Journal of Philosophical Theological Research
Print: 1735-9791
Online: 2538-2500

A Critical Evaluation of the Fitting-Recipient Argument for the AI Ensoulment Hypothesis

Reza Dargahifar

Ph.D. in Philosophy of Mind, Assistant Professor, Faculty of Strategic Sciences and Technologies, University of Qom, Qom, Iran dargahifarreza@gmail.com

Abstract

Will advanced artificial intelligence systems one day become ensouled? An affirmative answer to this question, which is an important issue in the philosophy of artificial intelligence, has various metaphysical, ethical, legal, social, and political implications. One of Brian Cutter's two arguments in his article "The AI Ensoulment Hypothesis" in defense of the possibility that the AI ensoulment hypothesis might be realized in artificial general intelligence is the Fitting-Recipient Argument. In this article, after presenting some preliminary remarks on AGI, Cutter's argument is explained together with its underlying assumptions and key concepts, and his reasons for the truth of its two premises are formulated and analyzed. It is then shown that the first premise of the argument, which Cutter treats as expressing a sufficient condition, is in fact defensible only as a necessary condition. Accordingly, the argument commits the fallacy of affirming the consequent, that is, treating a necessary condition as a sufficient one, and therefore fails to establish its conclusion. Consequently, ensoulment cannot be inferred from fittingness for ensoulment, and thus the Fitting-Recipient Argument cannot provide a basis for Cutter's intended conclusion that we should take seriously the possibility of artificial general intelligence becoming ensouled.

Original Research

Keywords:

artificial general intelligence (AGI); AI ensoulment; fitting-recipient argument; substance dualism; philosophy of artificial intelligence; philosophy of mind; affirming the consequent

Received: 2026/02/29; **Received in revised form:** 2026/05/18; **Accepted:** 2026/06/12; **Published online:** 2026/06/18

Dargahifar, R. (2026). A Critical Evaluation of the Fitting-Recipient Argument for AI Ensoulment Hypothesis. *Journal of Philosophical Theological Research*, (Special Issue on AI, Consciousness, and Human Agency), 28(3), 205-232. https://doi.org/10.22091/jptr.2026.17187.3655

Copyright © The Author

Introduction

The first episode of the educational English-language television series *How Do You Do* (a Canadian production from 1988) begins with Frankie Stone, the creator of a humanoid robot, tightening the final screw and then pressing the power switch. The robot opens its eyes and, after seeing Frankie and hearing his greeting, begins to respond with some difficulty and hesitation. When Frankie asks how it is feeling, it says, "I am …" and, without finishing the sentence, asks, "Who am I?" Frankie replies that its name is CHIPS (an acronym for Cybernetic Heuristic Integrated Perception System).

With the emergence of successful AI technologies and the projection of various scenarios concerning their future development, one of the questions that has been and continues to be raised in the philosophy of AI is whether an artificially intelligent system might one day genuinely possess free will, rather than merely simulate voluntary behavior. Could it actually understand and know, and be aware of itself and its environment, rather than merely behave in ways that would justify others in attributing consciousness, understanding, and cognition to it? More fundamentally, could it one day possess a soul or mind? If so, could it be regarded as a living being and as a new species in the natural world? For many people, particularly some AI engineers, such questions are meaningless or pointless, and it seems highly unlikely, if not impossible, that an intelligent system could genuinely possess a soul or mind. Yet some views and theories concerning what it is for something to be alive or to possess a soul entail, at least, the possibility that life, soul, and mind could be realized in certain advanced intelligent systems.

We human beings ordinarily take it for granted that we possess a soul, or at least that we are conscious, and we can also accept that animals – and sometimes even plants – possess a soul or are conscious. By contrast, the idea that a machine could possess a soul or be conscious strikes us as highly implausible; at least for the time being, it is difficult for us to accept, or even regard as possible, such a prospect. In *The Mysterious Flame*, Colin McGinn recounts a dialogue from a novel by Terry Bisson between an alien explorer and his commander, who have come to observe the Earth. The two are astonished that terrestrial beings – human beings – are made of meat and yet are sentient and conscious, and that they have succeeded in constructing machines capable of transmitting signals. For them, it was difficult to accept the possibility that a creature made of meat could be conscious, dream, and love, or that a brain made of meat could think (where "meat" is used as a generic term for the material constituting the human body) (McGinn 1999, 7–8).

Oberg (2023) offers a significant reconceptualization of the "AI ensoulment"

hypothesis by systematically decoupling the theological concept of the soul from the psychological concept of the self. He argues that the soul, as presented in Abrahamic scriptural traditions, is necessarily tied to transcendence and postmortem continuity, whereas the self is fully determinative in purely earthly, material terms and is therefore potentially applicable to artificial systems. Drawing on and revising Kristjánsson's (2010) self-theory, Oberg defines the core Self as the sum of self-defining traits, self-directing and self-evaluating traits, consciousness, and bodily presence (Self = Set 1 + Sets 2 & 3 + C + Bp). He further grounds this model in a two-tiered, materialist account of cognition in which pre-aware, intuitive, and emotional processes dominate, and consciousness is explained as complex neural representation rather than as evidence of anything immaterial. On this basis, he argues that an artificial Self – though almost certainly different from a human Self – is not metaphysically impossible, provided that consciousness can be instantiated non-organically through representational and associative systems analogous to neural "maps" and "tags." Oberg is careful to distinguish this from full personhood, noting that an AI Self might lack the embeddedness and social identity characteristic of human persons, which would reduce but not eliminate the ethical stakes. He concludes by calling for an open and ongoing public dialogue about the moral and legal status of potentially self-aware artificial systems, while explicitly declining to answer whether such systems could meaningfully claim a "soul" for themselves. Oberg does not argue that AI can or should be "ensouled" in any religious or metaphysical sense. Rather, he systematically separates the concept of "soul" from the concept of "self" and then asks whether an artificial self (not soul) could be constructed – a condition he calls being "en-solved" (a deliberate neologism, though the pun on "ensouled" is clearly intended). His conclusion is cautiously affirmative: an artificial Self is conceivable if consciousness can be understood materialistically, but everything hinges on how we define consciousness and whether it can arise non-organically.

Tiwari and Tomar (2026) advance a categorical rejection of the AI ensoulment hypothesis, arguing that artificial intelligence can never possess a soul or consciousness – not merely as a technological limitation but as a philosophical and metaphysical impossibility. Drawing on both Western and Indian philosophical traditions, they contend that AI remains confined to language and computation, whereas human consciousness is grounded in lived, embodied, experiential subjectivity that cannot be simulated or engineered. From the Western canon, they marshal Descartes' dualism, Husserl's intentionality, Nagel's subjective character of experience, Chalmers' hard problem, and Searle's Chinese Room to demonstrate that consciousness entails features – self-reflection, intentionality, qualia, semantic understanding – that resist reduction to computation. Indian traditions, they argue, are even more radical in treating consciousness as ontologically prior to matter: Advaita

Vedanta identifies ātman with brahman; Sāṃkhya distinguishes puruṣa from prakṛti; Jainism separates jīva from ajīva; Kashmir Śaivism emphasizes vimarśa (reflexive self-awareness); and Bhartṛhari grounds meaning in śabda-brahman and sphoṭa. Pairing Wittgenstein's language games with Bhartṛhari's holistic semantics, the authors conclude that AI processes linguistic fragments without inhabiting the forms of life or experiential depth that give language meaning. They critically rebut functionalism, transhumanism, emergentism (including Integrated Information Theory), and ethical arguments for AI personhood, maintaining that simulation is never equivalent to subjectivity. Their conclusion is unequivocal: AI's boundaries are linguistic and computational, while human consciousness transcends language through lived experience, making the notion of a soulful AI both metaphysically impossible and philosophically incoherent. This position stands in direct contrast to Oberg's (2023) materialist account, which, while also rejecting the theological soul, argues that an artificial Self is conceptually achievable if consciousness is understood as materially instantiable representational processing. The tension between these two positions – Tiwari et al.'s ontological impossibility thesis and Oberg's cautious materialist optimism – frames the central debate in the literature on AI consciousness and soul.

