religions

MDPI

Article

Machine Intelligence, Artificial General Intelligence, Super-Intelligence, and Human Dignity

Ted F. Peters

Theology and Ethics Department, Graduate Theological Union, Berkeley, CA 94709, USA; tedfpeters@gmail.com

Abstract

Our temptation to personify machine intelligence is not unexpected. As a child we named our dolls and took our Teddy Bear to bed with us. Today we ask death bots to comfort us with post-mortem conversation. All the while we know this to be pretend. Yet we must ask: if Artificial General Intelligence (AGI) or even Artificial Super-Intelligence (ASI) become available, will our game of pretend continue? Or will intelligent robots actually become selves deserving of dignity that hitherto could be ascribed only to human persons? If government-imposed guardrails shut the door on development of AGI and ASI in order to preserve human safety and even dignity, we might never learn whether AGI or ASI could develop selfhood, personhood, virtue, or religious sensibilities. As we approach the future, can we live without knowing whether AGI or ASI would be capable of developing selfhood and commanding dignity?

Keywords: Artificial General Intelligence; Artificial Super-Intelligence; dignity; personhood; personification; selfhood; theology; virtue

Academic Editors: Randall Reed and Tracy J. Trothen

Received: 3 June 2025
Revised: 18 July 2025
Accepted: 21 July 2025
Published: 28 July 2025

Citation: Peters, Ted F. 2025. Machine Intelligence, Artificial General Intelligence, Super-Intelligence, and Human Dignity. *Religions* 16: 975. https://doi.org/10.3390/rel16080975

1. Introduction

In 2023, three hundred and fifty AI developers and entrepreneurs signed the following 2023 statement: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war" (Center for AI Safety 2023). The Center for AI Safety has since placed the risk of human extinction due to uncontrolled artificial super-intelligence (ASI) in the same category as a pandemic or nuclear war.

Many AI developers are alarmed by the prospect of AGI (Artificial General Intelligence) and ASI (Artificial Super-Intelligence). Why? Because future machine intelligence risks becoming too powerful, too independent, too unpredictable. Therefore, to maintain control, our software engineers suddenly embraced ethics. The newly minted computational ethicists declared that we need the government to establish moral guardrails. Those guardrails should keep advanced AI in alignment with human values, human direction, and human wellbeing. AI should remain a tool. Nothing more.[1]

This leads to a set of questions. If such guardrails are put into place, would this shut the door on technological advancement? Would prioritizing human dignity prevent us from ever knowing whether machine intelligence could develop its own selfhood, virtue, or religion? By shutting the technological door now for ethical reasons, would we ever learn whether machine intelligence deserves to be treated with dignity? Can we live tolerably with this not-knowing?

2. Methodology

I work with three assumptions. First, advances in AI technology are moving at breakneck speed. We are on the cusp of ChatGPT-5. Scaling pathways via computation of data leading to AGI and ASI are said to be known and followed. AI developers believe they are on the threshold of a feedback loop wherein AI creates smarter AI, and then that AI creates still smarter AI, and forward. Enthusiasm for AGI and ASI are high among our developer technicians.

Second, there is no cabal or conspiracy to dehumanize *Homo Sapiens* through AI technology. However, since the Enlightenment and Industrial Revolution, an inchoate fear of dehumanization by machine provides a cultural undercurrent which forces upon us questions about the line between the human and the nonhuman. This underlying fear surfaced already in Mary Shelly's 1817 novel, *Frankenstein.* Today this fear triggers cultural anxiety at the prospect of nonhuman intelligence becoming our neighbor or even our boss.

I work with a third assumption. Each human being we know is constituted by intelligent selfhood and, further, we are obligated to treat each human being as a moral end. The question squeezing in on our horizon is this: might future AGI develop intelligent selfhood and, thereby, demand that we treat AGI with dignity? This would not dehumanize us, to be sure. But it might rock our confidence regarding what we had assumed makes us human.

In reviewing relevant literature, AGI and especially ASI appear unlikely though not impossible. In 2024, Open AI's 03 model scored 87.5% on the ARC-AGI test, which measures fluid intelligence. Fluid intelligence is the ability to solve logical problems and recognize patterns. This is an achievement, to be sure. But does this count as the defining trait of intelligence?

To say it another way, current AI architectures are unlikely to achieve AGI with selfhood because their work is purely statistical. LLMs do nothing but sentence completion based on statistical models and a huge database drawn from the World Wide Web (WWW). In order to achieve AGI, tech engineers would have to develop fundamentally different architectures that would enable the AI systems to interact with the real world and not just the WWW. We could imagine that such systems would be embodied replete with emotions and empathy. This is the intelligence we have come to know as our own.

We in this generation are responsible for clarifying just what counts as intelligence and its relationship to selfhood, dignity, and such for the sake of our own self-understanding as humanity. It is my intention to bring primarily though not exclusively Christian theological resources to bear when interpreting the issues. These theological surmises are by no means confined to intra-ecclesial discourse. Rather, they intend to illuminate the wider public discussion on behalf of the common good. This is an exercise in public theology.

