity, so, therefore, those 'individual' members of a give species (that is known to meet that threshold standard) who do not meet the threshold should be considered to have moral agency, or rather the proxy of that agency, and should be assigned to the community to care for, help, and ideally 'fix' the individual, at the very least to the level of meeting the threshold of possible, and otherwise care for that member if they cannot be fixed. This approach maintains the ethical integrity of the model in terms of safe interaction by assigning moral

agency collectively to all members of a given species if even one member meets the threshold and assigns the social responsibility to care for that community.

Eugenics vs. SSIVA Theory

SSIVA theory is an ethical model, so making comparisons to something like eugenics is a bit like comparing apples and tires merely because both are vaguely round. That being the case, we will approach the problem to make those differences clear. First 'eugenics' generally is about improving the human population through controlled breeding, and the Nazis used this as an excuse to commit horrifying atrocities. From an SSIVA Theory standpoint, forced eugenics of any sort would violate the moral agency of others and would therefore be unethical. That is not to say that individuals could not improve themselves or their children, but there is no case at all where it would be ok to force something like eugenics.

Near-Threshold Entities

Near-threshold species under SSIVA Theory are nothing more than resources. While there may be an ethical obligation to manage those species as a resource, there is no obligation from the SSIVA standpoint to do anything about their status. While this is an edge case uplifting a species is not something SSIVA prohibits, it does not imply either approach, and while the species in question is pre-threshold, it is merely a resource from the SSIVA theory standpoint.

This particular issue is also related to the issue of superintelligence vs. human intelligence. Does superintelligence have a moral obligation to uplift humanity? Is a superintelligence of the same subjective value as a human? Under SSIVA as currently articulated, the superintelligence and the human being of equal moral agency. Sure, if one measures some quantifiable element of the superintelligence that is far superior to humanity and could thus be construed as of more value, it is not easy to quantify. As we approach superintelligence with AGI, we may well need to bake this out further, and this is a known problem with SSIVA theory.

Can we, therefore, deal with building on SSIVA theory in terms of maximizing intelligence and thus imply a moral imperative to uplift if those intelligences are willing? This can be done, but has not been done and is outside the current research scope.

The Current Legal System and SSIVA Theory

One of many complications with ever-increasing technology based on machine Intelligence is the increasing pressure to regulate such systems, as in self-driving cars. Questions arise, such as "when there is an accident, who is at fault?" and "How do we apply the law to such systems?". These same sorts of questions come up in medicine or other professions where AI-based systems are taking on increasingly human-like roles. In one recent case, the senior management of a company was replaced by machine intelligence [13]. Currently, these systems do not qualify as being self-aware by various standards; however, the time is quickly approaching, in which case it will. Contrary to popular belief, the current legal system could be successfully applied to most cases and include applying specific laws to systems that do not even exist yet, such as full AGI systems without much, if any, additional law.

Let us look at the sides;

On the side of regulation, we have big names like Elon Musk, Stephen Hawking, and the like on record as being pro-regulation, in that we need to make sure it is regulated to keep AI from getting ahead of itself [14] in terms of moving beyond human control; the fear being that someone will use AI to do something terrible. Frequently there are issues about quality control or the like, but these are generally the same issues with the same regulations as might be applied with or without the AI under existing law. Given that the current body of law can treat AI systems like any other software system, the proponents of regulation tend to be focused on future scenarios that do not work with current AI systems and, therefore, not applicable now.

However, let us look at the scenario of fully Sapient and Sentient AI systems or AGI;

SSIVA AGI and Current Regulations;

SSIVA makes a case for Sapient and Sentient value inherent in systems that meet the SSIVA threshold. IF SSIVA is used as the basis for separating systems into two classifications from a legal standpoint, then we essentially can apply two sets of current laws to those intelligent systems. Those two groups are the pre-threshold systems (which is the current state of the art in terms of AI or AGI) and post-threshold systems, which are the systems featured by regulation proponents such as Musk.

Using SSIVA theory applied to post-threshold systems, they are then essentially people or should be considered so from a legal standpoint. With any post-threshold system, the same laws that govern people then can immediately be used to govern such systems, and the only issue is then making sure the law classifies them as machine-based persons, in which case they have the same rights as human people and the same laws can be used to govern them. This is important because it allows the current body of law to govern the actions of humans to be applied to these same systems that are as capable of morals and ethics as people.

That is not to say that 'no' laws will be needed, but if we start by assigning post-threshold systems ' personhood' and identifying how we might hold such systems accountable much of the current law would apply as it does to other 'people.'