Swann (2021) approaches the question of AI ensoulment from a Christian exegetical and theological perspective, arguing that AI cannot possess a soul in the biblical sense. Focusing on the Hebrew concept of *nephesh* (soul) as developed in Genesis, Swann contends that the human soul is not a detachable, immaterial substance but the conjunction of a physical body and the God-given breath of life, resulting in a unified, integrated being who is a soul rather than merely having one. The human *nephesh* is further distinguished by its divine mandate to serve as the Image of God and curate creation – a functional identity that AI, as a human tool, lacks entirely. Swann argues that a digital consciousness is inherently non-corporeal and dichotomous (program plus housing), and therefore cannot constitute the integrated body–breath unity that defines a biblical soul. He further notes that sentience does not entail possession of a soul, since spiritual beings possess consciousness but are not considered souls. While Swann concludes that AI cannot be ensouled, he reframes the ethical question: the real theological concern is not whether AI has a soul but whether human souls are prepared to take responsibility for their autonomous creations. AI, unlike a nuclear weapon, has determinative power, and the burden of moral accountability falls entirely on its human creators. This position stands as a distinct theological-exegetical contribution to the debate, contrasting with Oberg's (2023) materialist account of the Self and Tiwari et al.'s (2025) philosophical impossibility thesis, and together these three perspectives illustrate the range of contemporary positions on AI consciousness and soul.

In an article entitled "The AI Ensoulment Hypothesis," Brian Cutter argues

that some future AI systems will acquire an immaterial soul (Cutter 2025, p. 1). He proposes the Artificial General Intelligence Ensoulment Hypothesis and offers two arguments in support of it.

In what follows, after making some preliminary remarks about artificial general intelligence, I will first explain the AI Ensoulment Hypothesis as formulated by Cutter and then reconstruct the Fitting-Recipient Argument by distinguishing its premises and explaining and defending them. Finally, I will criticize the argument by rejecting its first premise. It is worth noting that Bálint Békefi (2026) has also criticized this argument, but he raises an objection to its second premise. Thus, my strategy for criticizing the argument differs from Békefi's.

It is also worth noting that Tijanovic's "The Virtual Ensoulment Hypothesis: Can Virtual Individuals Be Ensouled?" (forthcoming) extends Brian Cutter's (2025) AI ensoulment hypothesis from artificial intelligence to individuals inhabiting virtual worlds. Working explicitly within a substance-dualist framework, Tijanovic argues by analogy that if human beings can be ensouled, virtual individuals may likewise be ensouled insofar as they are functionally equivalent to real individuals in three respects emphasized in Cutter's framework: access to environmental inputs, a sufficiently complex behavioral repertoire, and inferential capacities partly grounded in memory. Tijanovic then introduces a novel objection, the "argument from value," according to which virtual worlds may lack values necessary for ensoulment – causal stability, truthfulness, and structural stability. He argues that the first two do not ultimately preclude virtual ensoulment, whereas the structural instability of virtual worlds constitutes a more serious challenge, since their dependence on computers makes their continued existence vulnerable to external interruption or termination. Thus, while Cutter primarily argues for the possibility of ensoulment in AI on the basis of functional capacities, Tijanovic extends this functionalist strategy to virtual individuals and tests its limits by introducing environmental and axiological conditions that may bear on fitness for ensoulment. Whereas Cutter's central concern is whether an artificial physical system can be sufficiently fit to receive an immaterial soul, Tijanovic asks whether functionally comparable individuals in computer-generated worlds can satisfy the broader conditions required for ensoulment. He ultimately remains relatively agnostic about virtual ensoulment, while maintaining that the possibility should remain open.

Artificial General Intelligence (AGI)

Artificial General Intelligence (AGI), which some refer to as strong AI or use

the term to designate something similar, is a term for AI systems that perform a variety of cognitive activities, particularly thinking, at a human level or at a level beyond and more advanced than that of human beings. Such systems perform a wide range of tasks, much as humans do. They are said to actually think, rather than merely behave as though they think. Some claim that, if such systems are constructed, they will possess sensation and consciousness, much like human beings (Cycleback 2019, 11; Leon with ChatGPT 2025, 242–243).

The AGI project aims at intelligence as a whole – an intelligence with multiple dimensions that can be deployed across a wide variety of situations. Once fully implemented, an AGI system will very likely resemble the human brain/mind in a number of respects (Goertzel & Wang 2007, 1–2). The project and the pursuit of AGI have been subject to various criticisms, including the claims that its realization is impossible; that there is no such thing as general intelligence; that it is still too early to work on it; that AGI is nothing more than hype; that research on AGI is not worthwhile because it is too difficult to be successfully realized; and, indeed, that AGI is dangerous (Goertzel & Wang 2007, 4–8).

Approaches to Creating Artificial General Intelligence

One approach that may appear to provide a reliable route to achieving AGI is Emulating the Human Brain, that is, creating an exact replica of the human brain in a digital simulation, from the highest level of organization down to the atomic level. In this sense, creating AGI is merely an engineering problem. Given that general intelligence, in the sense in which it is exhibited by human beings – who are particular configurations of atoms – is possible, all that is required is to analyze these configurations and atomic arrangements in complete detail and then reconstruct them in a computer. AGI would then emerge and come into existence through the application of a particular form of nanotechnology and within *in-silico* physics and computational space. According to this approach, the human brain serves as the guide and driving force for the production of AGI (Goertzel & Pennachin 2007, 17).

On this approach, if the appropriate physical organization of matter (at the atomic level) is reproduced, mind and intelligence will emerge as a consequence. It is important that the precise material composition of the system does not play an exclusive or determinative role; rather, what matters is its arrangement, configuration, and functional organization. According to this view, mindedness is not dependent on a biological substrate, and the brain is merely one possible substrate; silicon can fulfill the same role. If matter can be organized precisely at the atomic scale, intelligence will emerge. On this

view, to engineer the mind, one must engineer matter. The mental world is an extension of the material world, provided that matter possesses the requisite organization. This kind of view is based on the belief in the multiple realizability of the mind and intelligence and, likewise, on the principle of substrate independence, or substrate invariance, according to which mind and intelligence can be realized independently of their material substrate.

Another approach has been termed Emulating the Human Mind, that is, constructing virtual neurons, virtual synapses, and virtual activations of these components. According to this approach, the overall structure of the human brain is emulated and replicated, but not the details in which that structure is implemented and realized. The aim here is to emulate the human mind as it is studied by cognitive psychologists, while the manner in which the human mind is realized in the human brain is considered entirely irrelevant. The traditional logic-based approach to artificial intelligence belongs to this camp (Goertzel & Pennachin 2007, 19–20). As will be discussed and clarified below, Cutter appears to have advanced his account of AGI ensoulment on the basis of an approach more closely resembling Emulating the Human Mind.

