A Kantian approach to the possibility of mechanical consciousness[a]

Sung-Yeop Jo[*]

Abstract

Il progresso dell'intelligenza artificiale (IA) mette in discussione le definizioni tradizionali di coscienza. Questo articolo applica la filosofia di Immanuel Kant per sostenere che, sebbene la capacità di un'IA di sintetizzare i dati sensoriali in rappresentazioni oggettive soddisfi il criterio funzionale di una coscienza rudimentale, come delineato nella sua Deduzione A, ciò deve essere distinto dalla coscienza di sé più robusta della Deduzione B. Questa consapevolezza di ordine superiore, identificata con il "penso" o l'appercezione originale, è un atto spontaneo di un sé unificato e la base per un vero e proprio agire. Kant fonda i diritti morali e legali sullo status di persona, che richiede tale agire autonomo e autocosciente. Essendo un sistema che segue esclusivamente delle regole, un'IA non ha la capacità di formulare le proprie massime da una prospettiva in prima persona. Pertanto, sebbene un'IA possa essere considerata "cosciente" in senso kantiano limitato, non può qualificarsi per lo status etico o giuridico di persona.

Parole Chiave: intelligenza artificiale, coscienza, appercezione, persona, agire

The advancement of artificial intelligence (AI) challenges traditional definitions of consciousness. This article applies the philosophy of Immanuel Kant to argue that, while an AI's capacity to synthesize sensory data into objective representations meets the functional criterion for a rudimentary consciousness as outlined in his A-Deduction, this must be distinguished from the more robust self-consciousness of the B-Deduction. This higher-order awareness, identified with the 'I think' or original apperception, is a spontaneous act of a unified self and the basis for genuine agency. Kant grounds moral and legal rights in the status of personhood, which requires this autonomous, self-conscious agency. As a purely rule-following system, an AI lacks the capacity to formulate its own maxims from a first-person perspective. Therefore, while an AI may be considered 'conscious' in a limited Kantian sense, it does not qualify for the ethical or legal status of a person.

Keywords: artificial intelligence, consciousness, apperception, person, agency

[a] Received on 28/03/2025 and published on 09/12/2025.

[*] Ludwig-Maximilians-Universität München, Chair of Metaphysics and Ontology, e-mail: sung-yeop.jo@outlook.com.

1. Introduction

Certain beings are typically regarded as conscious, yet it is still far from obvious what consciousness itself is or how the term ought to be defined. This long-standing philosophical challenge has acquired new urgency with the rise of artificial intelligence (AI) models that exhibit human-level cognitive capacities. Indeed, given the variety and quality of their outputs, asserting a fundamental difference between their intelligence and our own has become increasingly difficult.

The central task of this article, therefore, is to employ the framework of Kantian philosophy to clarify the essential conditions of human consciousness and subsequently determine whether these same conditions could, at least in principle, apply to AI. To address this issue, I will return to Kant's *Critique of Pure Reason* (CPR), which offers a profound analysis of the necessary conditions for consciousness. My focus will be the A- and B-Deductions, where Kant defends the actuality of human consciousness and its necessary unity against the empiricist critics, most notably David Hume. In response to Hume's bundle theory, which reduces the thinking self to a mere collection of fluctuating perceptions, Kant shifts focus to the mental act of synthesis, arguing that it is through this original act that disparate sense perceptions are combined into a unified, objective representation. The necessary unity of consciousness for objective representation – the principle of apperception (A117) – constitutes the very foundation of the Kantian theory of consciousness.

The challenge, however, is that the Kantian principle of apperception also seems applicable to advanced AI. An AI model that can cognize objects in a visible image, for instance, seems to perform the very synthesis of sensory data into objective representations that Kant describes. This raises two crucial questions: First, if we accept the premises of Kant's philosophy of consciousness, on what grounds could we then deny that an AI with this capability possesses a form of consciousness? Second, if we were to grant it consciousness, would it follow that such an entity also deserves ethical consideration?