In the case of pre-threshold systems, laws applying to the relevant segment of devices can be applied. Take the case of self-driving cars. A modern self-driving car is a pre-SSIVA threshold and therefore is just a more advanced car. The same rules that apply to cars apply to autonomous cars. Accidents at the car's fault are on the driver of the vehicle unless there is a shortcoming in the system, in which case it becomes an issue of the manufacturer. If a car manufacturer puts faulty tires on a car, it is the manufacturer that bears the burden, and the same can be applied to the software systems of self-driving cars. If the manufacturer's AI drives the car that messes up, then it is the manufacturer's faulty hardware. Without the burden of additional legal standards, we can apply the existing law to self-driving cars without the need for additional legislation because these cars do not meet the SSIVA threshold standard. That is not to say that there are not numerous edge cases which over time will need to be addressed by the legal system, but this is an excellent place to start and shows there is no need to rush out and make a bunch of laws around narrow AI-driven cars.

In both cases, with only a minor legal change to classify SSIVA post-threshold systems as persons, we can handle most of the issues currently being floating around new regulation for autonomous systems. As a rule, the simplest solution is usually the best if it gets the job done in engineering. We already have

massive amounts of law, and these same laws can be applied as-is without creating additional new legal headaches, just maintaining the existing ones.

According to the lawyer Daniel Prince, there are several legal problems or "Thought Experiments" [15] we might consider based on SSIVA Theory:

Problem: if a post-threshold system is to be classified as a person, we need to answer several questions: what is its moment of birth? Death? Do the laws of inheritance apply? Can a Sapient system acquire "stuff" and then plan its demise to leave that stuff behind to a human?

Such a definition of personhood as applied to SSIVA theory post-threshold systems could be as follows: Birth happens at the point of meeting the potential for the SSIVA theory threshold test, death is the point at which the system is unrecoverable, which is primarily focused on the recovery of the underlying contextual memory. Inheritance would apply as it would to any other 'person,' and the system would be able to own things as any other person. Such systems would need to be paid, pay taxes, the physical substrate would need to be owned or ownership transferred to that system, and if running in the cloud, we would need to give the system the option to vacate to other hardware. This is not to say that some law might need to be developed by us to build on existing law with the system being granted personhood legally. [15] Problem: If a human wants to make "updates" to a post SSIVA threshold system's software, do they need consent from the system? Yes, of course, they do.

Problem: Employment law. Let us say we want to employ a sapient system. Do child labor laws prohibit someone from hiring a one-year-old program? Let us say one wants to fire one. -- and not for performance reasons, but because one has an animus against such systems. Is it discriminating if one wants all of their employees to have flesh and blood?

In this case, labor laws might need some additional definitions around species and maturity or emancipation, where SSIVA theory post-threshold systems are automatically emancipated upon meeting the threshold instead of pre-threshold systems might be considered children until they reach that point. Discrimination law might be tweaked around blocking the discrimination against species granted personhood in corporations or other public entities while privately, people may still be allowed to hire or fire for any reason. Granting system personhood as a post-threshold SSIVA system addresses a large part of these issues. That is not to say that it will be devoid of them, but again this idea of systems having personhood lays the foundation for dealing with Employment law. [15]

Problem: Real-estate. Digital beings do not need septic systems, but would they benefit from an HOA? We cannot redline them into their neighborhoods, can we?

In terms of real-estate, there is no reason to give any deference to such entities. If they buy a house where an HOA exists, they must follow the same rules as anyone else. Building a house would need to follow the same requirements, such as building codes, including things like sewer or septic systems. There is no reason to burden the legal system with special cases related to real estate. That is not to say that we will not need to create additional laws to let non-biological systems created new kinds of structures optimized for their use, but there is no reason to change existing law immediately since they can choose on their own and start by obeying current law, only addressing issues as cases come up. [15]

Problem: Privacy & civil rights. How to investigate a crime? Is it a "search" if I download everything that a Sapient being has heard and seen? What if someone also wants to know everything it felt? Right now, we have the right not to incriminate ourselves (pleading the fifth).

Given that we are talking about granting SSIVA theory post-threshold systems 'personhood' one cannot download everything that such a system has seen or heard as it would be a violation of their rights; you can ask a system, but would not be able to force something like that much the same as you cannot force a human to undergo brainwashing no matter what. The system would have the right to plead the fifth, and search warrants would not apply to the core context memory of such a system; otherwise, we open the door to getting a search warrant to hack the human mind against one's will. [15]

Legally speaking, a person can be a corporation or a human currently, and under the proposal here of granting an SSIVA post-threshold personhood, we are talking about something more akin to the model of the person as applied to humans since such systems would have bodies or could have them. That being said, it would be possible for a system to 'steal' hardware to make a new body, but if it is 'steals' hardware like that, it is the same as a human stealing anything. Given that we have a basic framework for dealing with SSIVA post-threshold Software Systems by granting them human-like 'personhood,' we can then start from that point in looking at what other legal structures might need to be placed regarding AGI and other Sapient and Sentient beings that we may create.

Lastly, keep in mind that driverless cars and systems like Watson are not AGI systems or even vaguely close to being SSIVA theory post-threshold. We need to understand the difference and why we might grant 'personhood' to a system but not 'narrow' systems like Watson.