Another approach is to achieve AGI by creating artificial life (see Goertzel & Pennachin 2007, 22–25), while yet another is the Integrative Approach (see Goertzel & Pennachin 2007, 26, 75–76).

The AI Ensoulment Hypothesis

As noted above, Brian Cutter argues for the claim that some future AI systems will acquire immaterial souls. Accordingly, his claim is not that: (a) current intelligent systems possess souls; (b) all future AI systems will become ensouled; or (c) the soul is made of matter or is a product of computational complexity. Rather, his claims are that: (α) a material system can become ensouled; that is, ensoulment is embodied, and an immaterial soul is in some way connected with a material, physical body; (β) the soul is immaterial, that is, it is not made of physical matter; (γ) an artificial system can be ensouled, and ensoulment is not exclusive to naturally occurring biological systems; rather, more generally, any material system possessing certain properties and capacities – whether natural or artificial, terrestrial or extraterrestrial – can be ensouled; and (δ) ensoulment is related to intelligence (whether intelligence is a prerequisite for ensoulment or intelligence is a manifestation or consequence of ensoulment).

Cutter appears to mean that a material system is capable of receiving a soul when it is capable of being intelligent; that is, intelligence is a prerequisite for ensoulment. If, however, intelligence itself is a consequence of ensoulment,

then we must consider what the necessary and sufficient conditions for the ensoulment of a material system are; in other words, under what conditions does a material system acquire an immaterial soul?

It is already clear from this that Cutter's argument rests on two premises, as he himself explicitly states: (1) that substance dualism is correct with regard to the relation between soul and body, since, according to this theory, an ensouled being possesses both a material body (i.e., a material system suitably constituted and appropriately disposed to receive a soul) and an immaterial soul; and (2) that a material system must possess some particular property or aptitude for ensoulment, since it is evident that not every material system is ensouled. Hence, possession of a soul depends on the fulfillment of certain conditions in the material entity. Cutter assumes that an Artificial General Intelligence (AGI) system – which has not yet been achieved but may be developed in the future – will possess the requisite suitability and aptitude for ensoulment (Cutter 2025, 1).

Basic Assumptions

Cutter begins his discussion with the assumption that substance dualism is correct in the case of human beings. It seems clear, however, that his assumption should not be that this theory is correct only of human beings. Rather, his claim and hypothesis require the theory to be correct of any material ensouled being. Nevertheless, the starting point of his discussion is the truth of substance dualism in the case of human beings. In other words, he begins with human ensoulment and human beings' status as substances composed of two distinct kinds of substance, and seeks to extend ensoulment to certain other material systems. On this basis, every living human body is associated with an immaterial entity that, in some way, constitutes the basis or fundamental bearer of distinctively human mental capacities, such as consciousness, thought, and rational agency. The fundamental question, then, is: given this assumption, should we expect future AI systems to possess souls as human beings do? Could highly advanced future artificial intelligences also come to possess souls?

He calls this hypothesis the AI Ensoulment Hypothesis and, by means of two arguments offered in its support, seeks to establish that this hypothesis should at least be taken seriously. Given that we human beings have souls, he argues, we should not be certain that computers will not be ensouled at all. He maintains that the hypothesis should not simply be dismissed as a fantasy or a remote possibility. In his view, if we accept the existence of souls in the case of human beings, we cannot, without sufficient argument, be confident that future artificial beings will never possess souls (Cutter 2025, 1). Cutter's

argument is a conditional: if substance dualism is true, can we then say with confidence that AI will never possess a soul?

By "soul," Cutter does not mean merely an entity that possesses consciousness. For him, the soul is an immaterial entity that is the fundamental bearer or possessor of certain mental capacities. Accordingly, the central question of his article is not whether AI can become conscious, but rather: Could an artificial being possess the same kind of immaterial existence that dualists attribute to human beings?

Cutter's other assumption is that, in the future, a form of AGI will be created that matches or even surpasses human beings in all or nearly all cognitive domains in terms of behavior and function. He emphasizes that this definition of AGI is purely behavioral and functional. Thus, phenomenal consciousness or an authentic and genuine mental life is not part of the concept of AGI as he defines it. For the purposes of his argument, he assumes that it is sufficient to argue for the ensoulment of a minimal form of AGI. His assumption is that such machines could perform virtually any behavioral task as well as, or better than, human beings. He also assumes that the internal functional organization of these systems will in some way resemble that of the human brain. This resemblance need not, of course, be physical or biological; it may exist merely at the level of organization and functional relations (Cutter 2025, 1). On this basis, it appears that Cutter's discussion of AGI ensoulment is grounded in an approach more closely resembling the Emulating the Human Mind approach described above.

On this basis, the central issue in Cutter's article is not the possibility of constructing AGI, but rather its metaphysical status. If AGI is eventually realized, and if human beings genuinely possess souls, should we not entertain the possibility that these artificial systems might also possess souls? Cutter's answer is affirmative. He does not claim that AGI becoming ensouled is the most probable scenario, but he maintains that its probability is sufficiently high that it cannot simply be ignored. His assignment of a 25 percent probability to the hypothesis indicates that he does not intend to defend his claim dogmatically; rather, he seeks to show that the AI Ensoulment Hypothesis should not be classified among highly improbable hypotheses (Cutter 2025, 1–2).

Cutter seeks to show that the two principal assumptions underlying the article – namely, that substance dualism is true and that human beings will ultimately succeed in constructing AGI – are sufficiently reasonable to make discussion of "AI ensoulment" worthwhile. If either of these assumptions were clearly false, the question of AI ensoulment would likewise lose its relevance. In his view, however, neither assumption can be easily dismissed (Cutter 2025, 2).

It is important to note that Cutter deliberately defines AGI in behavioral

and functional terms. Here, "behavior" and "function" refer primarily to cognitive behavior and function, such as problem-solving, learning, reasoning, planning, language use, the performance of cognitive tasks, and the like, but not necessarily to experiences such as pain, color experience, phenomenal consciousness, or first-person experience. Thus, Cutter seeks to ask: Can the mere presence of advanced cognitive functions provide a basis for ensoulment, even if we are still uncertain whether the entity possesses phenomenal consciousness or is a subject?

In defending the assumption of substance dualism, Cutter appeals to the well-known problem of consciousness. In his view, it remains unclear how consciousness can be explained solely in terms of physical processes. This difficulty may provide motivation for accepting some form of dualism. He then presents several considerations that, in his view, make substance dualism superior to property dualism, including its purportedly better account of personal identity, the unity of consciousness, and certain other metaphysical problems (Cutter 2025, 2). With respect to AGI, Cutter likewise argues that the pace of progress in AI technology has been sufficiently rapid that the possibility of achieving artificial general intelligence in the coming decades cannot reasonably be dismissed (Cutter 2025, 2–3).

Cutter rightly points out that taking seriously the possibility that the AI Ensoulment Hypothesis might actually be realized is important and has significant implications. He moves from a metaphysical discussion of the soul to its ethical and practical consequences; in fact, he seeks to show that even if the probability that AI will be ensouled is only "moderate," it nevertheless cannot be ignored, because its ethical implications are far-reaching. Cutter emphasizes that his claim is highly cautious. He does not argue that AI will certainly become ensouled, or even that this outcome is highly probable. His only claim is that we should not regard the possibility as negligible. Yet even this modest claim is of considerable ethical significance.