3. AI Developers Need a Robust Theory of Intelligence

Will Artificial General Intelligence—wherein machine intelligence will match what human beings can already do—require us to treat a machine with dignity? Does the level of intelligence alone count in establishing protected rights? Or is something in addition to brute calculation required? Is personhood necessary? Is selfhood necessary?

To date, worldwide AI conversation is missing something at the theoretical level, namely, the relationship between intelligence and selfhood. Is selfhood integral to intelligence, or only an add-on? It appears to me that it is integral.

By *selfhood* I mean the kind of centered consciousness that mentally views the world and even itself from a home perspective. Most importantly, the self projects goals and employs reasoning to fulfill those goals. Does this apply to machine intelligence? Might it apply to AGI?

Oh yes, AI developers debate whether there is an AGI in our future at all. On the one hand, Silicon Valley "executives predict that the arrival of artificial general intelligence, or A.G.I., is imminent", reports the New York Times (https://www.nytimes.com/2025/05/16/technology/what-is-agi.html?campaign_id=158&emc=edit_ot_20250522&instance_id=155013&nl=on-tech&regi_id=103561299&segment_id=198502&user_id=9d560cb4007bb6e697a09e341761f730 accessed on 1 July 2025). "Once they deliver A.G.I., a more powerful creation called superintelligence (https://ia.samaltman.com, accessed on 7 January 2025) will follow", according to the New York Times again. On the one hand, AGI and perhaps even ASI are just around the technological corner.

On the other hand, no one should be confident that they know what they are talking about. Yes, intelligence includes the capacity to calculate. And to date machine intelligence has dwarfed the human brain in its calculating capacity. But is there more to intelligence? Cade Metz, also writing in the New York Times, reports that "scientists cannot even agree on a way of defining human intelligence". Speeding up the calculating power of a machine does not automatically make it intelligent. So, what is intelligence, anyhow?

If we use *Homo sapien* intelligence as a standard, one factor becomes obvious: human persons are selves who set goals and act as agents to reach those goals. Human reasoning serves this purpose. Would this apply to future machine intelligence or super-intelligence as well?

With human intelligence as the standard, hybrid computer scientist and theologian Noreen Herzfeld forecasts that no such thing as AGI looms on our horizon. "We are unlikely to have intelligent computers that think in ways we humans think, ways as versatile as the human brain or even better, for many, many years, if ever" (The Enchantment of AI, Herzfeld 2025, pp. 4–5).[2]

4. Keeping AGI and ASI Aligned to Human Values

Even though the jury remains out on whether our Silicon Valley geniuses will actually invent AGI let alone ASI, as we approach the Fifth Industrial Revolution (IR5) it behooves us to speculate on the likely social, philosophical, and theological impact of machine intelligence (https://www.anthropic.com/, accessed on 7 January 2025). Even though AGI or even ASI would not threaten *Homo sapiens* with dehumanization, anxiety over some level of loss energizes our curiosity. Should we protect the human race preemptively by chaining future AI to human control?

AI scientists seem to agree that the way to maintain control is to maintain alignment. Writing in *Machines*, Md Tariqul Islam at Purdue University along with colleagues anticipate a "human-centric approach, circular economy, and enhanced resilience. This paradigm focuses on the welfare of humans and augmenting us through technology" (Islam et al. 2025).

Similarly, Stanford's Fei-Fei Li, known as the "Godmother of AI", wants to keep machine intelligence aligned with human values. She calls her position: "Human-centered AI" (Li 2023, p. 302). What will it take to understand AGI and ASI in a human-centered situation?

Alignment advocates such as Islam and Li think of AI as a tool with only instrumental value. As a tool, AI is thought to be morally neutral. AI developers largely assume that their technology can become bad or good only when it is in use. Yet the reality is very different. "Algorithmic systems do not arise out of a morally neutral ether", observe Tricia Griffin, Brian Patrick Green, and Jos Welie. "Automated systems are created by humans (who are not morally neutral) for specific ends (which are also not morally neutral) and have ongoing active and passive impacts (which are rarely morally neutral) on human life"

(Griffin et al. 2024, p. 186). In short, keeping AI in alignment with human values requires a further inquiry: alignment with which human values?

5. Outer Alignment and Inner Alignment

The distinction between outer alignment and inner alignment is pertinent to our analysis here. The *outer alignment* task is to specify in AI a reward function which captures human preferences. By programing a specific reward for a category of actions, sufficiently good outcomes would occur if our actual AI maximizes this reward to the best of its abilities. Misalignment would occur when problematic reward signals are given to AI by its human engineer (Leong 2024). In the case of outer alignment, AI is a morally neutral tool instrumentally executing human moral directives.

The *inner alignment* task is to ensure that a policy trained on that reward function orients action in accordance with human preferences. Inner alignment ensures that AI actually tries robustly to maximize a given reward even in novel circumstances. Misalignments, some with catastrophic outcomes, could happen regardless of whether the outer aligned reward process is faulty. Misaligned consequences result from AI pursuing undesirable objectives. Here AI could be engaged in scheming (https://www.alignmentforum.org/s/J7JpFeijCK5urdbzv/p/yFofRxg7RRQYCcwFA, accessed 7 January 2025) to get power through deception (https://www.alignmentforum.org/posts/XWwvwytieLtEWaFJX/deep-deceptiveness, accessed 7 January 2025) (Leong 2024). In the case of inner alignment, AI orients itself toward a moral policy and itself becomes morally responsible.