To develop the full potential of the Kantian approach to the possibility of mechanical consciousness, this paper will proceed in two main parts. First, drawing on a selective reading of Kant's A-Deduction, I will establish that the Kantian criterion for consciousness is the mental ability to identify empirical objects by applying concepts to sense perceptions. I will then argue that advanced AI models that meet this criterion should be recognized as possessing a rudimentary form of consciousness. Second, I will sharpen and defend this claim by clarifying that this rudimentary consciousness is conceptually distinct from the more robust *self-consciousness* that Kant identifies as a necessary condition for genuine agency. This distinction explains why such a "conscious" AI would not have a claim to human-like rights, such as those for their intellectual contributions to research. I will develop this crucial difference between consciousness and self-consciousness by highlighting the specific developments in Kant's argument from the A- to the B-Deduction.

2. Kant's Concept of Consciousness in the A-Deduction

I. Synthesis as the Ground of Consciousness

If consciousness is a capability of the mind, it must manifest itself through mental activities. The crucial question, however, is what specific activities serve as this manifestation and how they can provide reliable evidence for the actuality of consciousness. In the A-Deduction, Kant provides a clear answer to this inquiry, writing as follows:

> Without consciousness that which we think is the very same as what we thought a moment before, all reproduction in the series of representations would be in vain. For it would be a new representation in our current state, which would not belong at all to the act through which it had been gradually generated, and its manifold would never constitute a whole, since it would lack the unity that only consciousness can obtain for it. […]
> For it is this *one* consciousness that unifies the manifold that has been successively intuited, and then also reproduced, into one representation. This consciousness may often only be weak, so that we connect it with the generation of the representation only in the effect, but not in the act itself, i.e., immediately; but regardless of these differences one consciousness must always be found, even if it lacks conspicuous clarity, and without that concepts, and with them cognition of objects, would be entirely impossible (A103; emphasis in original)[1].

These two passages mark the first time in the A-Deduction that Kant uses the term 'consciousness'. Here, he continues his analysis of objective cognition by examining the contribution of concepts to synthetic judgment *a priori*, building upon his preceding discussions of the synthesis of apprehension in intuition and the synthesis of reproduction in imagination. The necessary contribution of consciousness to cognition, Kant argues, is to produce a mental representation of an empirical object by applying a concept to a manifold of sense perceptions, which has already been provisionally combined by the prior syntheses of apprehension and reproduction. In the B-Deduction, Kant employs the example of cognizing a weighty body to illustrate this synthetic act of our mind. When one holds a body – such as a stone – and feels its downward pressure, one forms the judgment that *it is heavy*. However, as Kant points out, the raw sensations alone would amount to nothing more than a subjective sequence of perceptions, such as «If I carry a body, I feel a pressure of weight». Arriving at the objective representation, «It, the body, *is* heavy», requires an additional mental act to combine these underlying sense perceptions into a single, united representation (B142; emphasis in original). This demonstrates, for Kant, that forming an objective representation necessitates a synthetic combination of disparate perceptions – an act performed by consciousness.

The cautious nature of Kant's inference to the principle of apperception bears repeating. He does not directly address the presence of consciousness; rather, he infers its necessary unity from the given fact that a manifold of sense perceptions is *found combined* in individual representations of our mind. Instead of attempting to access consciousness firsthand, which Kant regards as too "weak" to be an object of distinct reflection, he posits the principle of apperception as a necessary condition for objective representation, taking the reality of that representation as his argument's starting point. This marks an

[1] I. Kant, *Critique of Pure Reason*, in *The Cambridge Edition of the Works of Immanuel Kant*, edited by P. Guyer and A.W. Wood, Cambridge University Press, Cambridge 1998.

argumentative strength of the Kantian approach to consciousness. Unlike the attempts to ground consciousness in empirical observation via inner sense – a method whose reliability is philosophically dubious – Kant's principle of apperception is not an empirical finding. Rather, it is the result of a transcendental inference to the necessary conditions for our objective representation. It is the undeniable actuality of the empirical representations of objects that justifies the positing of this 'weak' consciousness and its synthetic function.