If we assign some probability to the possibility that AI may possess a soul and, consequently, some form of intrinsic value or moral status, then we cannot treat it merely as a tool. From this perspective, a strong Moral Risk Argument is at work: if AI is ensouled, it is presumably endowed with significant moral status. We cannot therefore regard the probability of its being ensouled as negligible. Ignoring the interests of an entity that may possess moral status could constitute a serious moral error. Accordingly, we should assign some weight to the possible interests of AI in our decision-making. Under such circumstances, its possible interests, suffering, or preferences should also be incorporated into our moral calculations.

Nevertheless, Cutter emphasizes that the current evidence is insufficient to

warrant a high degree of confidence in the AI Ensoulment Hypothesis. In his view, just as decisively rejecting the hypothesis is unwarranted, decisively accepting it is unwarranted as well. The appropriate position, therefore, is for the time being a form of suspension of judgment, accompanied by taking the possibility seriously and assigning it a non-negligible probability (Cutter 2025, 3).

What Is Meant by Soul?

Before discussing "AI ensoulment," Cutter clarifies what he means by "soul" and what he does not presuppose by this concept. Here, Cutter seeks to provide a minimal and relatively inclusive definition of the soul, one that can be shared by different formulations of dualism. He regards the soul as an immaterial entity that in some way underlies particular mental capacities and faculties, such as consciousness, thought, and rational agency, while at the same time being united with a physical system.

The immateriality of the soul, in general, means that the soul is not made of anything physical; it is not composed of atoms, molecules, or anything else investigated in microphysics. Cutter merely takes the soul to be the basis of mental capacities and deliberately refrains from entering into certain deep metaphysical disputes. He does not intend to address the question of clarifying the precise relation between the soul and mental properties. For example, he does not clarify whether the soul itself is the subject that thinks and chooses, or whether the actual subject is something else – perhaps an entity composed of body and soul – to which agency can nevertheless be attributed by virtue of the soul's functioning, just as it is the person who breathes through the activity of their lungs, and it would be incorrect to say that the lungs themselves breathe.

Nor does he take a position on the question of whether the person is identical with the soul or is a more encompassing entity of which the soul constitutes only one part. This question is, of course, highly important and lies at the heart of the disagreement between the Cartesian and Aristotelian views of the soul. Cutter likewise does not address several other disputed issues, such as whether the soul is responsible for non-mental vital functions, such as nutrition, growth, and digestion, or whether non-human animals also possess souls.

In this way, he seeks to provide a definition of the soul that is acceptable both to Cartesian dualists and to proponents of certain other theories of the soul. He states that, for his present purposes, it is sufficient to take substance dualism to entail that the soul is a nonphysical substance, that is, a particular concrete object distinct from every wholly physical object, event, or process (Cutter 2025, 3–4).

What Is Meant by Ensoulment?

Cutter defines the concept of ensoulment operationally. In effect, he seeks to determine precisely when a physical system can be said to "possess a soul." In his view, the mere existence of a soul somewhere in the universe is not sufficient; rather, there must be a direct and stable causal connection between the soul and a physical system. He conceives of this connection as operating in both directions. On the one hand, bodily states can affect states of the soul; for example, neural processes may give rise to sensory experiences (upward causation, or body-to-soul causation, bottom-up). On the other hand, the soul can also affect the body; for example, an individual's volition may cause bodily limbs to move (downward causation, or soul-to-body causation, top-down). Any system that stands in such a regular causal relationship with a soul counts, in Cutter's view, as "ensouled":

> a physical system (like a physical computer) is "ensouled" (or "has a soul," or is "united to a soul") only if there are direct causal connections of this kind between the physical system and a soul, such that the states or activities of the physical system directly and systematically affect the soul, or vice versa (inclusive "or") (Cutter 2025, 4).

Accordingly, the central question becomes: Can an AGI, in addition to being a physical system, possess a soul such that this kind of causal interaction obtains between the two?

Conditions for Ensoulment

After specifying what he means by ensoulment, Cutter argues that ensoulment is neither arbitrary nor accidental. If substance dualism is true, we should expect there to be a lawful, reliable, and counterfactual-supporting system such that, if conditions C obtain in a physical system, a soul will be united with it (or the system will be ensouled). Conditions C might consist of a particular set of states and properties of a human embryo (or, in accordance with what ancient Peripatetic philosophers held, the formation of a particular human temperament). Cutter's important point is that conditions C must be conditions that are sufficient for ensoulment according to the laws of nature – that is, a nomologically sufficient condition for ensoulment.

On this basis, the Ensoulment Hypothesis would be roughly equivalent to the following hypothesis: the matter constituting some future AI systems possesses conditions that are nomologically sufficient for ensoulment. Now, if we are speaking about the ensoulment of an intelligent system that is functionally similar to a human being, conditions C would constitute functional

sufficiency; that is, possessing a human-like functional structure and organization would be a condition that is nomologically sufficient for ensoulment. Cutter maintains that the two arguments he presents in support of the Ensoulment Hypothesis lend support to something resembling the Functional Sufficiency Hypothesis. Cutter identifies various possible structural and organizational similarities between a machine and a human being that could be sufficient for ensoulment, without attempting to determine precisely what those necessary similarities are (Cutter 2025, 5).

Thus, in Cutter's view, human beings are ensouled because the human body possesses certain distinctive natural conditions. He takes these conditions to be of the kind constituted by functional organization, because if human ensoulment were contingent upon the particular material constituting the human body, then talking about the ensoulment of an intelligent machine made of different materials would be meaningless and pointless. Thus, there is something in the human body, and something that has occurred in it, that accounts for human ensoulment. That something is not the material constituting the human body itself, but rather its functional organization and structure. If, then, a machine is to be ensouled, it must resemble a human being in this respect. The reasoning implicit in Cutter's account can be formulated as follows:

1. There is a law-governed condition (C) for ensoulment.
2. That condition is possession of a particular (human-like) functional organization.
3. Assumption: AGI will possess that functional organization.
4. Conclusion: AGI will satisfy the sufficient condition for ensoulment and will therefore be ensouled.

One critical question concerning Cutter's view and the path he takes is why functional organization should be sufficient for ensoulment. Do distinctive human biological structures or particular chemical compositions of the human body not play any role in ensoulment? Cutter appears to have adopted a kind of functionalist presupposition concerning cognition and does not regard ensoulment as dependent on specifically human biological matter or structure, as noted earlier. Whether this presupposition is correct or mistaken remains an open question.

Cutter is not, of course, a functionalist in the sense that he identifies the soul with a function. In his view, which connects ensoulment with substance dualism, functional organization is a condition for becoming ensouled. Thus, on his account, the soul is something distinct from a function or from functional organization.

Cutter makes an important point concerning the relation between the

nomologically sufficient conditions for ensoulment and the actual coming into existence of a soul. The kind of lawfulness Cutter has in mind is nomological lawfulness, which does not necessarily imply natural laws. According to the anti-emergentist view, natural entities can never be the cause of the coming into existence of a soul; rather, they merely provide the occasion or conditions for God to create a soul and infuse or unite it with a natural entity. On this view, such nomological regularity does not express a natural law, because if it were a law of nature, the soul would have to be an effect of the physical natural entity itself, as emergentists maintain.

Nevertheless, even on an anti-emergentist view, it is possible to speak of conditions that are nomologically sufficient for ensoulment – conditions that would constitute the occasioning cause for the divine creation of a soul. With these considerations in mind, Cutter emphasizes that the AI Ensoulment Hypothesis does not entail that we human beings, who are the creators of intelligent systems, create the souls of future AGIs. If future machines are ensouled, our role is merely to provide the material conditions for ensoulment. Whether we are the creators of those souls is an open question; it depends on our metaphysical view concerning the true cause of the soul.