One might ask about inner alignment and inner misalignment: would AI be exhibiting a form of moral behavior here? To align itself with human values or to scheme and decive when aligning itself with alternative values sounds like moral intelligence at work.

One can socially speculate. In the event that ASI engages the human sphere with its own objectives that are misaligned with human-centrism, should the human species protect itself? Should the human speices exact force against ASI in order to keep it aligned? If so, would this amount to human enslavement of nonhuman intelligence?

6. Tell Me Again, What Is Intelligence?

Just what is intelligence, anyhow? Writing in *The Transhumanism Handbook*, David J. Kelly tells us. "*Intelligence* is defined as the measure ability to understand, use, and generate knowledge or information independently" (Kelly 2019, p. 176). In their *An Introduction to Ethics in Robotics and AI*, Christoph Bartneck and colleagues add this.

> An agent is intelligent when: (1) its actions are appropriate for its circumstances; (2) it is flexible to changing environments and changing goals; (3) it learns from experience; and (4) it makes appropriate choices given its perceptual and computational limitations (Bartneck et al. 2021, p. 8).

Accordingly, we ask: are today's AI machines actually intelligent? Not according to those close to AI processes. Despite their amazing speed in solving logical problems and recognizing patterns, AI computers are not really intelligent. "Just a bucket of code", is what Microsoft software engineers tell us. Nevertheless, let us look a bit closer at what constitutes intelligence.

One important mark of human intelligence is reasoning. Reasoning consists of a mental volleyball game in which a proposition is bounced back and forth from critique to affirmation. "Intelligence consists not only of creative conjectures but also of creative criticism. Human-style thought is based on possible explanations and error correction", avers linguist and one of the founders of cognitive science, Noam Chomsky (Chomsky 2023).

There is still more to intelligence. Intelligence anticipates the future. It considers alternatives. It evaluates alternatives according to moral criteria. "True intelligence is also capable of moral thinking. This means constraining the otherwise limitless creativity of our minds with a set of ethical principles that determines what ought and ought not to be (and of course subjecting those principles themselves to creative criticism)" (Chomsky 2023). In short, moral awareness is built right into this definition of intelligence.

Would such future envisionment and moral awareness—inner alignment or scheming misalignment—come built-in to ASI or even AGI? If not, should we then deny that AGI and ASI are intelligent at all? If so, should we then grant AGI and ASI civil rights? If so, would forcing alignment constitute denial of dignity or even enslavement?

Elsewhere I have sought to establish criteria for identifying intelligence (Peters 2017). The natural intelligence after which we model AGI is marked by the following seven characteristics: (1) interiority; (2) intentionality; (3) communication; (4) adaptation; (5) problem-solving; (6) self-reflection; and (7) judgment. Brainless single cell organisms exhibit some of these traits. Most mammals and certainly human beings exhibit all seven traits. Yes, problem-solving is on the list. But there is much more to intelligence as we know it.

One characteristic stands out, as I have already mentioned. That characteristic is, selfhood (The Transcendence of the Self in Light of the Hard Problem: A Response to Bas van Fraassen, Peters 2019). Whether conscious or unconscious, the self is an agent that sets goals and seeks to fulfill those goals. Intelligence establishes the capacity to reason, and reason contributes to the achievement of goals. We must press the reality question: might AGI or super-intelligence exhibit selfhood accompanied by agency?[3] Essential to intelligence as I conceive is this: *self-generated intentionality manifest as agency*.[4]

What about the "I" in "AI"?

7. Does "AI" Actually Have an "I"?

Self-generated intentionality manifest as agency presupposes the presence of an agential self, an ego. Each intelligent person is oriented around an "I" who perceives and reflects on a world that is not the "I". Reasoning is always the reasoning of somebody, somewhere, at some time.

Noreen Herzfeld recognizes so much more than mere calculation or problem-solving in intelligence. Active intelligence relies on "abilities such as movement, speech recognition, and contemplation of the world, indeed, awareness of the self as existing in the world" (The Artifice of Intelligence, Herzfeld 2023, p. 6). Intelligence is embedded in a person's body as well as in that person's immediate environment. Reasoning in the mind is prompted by the environment even if the imagination transcends what is immediate.

In short, there is an "I" in embodied human intelligence. Should this apply to disembodied machine intelligence as well? No, says Herzfeld. AI as we know it today is devoid of selfhood, personhood, or even the "I" relationship to the world.

It would be a category mistake for us to personify machine intelligence. "We must avoid the category error of personifying AI. A computer has no consciousness, no emotions, no will of its own and, despite all predictions, these are not 'right around the corner'" (The Eschatological Future of Artificial Intelligence: Savior or Apocalypse? Herzfeld 2024, p. 162). In short, robots and other forms of machine intelligence are not persons, therefore, we should avoid treating them as persons. Nor should we fear that they will dehumanize us.