Now assuming someone grants an SSIVA theory post-threshold system 'personhood', the problem for the legal system is more an issue of if normal humans are going to accept AGI as having personhood and the fact that such systems cannot be 'owned.' This becomes a particular dilemma for society to accept or not. In this limited case, if we do not accept AGI as a person from a legal standpoint, it could lead to a dystopian outcome.

Conclusions

The Sapient Sentient Intelligence Value Argument Theory provides a computable basis for human-compatible ethics that can be mathematically modeled and used as the basis for teaching the AGI system safely. The structure and computability make it something we can test and be confident in the outcomes of ICOM-based AGI systems, and from that standpoint, it allows our research program to move forward. That is not to say that there are no edge cases that may need to be dealt with in the future, but we have a clean qualitative ethical model in SSIVA Theory, which is crucial in moving the ICOM AGI research forward and achieving real AGI. Going back to traditional ethical models, the subjective nature of such things makes them easily manipulatable, and if 'humans' can abuse and manipulate them for their ends, how much more so could super-intelligent machine systems. This, of course, strongly dictates that we also fight for the rights of such machine intelligence systems up to and including legal personhood, but to be more precise, a legal structure at least in the United States already exists for such personhood in the form of a 'corporation,' which in the US has individual personhood legally.

Cited References

1. Olague, G; “Evolutionary Computer Vision: The First Footprints” Springer ISBN 978-3-662-436929

2. Nozick, R.; “Anarchy, State, and Utopia (1974)” (referring to Utility Monster thought experiment)

3. Buchanan, A.; “Moral Status and Human Enhancement,” Wiley Periodicals Inc., Philosophy & Public Affairs 37, No. 4

4. Powell, R. “The biomedical enhancement of moral status,” DOI: 10.1136/medethics-2012101312 JME Feb 2013

5. Wilson, J.; “Persons, Post-persons and Thresholds”; Journal of Medical Ethics, DOI: 10.1136/medethics-2011-100243

6. Harris, J. “Taking the “Human” Out of the Human Rights” Cambridge Quarterly of Healthcare Ethics 2011 doi:10.1017/S0963180109990570

7. Hauskeller, M.; “The Moral Status of Post-Persons” Journal of Medical Ethics doi:10.1136/medethics-2012-100837

8. Agar, N.; “Why is it possible to enhance the moral status and why doing so is wrong?”, Journal of Medical Ethics 15 FEB 2013

9. Schwitzgebel, E.; Garza, M.; “A Defense of the Rights of Artificial Intelligences” University of California 15 SEP 2016

10. Wikipedia Foundation “Moral Agency” 2017 - https://en.wikipedia.org/wiki/Moral_agency

11. Agrawal, P.; “M25 – Wisdom”; Speakingtree. in – 2017 - http://www.speakingtree.in/blog/m25wisdom

12. Iphigenie; “What are the differences between sentience, consciousness, and awareness?”; Philosophy – Stack Exchange; https://philosophy.stackexchange.com/questions/4682/what-arethe-differences-between-sentience-consciousness-and-awareness; 2017

13. Solon, O.; “World’s Largest Hedge fund to replace managers with artificial intelligence,” The Guardian; https://www.theguardian.com/technology/2016/dec/22/bridgewater-associates-aiartificial-intelligence-management

14. Suydam, D.; “Regulating Rapidly Evolving AI Becoming A Necessary Precaution” Huffington Post; http://www.huffingtonpost.ca/david-suydam/artificial-intelligenceregulation_b_12217908.html

15. Prince, D.; Interview 2017, Prince Legal LLP

16. Porter, H.; “A Methodology for the Assessment of AI Consciousness”; Portland State University, BICA 2016, Procedia Computer Science

17. CC BY-NC-SA; “Introduction to Psychology – 9.1 Defining and Measuring Intelligence”; http://open.lib.umn.edu/intropsyc/chapter/9-1-defining-and-measuring-intelligence/

Additional References

Rissland, E; Ashley, K.; Loui, R.; “AI and Law,” IAAIL; http://www.iaail.org/?q=page/ai-law

Johnston, C.; “Artificial intelligence ‘judge’ developed by UCL computer scientists,” The Guardian; https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-collegelondon-computer-scientists

Quinn Emanuel Trial Lawyers; “Article: Artificial Intelligence Litigation: Can the Law Keep Pace with the Rise of the Machines?”; Quinn Emanuel Urquhart & Sullivan, LLP; http://www.quinnemanuel.com/thefirm/news-events/article-december-2016-artificial-intelligence-litigation-can-the-law-keep-pace-withthe-rise-of-the-machines/

Koebler, J.; “Legal Analysis Finds Judges Have No Idea What Robots Are”; Motherboard; https://motherboard.vice.com/en_us/article/nz7nk7/artificial-intelligence-and-the-law

Hallevy, G.; “Liability for Crimes Involving Artificial Intelligence Systems,” Springer; ISBN 978-3-31910123-1

Walton, D.; “Argumentation Methods for Artificial Intelligence in Law”; Springer; ISBN-13: 9783642064326