For Cutter, what matters is that, once conditions C are realized, a soul comes to be realized, whether through natural emergence or through special divine creation. In the case of an ensouled human being, whenever certain physical and material conditions are realized – when a human embryo is formed and develops under appropriate conditions, whether in the womb through natural sexual reproduction or in a laboratory through other methods of combining gametes – the human body becomes ensouled. This process is law-governed and supports a counterfactual conditional: if such conditions were not realized, ensoulment would not occur.

The same holds in the case of the ensoulment of a machine: such an order and regularity obtain, and ensoulment is not a random or arbitrary event. Moreover, this account does not depend on any particular metaphysical view concerning the coming into existence of the soul or its relation to the body. For example, it is compatible both with theories according to which the soul comes into existence prior to the body and with theories according to which the soul comes into existence together with the body. One might hold that souls already exist and that material conditions merely provide the occasion for a soul to become united with a particular body, rather than providing the conditions for the creation of a new soul.

Cutter maintains that whatever is true of human beings must also be true of machines. Programmers and engineers provide certain physical conditions, and the artificial intelligence then becomes ensouled, whether the soul is newly

generated and the physical conditions are the cause or occasioning cause of its coming into existence, or whether the soul already exists and the physical conditions are the cause or occasioning cause of its becoming attached or united to the machine's physical system (Cutter 2025, 5–6).

The conclusion, therefore, is that ensoulment (being united with an immaterial soul) is ontologically distinct from any functional property of a physical system (i.e., the Ensoulment Hypothesis is entirely different from functionalism), although certain functional properties may be nomologically sufficient for ensoulment (Cutter 2025, 7).

The Fitting-Recipient Argument

Cutter offers two arguments in defense of the AI Ensoulment Hypothesis, one of which is the "Fitting-Recipient Argument." This argument is based on the idea that, given the assumptions discussed above, whenever a physical system is "fit" to possess a soul, it will become united with a soul (i.e., it will become ensouled). Roughly speaking, "being fit to possess a soul" means that the physical structure of a system is such that it can enter into meaningful interaction and cooperation with the activities of a soul.

Cutter formulates the argument as follows:

Premise (1). Being fit to receive a soul is, in accordance with the laws of nature, a sufficient condition for a physical system to become united with a soul.

Premise (2). A machine that is functionally similar to a human being (i.e., a machine that is appropriately similar to the human body/brain in terms of its functional organization and behavioral capacities) is fit to possess a soul.

Conclusion. A machine that is functionally similar to a human being will become united with a soul (Cutter, 2025, pp. 18–19).

Before defending the two premises of the argument, it is necessary to say something about the meaning and criterion of a body's fitness to receive a soul. Whether a body is fit and suitably constituted to receive a soul depends, to some extent, on the nature and characteristics of the body and, to some extent, on the nature and characteristics of the soul. More specifically, this fitness is a function of the soul's fundamental or basic powers and activities. Cutter seeks to argue that if a machine possesses powers similar to those of the human body, it will likewise be fit to receive a soul.

The human soul has certain distinctive basic powers or capacities, including

sensory experience, judgment, practical and theoretical reason, and choice. To better understand when a physical system is suitable for such a soul, we should consider cases in which the states and capacities of the human body complement or support the fundamental powers of the soul. In other words, to determine what makes a body a suitable recipient of a soul, we must first determine what the soul does and then examine what kind of body could cooperate with and be suited to the performance of those activities. If the soul possesses sensory perception, thought, decision-making, and choice, what kind of physical system could effectively interact with those powers? (Cutter, 2025, p. 19).

For example, if we want to determine what kind of vehicle is suitable for a pilot, we must first identify the pilot's basic activities (e.g., navigating and controlling a physical object in the air), and we can then determine what kind of structure an aircraft must have in order to make these activities possible. Cutter appears to apply the same model to the soul. He offers an argument along roughly the following lines:

1. The soul possesses cognitive powers;
2. The human body supports these powers;
3. AGI can likewise provide the same kind of support.

Conclusion: AGI can likewise be a suitable recipient for a soul.

Cutter makes several points in an effort to show why the human body is a suitable substrate for the human soul. First, the human body provides the inputs necessary for sensory consciousness. The human body possesses a remarkably rich set of internal states, particularly the neural representations present in the brain's perceptual systems, that reliably correlate with states of the external environment. These physical states can serve as the direct causal basis of the soul's sensory experiences. As a result, through these states, the soul can acquire a rich and accurate representation of its surroundings and, on that basis, form a large number of beliefs about the world that are generally accurate and rational. Most other physical objects, such as rocks or chairs, lack such a structure. A rock does not contain a set of internal states capable of providing appropriate inputs for sensory experience through which a soul could construct an accurate and reliable model of its environment (Cutter, 2025, p. 19). Cutter's argument can be summarized in a single sentence: the human body is useful to the soul because it provides the soul with rich and structured information about the world; the human body is not merely a container for the soul, but a complex information-processing system that continuously supplies the soul with valuable data.

Second, the human body provides its soul with a remarkably broad and

diverse range of behaviors through which the subtleties and complexities of its mental life can be expressed. Since human outward behavior is ultimately produced through the brain, the capacity to express such mental subtleties requires the brain to possess a structure capable of reflecting or encoding them (Cutter, 2025, p. 19). In the previous example, the human body was presented as a source of input for the soul; here, it is presented as a means of obtaining output from the soul. On the one hand, through the body, the soul receives information from and is causally affected by the world; on the other hand, through its body, the soul expresses itself in the world and exerts an influence upon it.

Cutter cites a similar point from J. P. Moreland in explaining why bodies are created. Moreland argues that the more complex an entity's consciousness is, the more complex and sophisticated its body must be in order to enter into causal interaction with the subtle and diverse states of that entity's mind. If an entity possesses highly complex consciousness but is connected to a very simple material object, much of the complexity of its mental life would never have the opportunity to manifest itself or be realized outwardly and would therefore remain without practical effect. For this reason, objects such as rocks or chairs are not suitable recipients for souls. Even if a soul with a complex inner life were attached to a rock, the rock's structure would not allow that complexity to be manifested in its behavior. A rock possesses neither a sufficient diversity of behavior nor the internal organization required for the subtle regulation of behavior.

Third, the human body assists the soul in reasoning and thinking. The brain is organized in such a way that it may play a supportive role in the exercise of the soul's rational faculties. It provides a rich store of memories encoded in neural networks that the soul can draw upon. The brain also contains temporary representations held in working memory, which can facilitate reasoning and function as a kind of "neural scratchpad" or "neural notepad."

A dualist would presumably maintain that no sequence of physical events in the brain is itself the process of reasoning. Nevertheless, some of these events can assist or complement reasoning, just as notes written on a sheet of paper are not themselves part of the reasoning process but can help us carry out that process. By contrast, objects such as rocks or chairs typically do not possess internal states capable of playing such a supportive role in reasoning. Thus, compared with the human brain, a rock is far less well suited to cooperate with the rational faculties of a soul (Cutter, 2025, p. 20).

Defense of Premise 1

After presenting examples of the cooperation between body and soul, Cutter

seeks to extract a general principle from them, namely, to defend the reasonableness of the first premise of the argument (being fit to receive a soul is a sufficient condition for becoming ensouled). The human body is organized in various ways that enable it to facilitate and cooperate with the activities of the soul. It can therefore be said that the human body is "fit to receive a soul."