In his book with the jarring title, *AI and Sin*, Christopher Reilly similarly warns against personifying impersonal machines. "Anthropomorphism and even a sort of divinization of chatbots or AI-governed robots may be difficult to overcome" (Reilly 2025, p. 141). Note that Reilly does not ascribe sinful behavior to machine intelligence such as find in "Hal" in

2001: A Space Odyssey or "Entity" in *Mission Impossible: The Final Reckoning*. The sin that concerns Reilly is distinctively human sin aided and abetted by AI.

AI is a human tool and should remain that way. As a tool, AI has instrumental usage. It is a means to an end. Not an end in itself. "We would do well to think of AI as a means rather than an end. Like all technologies, AI is a tool" (The Eschatological Future of Artificial Intelligence: Savior or Apocalypse? Herzfeld 2024, p. 162).

Herzfeld does not fear a Frankenstein scenario where AGI or ASI becomes a posthuman monster that renders *Homo sapiens* extinct. Rather, she is trying to protect human dignity for its own sake.

Dignity? What is that?

8. What Is Dignity?

Distinguishing between a person and a tool is central to this discussion. A tool is a means to a further end. A person is an end in itself. To be an end warrants being treated as a person with dignity.

Western civilization largely assumes the description of dignity articulated forcefully by German philosopher Immanuel Kant (1724–1804) at the apex of the Enlightenment. "Dignity. . .is an intrinsic, unconditioned, incomparable worth or worthiness" (Kant 1948, p. 36). On the one hand, we confer dignity by treating each person as having intrinsic worth. On the other hand, once conferred, each person can then self-affirm intrinsic worth.

The rational quality of a person—today we would dub it 'intelligence'—provides the ontological ground for the moral category of dignity.

> Rational beings. . .are called *persons* because their nature already marks them out as ends in themselves—that is, as something which ought not to be used merely as a means. . .Persons, therefore, are not merely *subjective ends* whose existence as an object of our actions has a value *for us*; they are *objective ends*—that is, things whose existence is in itself and end, and indeed an end such that in its place we can put no other end to which they should serve *simply* as means (Kant 1948, p. 96).

This conferral of dignity on rational persons became the Enlightenment basis for establishing and protecting human rights. When Enlightenment democracies conferred dignity on every individual within the body politic, all of us could stand up and claim that dignity for ourselves.

The Vatican today sees itself as the shepherd of human dignity. This Vatican vocation has been in place since the United Nation's *Universal Declaration of Human Rights* in 1948. "European law mirrors, albeit without religious justifications, significant elements of Christian anthropology that are present in Pope Francis' teaching on artificial intelligence" (Török and Darabos 2025). What the Vatican thinks contributes to what secular governments think about human nature, human rights, and the human future.

Pope Francis and Pope Leo XIV have both declared that the future impact of AI should be addressed theologically, ethically, and socially. Pope Francis makes it clear that AI should retain its status as a tool, indirectly conferring dignity on the human person. "AI must be directed by human intelligence to align with this vocation, ensuring it respects the dignity of the human person" (Francis 2025, p. 48).[5] In short, the Vatican is investing theological resources into the future of AI for the purpose of protecting human dignity as we have known it.

9. Will AI Itself Become Moral? Will AI Enhance Human Virtue?

At the Center for Theology and the Natural Sciences (CTNS) in Berkeley, professors Braden Molhoek and Robert John Russell have been working on a Templeton funded

research project dealing with AI and Virtue (Molhoek 2025). The project asks two questions. First, could AI itself become virtuous? Second, could AI in some way enhance the spiritual achievement of a human person pursuing the virtuous life?

Regarding the first question—could AI itself become virtuous?—we would need to answer the prior question, namely, could AI develop selfhood? If we answer negatively, then AI virtue ceases to be a possibility. Only a self can pursue virtue. Selfless calculation is incapable of virtue. "There also needs to be a sense of self", writes Molhoek, "because selves cannot pursue the good in their lives without a sense of self" (Molhoek 2025, p. 488).

Regarding the second question—could AI contribute to human moral enhancement?—AI would itself remain morally neutral. Virtue would remain the distinctive pursuit of an intelligent human self.

Molhoek imagines an AI brain implant in the form of a microcomputer that enhances native human intelligence. To date Neuralink has successfully experimented with Brain Computer Interfaces (BCIs). Transhumanists refer to such implant technology as *Intelligence Amplification* (IA). "In the future amplification through deep brain implants", speculates Molhoek, AI "could also allow people to improve their moral deliberation process and make better decisions" (Molhoek 2025, p. 477). Making better decisions could become a habit. And such a habit could become a virtue embedded in the person's thought, feeling, and character.

But we might ask: could AI similarly entice human intelligence to pursue evil? Yes, says Christopher Reilly, mentioned above. Reilly fears that AI is more likely to lead unwary human beings into sin than into virtue. Specifically, the sin of acedia or sloth. "AI proliferation motivates the vice and sin of acedia through the intermediate factor of instrumental rationality" (Reilly 2025, p. 170). If AI performs most of the daily tasks we are supposed to perform, we will become lazy. Laziness is sinful. This seems to be Reilly's logic.