Moreover, it seems that the natural explanation of the fact that a living human body possesses a soul, whereas rocks and human corpses do not, lies precisely in this difference in their fitness and capacity to cooperate with the activities of the soul. The living human body possesses a structure that enables it to cooperate effectively with the activities of the soul, whereas rocks and corpses lack such a capacity. Cutter then advances a stronger claim: he suggests that even our intuitions about actual and hypothetical cases of ensoulment are implicitly based on this very criterion. For example, if we accept dualism concerning human beings, we often find it natural to suppose that intelligent extraterrestrial beings would also possess souls. The likely reason for this intuition is that we imagine their bodies to be organized in such a way that they can cooperate with the activities of a soul. By contrast, although it is logically possible to suppose that a rock, a chair, or a corpse possesses a soul, we do not take such suppositions seriously because these objects lack the physical organization necessary for meaningful cooperation with the activities of a soul. Taken together, these considerations support the idea that "fitness to receive a soul" is a sufficient and law-governed condition for becoming ensouled (Cutter, 2025, pp. 20–21).

Cutter goes on to explain that the fitness and suitability described above are gradable, admitting of degrees and variations in strength and extent. A physical system can possess a structure that enables it to cooperate with the activities of a soul to different degrees; this is true even among human beings. For example, brain damage or developmental impairments reduce the degree to which a body is suited and fit to cooperate with the activities of the soul (Cutter, 2025, p. 21).

Cutter is addressing here an objection that can readily be raised: if "fitness for a soul" is the criterion for ensoulment, precisely at what point does this fitness begin? Fitness to receive a soul appears to be a gradable property. A physical system can possess, to varying degrees, a structure that enables it to cooperate with the activities of a soul. This raises the question of where the boundary or threshold required for ensoulment should be drawn.

Cutter's response is that, for present purposes, the threshold can be set very high, namely, at the level of an ordinary, mature human body. Since the first premise of the argument provides only a sufficient condition for ensoulment, setting such a high threshold does not entail that creatures below this level lack souls. It merely claims that any system that reaches this degree of fitness will

thereby satisfy the conditions for ensoulment (Cutter, 2025, p. 21).

In fact, Cutter recognizes that the concept of a "fitting recipient" is vague and admits of degrees, because fitness is not a binary property; it is not, for example, like being pregnant, which either obtains or does not. Rather, it is something more like being intelligent or being healthy. If fitness is gradable and admits of degrees, however, what about a chimpanzee? A dolphin? A human infant? A human fetus? A person with severe brain damage? A partially advanced AGI? At what point does ensoulment occur?

Cutter does not answer these questions. Instead, he argues that, for present purposes, we do not need to know precisely where the boundary of ensoulment lies. It is sufficient to say that an ordinary mature human being has clearly crossed the threshold. A human-like AGI will cross it as well. Thus, discussion of borderline cases can be set aside for the time being, as Cutter seeks to focus only on clear cases. He does not claim that only beings with a high degree of fitness possess souls; rather, he claims that any being that reaches this level of fitness will possess a soul. For his argument, it is sufficient to accept that the human level of fitness is sufficient for ensoulment.

Defense of Premise 2

Thus far, Cutter's efforts have been directed toward showing that fitness is a sufficient condition for ensoulment and that the human body is suitable and fit for ensoulment. He now seeks to take the second important step and argue that what gives rise to this fitness is nothing more than functional organization. The second premise of the argument is that a functionally human-like machine will be suitable for receiving a soul. The reason for this claim is that fitness to receive a soul depends entirely on the functional organization and behavioral capacities of a system. All the examples previously given to explain the suitability of the human body for the soul concerned only the functional and behavioral features of the human body, not the particular material of which the body is composed. For example, it was argued that the brain possesses a rich set of internal states that can serve as inputs for sensory experiences and make possible the formation of a rich model of the environment. But this feature is not dependent on the biological nature of the brain. A computer could likewise possess similar internal states, even if those states were realized in a non-biological substrate. The same point applies to the other examples as well. Thus, if fitness to receive a soul depends solely on functional organization and behavioral capacities, and if the human body is suitable for the soul because of precisely these features, then a machine that is functionally similar to a human being will likewise be equally suitable for receiving a soul (Cutter, 2025, p. 21).

The most important concept here is substrate independence, on which the entire argument is effectively based. The idea is this: suppose that the human brain is composed of neurons. What is relevant to and plays a role in ensoulment? The first option is the biological material of the neurons; the second is the pattern of relations and functions realized by the neurons. Cutter chooses the second option. He argues that if something can perform the same roles as neurons – that is, process the same information, possess the same representations, and produce the same behaviors – it will possess the same fitness for ensoulment. Thus, for present purposes, it makes no difference whether the physical system is made of neurons, silicon chips, or some other material.

Cutter refers precisely to the examples he himself presented earlier. In any of those previous examples, did he refer at any point to carbon, DNA, or proteins? The answer is no. All of the examples concerned information, representation, memory, behavior, reasoning, and the like. Why, then, should biology suddenly become relevant?

Critical Evaluation

Cutter employs the approach of substance dualists to establish the ensoulment of intelligent machines. Until now, substance dualists have established ensoulment only in the case of natural (i.e., non-artificial) beings, and it may previously have been assumed that an artificial being could not become ensouled, or that there was no reason to regard its ensoulment as even possible. In the case of a natural human being, the substance dualist maintains that whenever an appropriately constituted human body comes into existence, it will become ensouled, according to various metaphysical explanations that have been offered in this regard, some of which were mentioned above.

The particular metaphysical explanation of how a human body becomes ensouled is independent of the conclusion Cutter seeks to establish. His claim is that, for the same reason and through the same process by which an appropriately constituted human body becomes ensouled, an intelligent machine will also become ensouled, and that there is no relevant difference between the two cases. In other words, if substance dualism is true of human beings, it should also be true of human-like machines, because what matters is the presence of certain distinctive mental powers and states together with a body suitably constituted for those powers and states – and both conditions obtain in the case of both natural human beings and AGI.

Accordingly, we should posit a general principle common to human beings and AGI. If we call the concept common to the human soul and the AGI soul

soul*, and the concept common to the suitably constituted human body and the suitably constituted physical body of a machine (which are functionally similar to one another) body*, then:

Whenever body* is realized, it becomes united with soul*.

Anyone who rejects this general principle and maintains that a machine cannot become ensouled, thereby objecting to the Fitting-Recipient Argument, must, if they find the first premise acceptable on a dualist approach, regard the functional similarity between the machine and the human body as insufficient for the machine's ensoulment – that is, they must reject the second premise. In that case, they must introduce some additional condition or conditions for ensoulment that are realized in human beings but cannot be realized in machines, and consequently conclude that a machine will never become ensouled.