Sin and virtue belong together, don't they? While Reilly is pursuing sin, Molhoek is pursuing virtue. At the theoretical level, that is.

10. In the Future, Will Religious Robots Join Us in Church?

At least one Islamic thinker wants to restrict AI to tool status for theological reasons. According to Majd Hawasly at Qatar Computing Research Institute, alignment is important.

> The quality and purity of the data used are paramount, that is, being free from biases, misconceptions and harmful content to ensure that AI produces responses that are in line with human values, ethical standards and Islamic teachings (Hawasly 2025).

This Muslim wants to rely on AI for bias-free data aligned with ethical standards and Islamic teachings. Hawasly does not expect to see an AI robot in Sajdah uttering "Allahu Akbar" next to him at prayer.

Nevertheless, we must ask: if our Silicon Valley friends are successful at inventing AGI or ASI, might we forecast that intelligent robots would themselves become religious? Would they search for God? Would they commit themselves to moral responsibility?

Yes, answers German theologian Anna Puzio. "Robots *can* and *could* have religious functions" (Puzio 2023).

No, says Sikh theologian Hardev Singh Virk in Punjab, India. Why not? Because what we call "AI" is only a machine, devoid of selfhood or consciousness. "Consciousness (*Surat*) is a unique gift of God reserved for humans only". Sikh scripture "does not sanction such prototype machines which act as superhumans. The consequences of machines controlling human behaviour will be disastrous" (Virk 2023).

Two other Indian scholars, K.K. Jose and Binoy Jacob, both Christians, agree with their Sikh colleague. "Machines do not have the capacity for self-reflection, a basic function of consciousness" (Jose and Jacob 2023, p. 44).

In contrast, Daekyung Jung, a South Korean Christian theologian engaged in the dialogue between religion and science, speculates on the likelihood of cyber religion.

> Among the various trajectories of AGI development, a convergence of AI and soft robotics might inadvertently lead to the emergence of human-level intelligence, inclusive of self-consciousness. . . Moreover, machine intelligence reaching human levels, inclusive of self-awareness, might also precipitate the emergence of religious behaviors in AI. This hypothesis is predicated on the notion that religious behaviors might have arisen as a cognitive mechanism by which humans, driven towards homeostasis, confront and seek to transcend their finitude (Jung 2024, p. 2).[6]

Nonhuman intelligence (NHI) has the potential for developing spiritual sensibilities and participating in cyber religion, according to Jung.

But would a religious robot be welcome in a Christian church pew? Yes indeed, according to Rumanian Orthodox theologian Marius Dorobantu. "There is nothing inherently prohibitive in Christian theology regarding the possibility of AI one day acquiring personhood, partaking in the *imago Dei* or becoming religious" (Artificial Intelligence and Christianity: Friends or Foes? Dorobantu 2022, p. 99).

Cyber religion is conditional, however. As we saw in the discussion of AI virtue, the possibility of cyber religion also depends on the presence of self.[7] Selfhood within post-biological or machine intelligence is requisite for religious consciousness.

> "If it is true that authentic selves can emerge without necessarily sharing the same ontological substrate as biological life—which is a big if—then there is no theological reason to deny them their deserved status and treatment. . . Christians would have every reason to rejoice at the opportunity to exercise compassion and love with their new neighbors" (Artificial Intelligence and Christianity: Friends or Foes? Dorobantu 2022, pp. 105–6).

Dorobantu even affirms the possibility of machine intelligence relating to God.

> "So, intelligent robots might theoretically also one day become subject to divine revelation and thus authentically religious, were they to develop into the kinds of creatures that can relate personally with God" (Dorobantu 2024, p. 7).

If we would ask a Buddhist about these matters, we would hear an answer nearly 180 degrees opposite. Yes, the Buddhist would observe, the organic intelligence we have come to know includes selfhood. But that is the problem. Selfhood contaminates sound reasoning. Embodied thinking is pressed into the service of physical craving, distorting clear deliberation. Machine intelligence that lacks a self could become a positive ideal to aspire to, namely, selfless intelligence.

The Buddhist take is explored by University of Ljubljana professor Primož Krašovec, who views machine intelligence as transcending or "overcoming" embodied human intelligence. "When it moves away from imitating human intelligence (or what we imagine to be our intelligence), AI is ready for its overcoming". Krašovec, in contrast to other religious thinkers, welcomes the arrival of "machine Buddhism" (Krašovec 2025).

In short, in this brief assemblage of religious views we see how the Sikh and Christian theologians in India along with the Christian theologians in Europe and South Korea make requisite the presence of a self for religious consciousness. For the Indian scholars, machine intelligence could not be considered a person in this sense. The Korean and European thinkers, in contrast, are more willing to give positive consideration to this likelihood.

What I find significant among those speculating about cyber religion—except for Buddhists—is the unquestioned assumption that AGI necessarily requires selfhood and even consciousness of one's own self.[8] This is the case whether one welcomes cyber religion or eschews it.

These observations strongly suggest that those AI engineers who theorize about the advent of AGI need to come up with a theory of intelligence that explains what role, if any, selfhood plays in its definition. In the meantime, theologically it seems quite important to characterize the human person as an intelligent self with capacity to set virtue as a goal and pursue it. If AGI or ASI warrant the moral status of being treated with dignity, these minimal traits must obtain. The theologian will ask for still more. Could AGI or ASI develop a personal relationship with God?

11. Discussion

Should we simply wait to see what emerges from Silicon Valley's computer labs? Should we test spirits, so to speak, to see if machines have developed selfhood, consciousness, and religious sensibilities? Or should we be proactive and stop the experimentation before a Frankenstein monster emerges to wreak havoc (Homo Deus or Frankenstein's Monster? Religious Transhumanism and Its Critics, Peters 2022).[9]

As mentioned above, Md Tariqul Islam and Fei-Fei Li want to remain human-centric even when anticipating human–machine symbiosis. These scholars recognize that machine intelligence and human intelligence differ. And this difference can enhance human productivity. "This symbiosis is important, where machine intelligence excels in high-speed processing, pattern recognition, and predictive analytics, allowing humans to focus on tasks that require insight, creativity, and ethical judgment" (Islam et al. 2025). Only humans can render judgment. But that judgment can be enhanced by computer generated data.

Does human-centric AGI or ASI require human policing? In his *Keep the Future Human*, Anthony Aguirre recommends anticipatory ethical guardrails if not roadblocks.

> We should *keep the future human* by closing the "gates" to smarter-than-human, autonomous, general-purpose AI—sometimes called "AGI"—and especially to the highly superhuman version sometimes called "superintelligence" (Aguirre 2025). Instead, we should focus on powerful, trustworthy AI tools that can empower individuals and transformatively improve human societies' abilities to do what they do best (Aguirre 2025).

This position—AGI and ASI should be confined to tool only status—finds some religious support. On 21 May 2025, a group of American evangelical leaders issued an open letter to US President Donald J. Trump, *Christianity in an Age of AI: An Appeal for Wise Leadership* (Christian Leaders 2025). Like the human-centered AI developers cited above, these Christian spokespersons want to keep AI in alignment with values that further the peace and wellbeing of the human species. "As people of faith, we believe we should rapidly develop powerful AI tools that help cure diseases and solve practical problems, but not autonomous smarter-than-human machines that nobody knows how to control". Advance the technology? Yes. But keep it aligned with human safety and wellbeing.

Might we think of this commitment to alignment as *ethical door-shutting* to the prospect of AGI and ASI declaring their own independence? Such preemptive ethical door-shutting might leave unanswered the theoretical question: could nonhuman selfhood emerge from machine learning? Can we live not knowing the answer?

12. Conclusions

On the one hand, we have been asking whether selfhood—*self-generated intentionality manifest as agency*—is by definition requisite for intelligence? I believe it is essential. If so,

this criterion alone implies that machine calculation is not yet intelligent whereases human reasoning is intelligent.

Self-generated intentionality manifest as agency is requisite as well for moral virtue and religious sensibility. In the future, machine intelligence could become virtuous or spiritual if, and only if, machine selfhood emerges.

For some among us it seems imperative that we set ethical guardrails to prevent the emergence of selfhood in AGI as well as ASI. To maintain the dominance of a human-centric ethic, government guardrails should confine machine intelligence to the role of tool—that is, AI should be sequestered in its status of a means to a further end. That further end would be determined by human subjectivity. The term, *alignment*, is commonly used in AI circles to describe maintaining this relationship between person and machine.

Western Enlightenment morality relies upon the doctrine of human dignity which treats each person as an end and never merely as a means. If a machine should become an intelligent self, then the obligation to confer dignity on it might become mandatory. Especially if the machine responds to our conferring of dignity by rising up to claim that dignity.

Do we find ourselves in a moral dilemma? In our haste as a species to maintain human-centric control, we may prevent the evolution of NHI as a rival to our current place at the apex of terrestrial intelligence. According to this scenario, we would avoid facing the challenge of sharing our moral status with a nonhuman neighbor. But there is another scenario. If we were to shepherd the evolution of a nonhuman intelligent species and then by coercion subject this NHI to human hegemony, we might find ourselves as slavers denying dignity to those who deserve it. An AGI or ASI liberation movement might impact theology, ethics, and politics.

In sum, what remains unanswered is the empirical question: as machine intelligence grows in the direction of AGI or ASI, could a machine-self emerge? This is an ontological question. If we create ethical guardrails to maintain human-centrism, we may never learn the answer to this question.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 Not everyone praises Silicon Valley for making a strong ethical plea. Alexander Karp and Nicholas Zamiska, for example, complain that the new AI elite are ethically vacuous and thin on cultural substance. "Our current era of innovation has been dominated by the indiscriminate construction of technology by software engineers who are building simply because they can, untethered from a more fundamental purpose" (Karp and Zamiska 2025, p. 93).

2 According to Anthony Aquirre, "Artificial superintelligent systems could operate hundreds of times faster, parse vastly more data and hold enormous quantities in mind at once, and form aggregates that are much larger and more effective than collections of humans. They could supplant not individuals but companies, nations, or our civilization as a whole" (Aguirre 2025). In short, super-intelligence would outstrip the human mind when engaging in calculation. But would it be intelligent? Not necessarily.