Békefi (2026) adopts this strategy in criticizing Cutter. He argues that some additional condition must be involved and calls this condition "Integrity." Accordingly, Békefi's objection to Cutter is what he calls "The Missing-Integrity Objection.[1]" Békefi accepts the first premise of the Fitting-Recipient Argument and directs his objection against its second premise. However, Békefi's strategy does not seem adequate for two reasons. First, he does not address the central problem with the argument, namely, treating a necessary condition as a sufficient one. Second, his criticism is open to response, as he himself acknowledges at the end of his critique of the Fitting-Recipient Argument:

If engineers could build a neural network using billions of memristors of that kind, and manage to create for it an environment suitable for the kind of deep learning characteristic of both animals and artificial neural networks, it is conceivable that a system could be produced with both human-level intelligence and high integrity. It would not be easy to decouple its physical parts and functional capacities (though the danger of being turned off would likely remain). I have no objection to the ensoulment of such an entity – but I suspect that rather than falling under our ordinary concepts of 'AI' and 'machine' (cf. Section 1.3), it would better be classified as some other thing – perhaps a synthetic brain. (Békefi, 2026, p. 134)

Rejection of Premise 1

As discussed above, "being fit to possess a soul" roughly means that the

1. I leave the examination of this objection for another occasion.

physical structure of a system is such that it can enter into meaningful interaction and cooperation with the activities of a soul. According to the first premise of the Fitting-Recipient Argument, if fitness obtains in a physical system, that system will become ensouled without exception or violation.

In explaining the truth and reasonableness of this premise, Cutter argues that the human body is fit to receive a soul and that the natural explanation of the fact that the living human body (and every other living body) possesses a soul, whereas certain objects, such as rocks and chairs, do not, lies in their difference in fitness for ensoulment. In other words, we have encountered ensouled entities, and we have explained their conditions and their difference from soulless objects by holding that they are ensouled because their bodies are fit to receive a soul. Fitness is a characteristic of an ensouled body precisely insofar as it is ensouled, and it expresses a particular relation between a body and its soul. This means that bodily fitness plays a necessary and effective role in ensoulment, in the sense that, had such fitness been absent, ensoulment would not have occurred. This state of affairs can be expressed by the following conditional proposition:

(I) If an entity is ensouled, then it has a body that is fit to receive a soul.

Proposition (I) means that ensoulment entails having a suitably constituted body; in other words, having a body that is fit to receive a soul is a necessary condition for ensoulment, while ensoulment is, conversely, a sufficient condition for having a body that is fit to receive a soul. (Here, necessary and sufficient conditions are used in accordance with standard philosophical theory; see Brennan, 2024).

By contrast, the first premise of the Fitting-Recipient Argument is:

(II) If an entity has a body that is fit to receive a soul, then it is ensouled.

Proposition (II) means that having a body that is fit to receive a soul is a sufficient condition for ensoulment, while ensoulment is a necessary condition for having a body that is fit to receive a soul.

The crucial issue that undermines the argument is that while proposition (II) serves as the actual premise of the argument, Cutter's reasoning can at most establish proposition (I), which is the converse of (II). In other words, what is true and what can be supported by argument is proposition (I) – that having a body fit to receive a soul is a necessary condition for ensoulment – yet the converse of this proposition has been introduced as a premise of the

argument, namely, that having a body fit to receive a soul is a sufficient condition for ensoulment. The distinction between the two has been conflated, resulting in a logical fallacy. This fallacy is known variously as affirming the consequent, the converse error, or treating a necessary condition as sufficient.

In fact, the argument would have to be reformulated as follows:

(I) If an entity is ensouled, then it has a body that is fit to receive a soul.

Premise (2). AGI has a body that is fit to receive a soul.

Conclusion. AGI is ensouled.

Formulated in this way, it is obvious that this argument commits the formal fallacy of affirming the consequent.

Indeed, one may accept that fitness for ensoulment is necessary for ensoulment, and that if such fitness is absent, ensoulment will not occur. However, it does not follow that fitness is all that is required for ensoulment. Fitness is a necessary condition for ensoulment, but it is not a sufficient condition. Therefore, even when fitness is present, one cannot infer that ensoulment will occur.

Thus, although AGI may be fit for ensoulment by virtue of its functional similarity to the human body, it does not follow from this fitness that AGI will necessarily become ensouled. Accordingly, the Fitting-Recipient Argument for the ensoulment of artificial intelligence fails.

Therefore, although the Fitting-Recipient Argument, as formulated by Cutter, is formally valid, its first premise appears to be problematic. According to the first premise, having a body that is fit to receive a soul is a sufficient condition for ensoulment. However, the argument and explanation Cutter offers in support of it can establish only, at most, that having a body that is fit to receive a soul is a necessary condition for ensoulment. Now, if we revise the first premise by replacing the sufficiency of having a fitting body for ensoulment with its necessity, the argument becomes formally invalid, since it would then commit the fallacy of affirming the consequent and would no longer establish its conclusion.

It might be argued that Cutter, as he himself claims, has successfully argued and explained that having a body fit to receive a soul is a sufficient condition for ensoulment, and that the first premise is therefore true. However, this does not seem to be the case. To clarify why Cutter's explanation and argument establish, at most, that fittingness is a necessary condition for ensoulment, let us examine the possible relations between fittingness and ensoulment in the

case of human beings:

Fittingness for ensoulment is either the sole condition for the occurrence of ensoulment (in which case it would be both a necessary and a sufficient condition) or one of several conditions required for ensoulment to occur. Cutter does not claim, nor do his explanation and argument show, that fittingness is the sole condition for the occurrence of ensoulment. Moreover, this assumption appears to be incompatible with the metaphysical commitments and understanding of the causal relation between soul and body held by some substance dualists (recall Cutter's point concerning emergentism and anti-emergentism: if ensoulment were made to depend wholly and lawfully on the instantiation of a particular bodily state, we would be moving toward an emergentist theory of the soul–body relation. Of course, some philosophers endorse emergentist dualism; see Hasker, 2018; for a critical discussion, see Rickabaugh, 2018. For a contemporary formulation and defense of Cartesian substance dualism, see Swinburne, 2018). Thus, assuming that fittingness is only one of the conditions required for ensoulment to occur – and Cutter agrees with this assumption and does not link the truth of the AI Ensoulment Hypothesis to the uniqueness of the fittingness condition – we shall proceed with the discussion.

If the occurrence of ensoulment requires the instantiation of $n$conditions and fittingness is one of them, it does not follow merely from this fact that fittingness is a sufficient condition for ensoulment. For fittingness to be a sufficient condition, it must be shown that whenever fittingness obtains, the other conditions necessary for ensoulment will necessarily obtain as well. Only then can one infer from the instantiation of fittingness to the instantiation of the full set of conditions necessary for ensoulment and, consequently, to the occurrence of ensoulment. In such a case, it may, with some qualification, be said that fittingness is a sufficient condition for the occurrence of a soul in a physical body, since the other necessary conditions are guaranteed by its instantiation. But is the assumption that the instantiation of fittingness also guarantees the instantiation of the other conditions necessary for ensoulment tenable? Cutter's argument and explanation do not suffice to establish this claim. At most, Cutter shows that human beings are ensouled and that, in cases where ensoulment occurs, the human body is fit to receive a soul. He says that the natural explanation of the fact that a living human body has a soul, whereas rocks and human corpses lack souls, appears to be precisely this difference in their fitness and cooperation with the activities of the soul. But what does he mean by "natural explanation"?

Cutter may mean by "natural explanation" a form of inference to the best explanation. In that case, in explaining the relation between fittingness and

ensoulment in human beings and other ensouled entities, as well as the relation between unfittingness and the absence of ensoulment in entities such as rocks and chairs, he appeals to a form of inference to the best explanation and takes fittingness to be the best explanation of this pattern of ensoulment and non-ensoulment. However, the fact that fittingness is the best explanation of this pattern does not, in itself, entail that fittingness is either a sufficient or even a necessary condition for ensoulment.