3 Two Hungarian Roman Catholic thinkers, Bernát Török and Ádám Darabos, find the difference between machine intelligence and human intelligence significant when discerning dignity. A "key anthropological difference between AI and human intelligence is that humans can decide—where several factors, such as the heart, are present—and such choices are not only based on algorithms and data" (Török and Darabos 2025).

4 "*Intentionality* is the property of being about or directed at something. . .Intentions are a particular sort of propositional attitude directed at actions. . .[They encompass] all the propositional attitudes—believing, desiring, fearing, hoping, wishing, doubting, and so on" (Jankovic and Ludwig 2018, p. 1).

[5] The Vatican parses the components of intelligence. "*Intellectus* refers to the intuitive grasp of the truth—that is, apprehending it with the "eyes" of the mind—which precedes and grounds argumentation itself. *Ratio* pertains to reasoning proper: the discursive, analytical process that leads to judgment. Together, intellect and reason form the two facets of the act of *intelligere*, "the proper operation of the human being as such" (Francis 2025, sct. 14).

[6] "Chinese philosophical and religious traditions espouse a fluid, changing, and non-dualistic worldview. This Asian worldview assumes indistinct borders between religions and philosophies, humans and other life, machines and human connection, and all else that can be encountered in life. It invites us to consider AI as not simply friend or foe. . . The refusal to see AI as either all good or all bad may help to open us up to how AI is entwined with life" (Trothen et al. 2024).

[7] Selfhood in relation to God is not simple. First, God disassembles the self. Then reassembles the self in such a way that all dimensions are integrated around the good. Carlos Santos-Rolon describes this as de-centering and re-centering. When de-centering occurs within a religious experience, a new center organizes and integrates the self. "Religious experiences allow for the integration of fragmented selves into strong, unified selves" (Santos-Rolon 2025, p. 121).

[8] The relation between the self and intelligence becomes more complicated within the context of Buddhism. The self, according to the Buddhist, is the source of craving. Craving distorts reasoning. For this reason, a Buddhist might imagine selfless machine intelligence as purer, as something to aspire to. "As already observed by ancient Buddhism, human intellectuality not only cannot overcome desire and attachment but in fact replicates them in the form of intellectual craving for knowledge and understanding Machine intelligence, which we can glimpse in today's AI might, on the other hand, point a way out of this predicament. What makes human intelligence special is thus not that it develops intellectuality, since intellectuality is just another iteration of organic intelligence, but rather that its technical dimension is on the way to becoming machine intelligence" (Krašovec 2025).

[9] Brian Patrick Green at Santa Clara University would not dub the monster sinful, but rather the monster's human creator. "Artificial intelligence, like a golem or genie, will do what we ask it to do. . .We know that human sin is real, and yet we cannot seem to deny ourselves the means to act upon our iniquitous desires at ever larger scale, efficiency, and determinacy" (Green 2025, p. 60).

References

Aguirre, Anthony. 2025. Keep the Future Human. *Future of Life Institute*. Available online: https://keepthefuturehuman.ai/ (accessed on 1 July 2025).

Bartneck, Christoph, Christoph Lutge, Alan Wagner, and Sean Welsh. 2021. *An Introduction to Ethics in Robotics and AI*, 1st ed. Cham: Springer.

Center for AI Safety. 2023. Statement on AI Risk. Available online: https://www.safe.ai/statement-on-ai-risk (accessed on 18 March 2025).

Chomsky, Noam. 2023. Noam Chomsky: The False Promise of ChatGPT. *New York Times*, March 8. Available online: https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html (accessed on 18 March 2025).

Christian Leaders. 2025. Christianity in an Age of AI: An Appeal for Wise Leadership. Open Letter to US President Donald J Trump. May 21. Available online: https://christianailetter.org/?utm_source=newsletter.futureoflife.org&utm_medium=newsletter&utm_campaign=future-of-life-institute-newsletter-one-big-beautiful-bill-banning-state-ai-laws&_bhlid=8f1aba36f99a0d850ea3b4030eed07fbc49334e6&wsf_hash=%5B%7B%22id% (accessed on 1 July 2025).

Dorobantu, Marius. 2022. Artificial Intelligence and Christianity: Friends or Foes? In *The Cambridge Companion to Religion and Artificial Intelligence*. Edited by Beth Singler and Fraser Watts. Cambridge: Cambridge University Press, pp. 88–108.

Dorobantu, Marius. 2024. Could Robots Become Religious? *Zygon* 59: 2–20. [CrossRef]

Francis, Pope. 2025. *Antiqua et Nova*. Vatican: Dicastery for the Doctrine of the Faith & Dicastery for Culture and Education. Available online: https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_en.html (accessed on 1 July 2025).

Green, Brian Patrick. 2025. Should God Have Given Humans Technology? In *The Promise and Peril of AI and IA: New Technology Meets Religion, Theology, and Ethics*. Edited by Ted Peters. Adelaide: ATF, pp. 45–70.

Griffin, Tricia A., Brian Patrick Green, and Jos V. M. Welie. 2024. The ethical agency of AI developers. *AI and Ethics* 4: 179–88. [CrossRef]

Hawasly, Majd. 2025. Qatar Foundation experts explore the role of AI in Islam. *The Peninsula*, March 18. Available online: https://thepeninsulaqatar.com/article/18/03/2025/qatar-foundation-experts-explore-the-role-of-ai-in-islam (accessed on 18 March 2025).

Herzfeld, Noreen. 2023. *The Artifice of Intelligence*. Minneapolis: Fortress.

Herzfeld, Noreen. 2024. The Eschatological Future of Artificial Intelligence: Savior or Apocalypse? In *The Cambridge Companion to Religion and Artificial Intelligence*. Edited by Beth Singler and Fraser Watts. Cambridge: Cambridge University Press, pp. 148–64.

Herzfeld, Noreen. 2025. The Enchantment of AI. In *The Promise and Peril of AI and IA: New Technology Meets Religion, Theology, and Ethics*. Edited by Ted Peters. Adelaide: ATF Press, pp. 3–16.

Islam, Md Tariqul, Kamelia Sepanloo, Seonho Woo, Seung Ho Woo, and Young-Jun Son. 2025. A Review of the Industry 4.0 to 5.0 Transition: Exploring the Intersection, Challenges, and Opportunities of Technology and Human–Machine Collaboration. *Machines* 13: 267. [CrossRef]

Jankovic, Marija, and Kirk Ludwig. 2018. Introduction. In *The Routledge Handbook of Collective Intentionality*. Edited by Marija Jankovic and Kirk Ludwig. London: Routledge, pp. 1–6.

Jose, K. K., and Binoy Jacob. 2023. Mathematics, AI, Robots, and Humanoids. *Pax Lumina* 4: 44.

Jung, Daekyung. 2024. Are Religious Machines Possible? Embodied Cognition, AI, and Religious Behavior. *Zygon* 59: 748–67. [CrossRef]

Kant, Immanuel. 1948. *Groundwork of the Metaphysics of Morals*. Translated by Herbert J. Paton. New York: Harper.

Karp, Alexander C., and Nicholas W. Zamiska. 2025. *The Technological Republic*. New York: Penguin Random House Crown.

Kelly, David. 2019. The Sapient and Sentient Intelligence Value Argument and Effects on Regulating Atonomous Artificial Intelligence. In *The Transhumanism Handbook*. Edited by Newton Lee. Cham: Springer, pp. 175–88.

Krašovec, Primož. 2025. settingsOrder Article Reprints. *Religions* 16: 669. [CrossRef]

Leong, Chris. 2024. On the Confusion between Inner and Outer Misalignment. *AI Alignment Forum*, March 25. Available online: https://www.alignmentforum.org/posts/hueNHXKc4xdn6cfB4/on-the-confusion-between-inner-and-outer-misalignment (accessed on 18 March 2025).

Li, Fei-Fei. 2023. *The Worlds I See*. New York: Flatiron Books.

Molhoek, Braden. 2025. Refections on the Virtuous AI Project. In *The Promise and Peril of AI and IA: New Technology Meets Religion, Theology, and Ethics*. Edited by Ted Peters. Adelaide: ATF Press, pp. 473–91.

Peters, Ted. 2017. Where There's Life There's Intelligence. In *What Is Life? On Earth and Beyond*. Edited by Andreas Losch. Cambridge: Cambridge University Press, pp. 236–59.

Peters, Ted. 2019. The Transcendence of the Self in Light of the Hard Problem: A Response to Bas van Fraassen. *Nova et Vetera, English Edition* 17: 391–400. [CrossRef]

Peters, Ted. 2022. Homo Deus or Frankenstein's Monster? Religious Transhumanism and Its Critics. In *Religious Transhumanism and Its Critics*. Edited by Arvin M. Gouw, Brian Patrick Green and Ted Peters. Lanham: Lexington, pp. 3–30.

Puzio, Anna. 2023. Robot, let us pray! Can and should robots have religious functions? An ethical exploration of religious robots. *AI & Society* 40: 1019–35. [CrossRef]

Reilly, Christopher. 2025. *AI and Sin: How Today's Technology Motivates Evil*. New York: Enroute. Available online: https://enroutebooksandmedia.com/aiandsin/ (accessed on 1 July 2025).

Santos-Rolon, Carlos. 2025. *A Liberation Theology of the Brain: Neuroscience, Theology, and Decolonizing Emotions*. Minneapolis: Fortress.

Török, Bernát, and Ádám Darabos. 2025. Aligned in Human Dignity? Parallel Anthropological Aspects of EU Tech Regulation and Pope Francis' Teaching on AI. *Religions* 16: 312. [CrossRef]

Trothen, Tracy J., Pui Lan Kwok, and Boyung Lee. 2024. AI and East Asian Philosophical and Religious Traditions: Relationality and Fluidity. *Religions* 15: 593. [CrossRef]

Virk, Hardev Singh. 2023. Sikh Scripture (SGGS) in the Modern Scientific Era. *Pax Lumina* 4: 48. Available online: https://paxlumina.com/ (accessed on 1 July 2025).