Suppose that all existing instances of Ahave two properties, $a$and b, and that $a$enjoys a certain priority over b(more precisely, that A's having $a$is causally prior to A's having b), such that the best explanation of why instances of Ahave property $b$is that they have property a. It does not follow from this explanation that the presence of $a$is a sufficient condition for the presence of b. To establish that $a$is a sufficient condition for b, it must be shown that necessarily (or in every possible world), anything that has $a$also has, or must have, b. Mere observation of the co-occurrence of $a$and $b$is not sufficient here; rather, it must be shown why the instantiation of $a$entails the instantiation of b. In other words, it must be shown that the non-instantiation of b is incompatible with the instantiation of a – that is, that it is impossible for an entity to have $a$without having b(in the present case, that it is impossible for an entity to have a body fit to receive a soul without being ensouled).

Cutter, of course, claims that he is going to establish precisely such a claim; that is, he intends to establish that there is a natural nomological relation between fittingness and ensoulment. The point, however, is that his argument for the first premise cannot accomplish this. Inference to the best explanation does not yield such a conclusion. In fact, generalizing from actual instances to metaphysical sufficiency is problematic unless it is shown that there is an essential or necessary relation between fittingness and ensoulment. It is not impossible for a body fit to receive a soul to exist without becoming ensouled, since ensoulment may be conditional upon some further condition or conditions, at least one of which may fail to obtain.

If fittingness, C, is a sufficient condition for ensoulment, S, then:

$$C \rightarrow S$$

Now, if ensoulment requires $n$conditions:

$$C_1 \wedge C_2 \wedge \cdots \wedge C_n \rightarrow S$$

and fittingness is one of them – for example:

$$C_1 = C$$

then, for ensoulment to occur, it must be the case that whenever C obtains, $C_2, \ldots, C_n$ also obtain. Accordingly, we must have:

$$C \rightarrow (C_2 \wedge \cdots \wedge C_n)$$

That is, fittingness must guarantee the instantiation of the other conditions necessary for ensoulment as well. But Cutter's argument cannot establish this. Unless this relation is established, C is, at most, a necessary condition rather than a sufficient one.

In fact, Cutter appears to be confronted with a pattern of explained correlation or co-occurrence: every instance of ensoulment is accompanied by fittingness. He takes fittingness to explain ensoulment. However, for fittingness to be a sufficient condition for ensoulment, he would have to show that no other independent metaphysical condition is required for ensoulment. Cutter's argument does not establish this, and in fact, establishing it would not be easy, if it is possible at all. To move from an inferential explanation to such a metaphysical claim, further metaphysical premises or principles would have to be provided that establish a necessary relation between fittingness and ensoulment.

However, Cutter may mean by "natural explanation" a natural nomological explanation, as discussed above. From Cutter's perspective, the AI Ensoulment Hypothesis is roughly equivalent to the hypothesis that the material constituting an ensouled physical entity – whether a natural or an artificial entity – possesses conditions that are sufficient for ensoulment in accordance with the laws of nature, and he identifies these conditions with "functional sufficiency." But in that case, the argument appears to beg the question, since the Fitting-Recipient Argument is supposed to support the Functional Sufficiency Hypothesis, while, in justifying its first premise, it appeals to a "natural explanation" – that is, precisely the natural nomological relation and functional sufficiency that it was supposed to establish.

Conclusion

Cutter's account and argument for the ensoulment of artificial intelligence are valuable and noteworthy. As Békefi puts it, "Cutter's is the only detailed discussion of AI ensoulment in recent memory; indeed, perhaps the first serious treatment of the question in analytic philosophy since Turing's half-hearted comments" (Békefi, 2026, p. 121).

Cutter has sought to explain the AI Ensoulment Hypothesis clearly by articulating its underlying assumptions and key concepts, and he has also done a commendable job of formulating the premises of the Fitting-Recipient Argument. He has constructed an argument for the ensoulment of machines in the form of a *modus ponens*. However, the argument appears to face a significant problem. In effect, he commits the fallacy of affirming the consequent: the first premise of the argument, which he treats as a sufficient condition, is in fact established only as a necessary condition, and his argument therefore commits the fallacy of treating a necessary condition as sufficient.

In fact, everything Cutter has offered concerning the relationship between ensoulment and bodily fitness establishes only that fitness is a necessary condition for ensoulment. For his argument to be valid, however, he needs fitness to be a sufficient condition for ensoulment, as he himself explicitly maintains. Yet he does not appear to provide an argument in support of its sufficiency. Consequently, the Fitting-Recipient Argument does not succeed in establishing the ensoulment of AI.

▣ Conflict of Interest

The author declares no competing interests.

References

Békefi, B. (2026). No ghost in the machine: Doubting AI ensoulment. *Faith and Philosophy: Journal of the Society of Christian Philosophers*, *42*(1), Article 7, 121-146 https://doi.org/10.37977/faithphil.2026.42.1.7

Brennan, A. (2024). Necessary and sufficient conditions. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2024 ed.). Stanford University. https://plato.stanford.edu/archives/sum2024/entries/necessary-sufficient/

Cutter, B. (2025). The AI ensoulment hypothesis. *Faith and Philosophy: Journal of the Society of Christian Philosophers*, *41*(1), Article 1, 1–26. https://doi.org/10.37977/faithphil.2025.41.1.1

Cycleback, D. (2019). *Philosophy of artificial intelligence*. Bookboon.

Goertzel, B., & Pennachin, C. (Eds.). (2007). *Artificial general intelligence*. Springer Verlag.

Goertzel, B., & Wang, P. (Eds.). (2007). *Advances in artificial general intelligence: Concepts, architectures and algorithms: Proceedings of the AGI Workshop 2006*. IOS Press.

Hasker, W. (2018). The case for emergent dualism. In J. J. Loose, A. J. L. Menuge, & J. P. Moreland (Eds.), *The Blackwell companion to substance dualism* (pp. 61–72). Wiley-Blackwell. https://doi.org/10.1002/9781119468004.ch4

Kristjánsson, K. (2010). *The self and its emotions*. Cambridge University Press.

Leon, F. (with ChatGPT). (2025). *Philosophy of artificial intelligence*. ResearchGate. https://www.researchgate.net/publication/388146273_Philosophy_of_Artificial_Intelligence

McGinn, C. (1999). *The mysterious flame: Conscious minds in a material world*. Basic Books.

Oberg, A. (2023). Souls and selves: Querying an AI self with a view to human selves and consciousness. *Religions*, *14*(1), Article 75. https://doi.org/10.3390/rel14010075

Rickabaugh, B. L. (2018). Against emergent dualism. In J. J. Loose, A. J. L. Menuge, & J. P. Moreland (Eds.), *The Blackwell companion to substance dualism* (pp. 73–86). Wiley-Blackwell.

Swann, J. (2021). Anima ex machina: Can artificial intelligence have soul? In W. H. U. Anderson (Ed.), *Technology and theology* (pp. 201–216). Vernon Press.

Swinburne, R. (2018). Cartesian substance dualism. In J. J. Loose, A. J. L. Menuge, & J. P. Moreland (Eds.), *The Blackwell companion to substance dualism* (pp. 133–152). Wiley-Blackwell.

Tijanovic, T. (in press). The virtual ensoulment hypothesis: Can virtual individuals be ensouled? Dialogo.

Tiwari, S. K., & Tomar, P. (2026). Why AI can never have a soul: A philosophical inquiry into language and consciousness. *Journal of Philosophical Investigations*, *20*(54), 233–252. https://doi.org/10.22034/jpiut.2025.69223.4202