In Kant's *Deduction*, consciousness is therefore presented not as a passive state, but as the act of synthesizing disparate sense perceptions into a single, conceptual representation. This active nature of consciousness is revealed in the application of *concepts* during the act of synthesis. As the preceding example in the B-Deduction implies, the *a priori* combination of sense data imparts objectivity to our representations. Their objectivity consists in the crucial distinction between a merely subjective, contingent association, where perceptions are simply simultaneous in the mind, and an objective synthesis, where those same perceptions are represented as necessarily combined in the object itself. The concept of a body, for instance, entails diverse properties, such as extension and mass. Therefore, by applying this concept, a conscious subject is able to unite the distinct sense perceptions of a certain extension and a certain mass into the single representation of one object. In performing «the synthesis of recognition in the concept» (A103), consciousness thus reveals itself as the faculty that makes objective recognition possible. It is through this specific synthetic act, which uses a concept as its rule, that an object can be empirically cognized[2].

The juxtaposition of Kant's inference to the necessary unity of consciousness in the A-Deduction with recent developments in AI technology leads us to a crucial philosophical question. To the extent that an AI is able to combine sensory data to identify objects, this ability seems *functionally similar* to the synthesis that Kant described. If we consistently apply Kant's line of reasoning, are we not then compelled to recognize this same capability in a non-human system as another form of consciousness?[3] Indeed, Kant's concept of consciousness in the A-Deduction provides a purely functional criterion for consciousness. On this view, consciousness is defined not by its biological substrate or private feeling but by the specific act of forming an objective representation through the application of concepts to sensory data. Therefore, any system that demonstrably performs this synthetic function – whether human or artificial – meets the condition for this rudimentary form of consciousness as established by the A-Deduction's logic.

3. Various Forms of Consciousness in Kant's Philosophy

I. The Logical Status of the Synthetic Act in the A-Deduction: Necessary or Sufficient for Consciousness?

[2] A significant point of scholarly debate regarding the A-Deduction is that Kant does not explicitly define what he means by "concept" in this context. This ambiguity has led to conflicting interpretations. For instance, Guyer argues that Kant is referring exclusively to the pure concepts of understanding, i.e. the categories. In contrast, Kitcher maintains that the argument also applies to, and even requires, empirical concepts, such as dog. See: P. Guyer, *Kant and the Claims of Knowledge*, Cambridge University. Press, Cambridge 1987, pp. 149-154; P. Kitcher, *Kant's Thinker*, Oxford University Press, New York 2011, pp. 126-142.

[3] This conceptual inquiry does not entail a technological evaluation of current AI models. The question of what conditions must be met for any being to be recognized as conscious remains philosophically valid, regardless of the actual capabilities of today's technology.

Before this 'Kantian' approach to the possibility of mechanical consciousness can be fully persuasive, several fundamental challenges must first be met. The first challenge is whether the mental act of combining a manifold of sense perceptions into an objective representation is a *sufficient* condition for attributing consciousness or merely a *necessary* one.

From a purely logical point of view, Kant's transcendental argument for the necessary unity of consciousness is sound if and only if the mental capability of synthetic combination is the sufficient condition for consciousness. In the A-Deduction, consciousness is described as *nothing more than* the functional foundation for the synthetic unity of sense perceptions as these are found combined in objective representations of our mind. Thus, Kant defines the transcendental apperception as «pure, original, unchanging consciousness», without which «no cognitions can occur in us, no connection and unity among them» (A107). The transcendental apperception functions as the foundation for all other forms of self-knowledge, which are acquired through empirical reflection on the contents of inner sense. For Kant, internal reflection is fundamentally empirical, as it is always situated in time and is therefore incapable of validating an *objective* representation of the self, in contrast to the *pure* consciousness of the transcendental apperception. Kant's transcendental apperception is thus "original" in a specific sense: it functions as the foundational premise for all subsequent philosophical discussions of consciousness and self-knowledge. The inference from the combination of sense perceptions in an objective representation to the transcendental apperception that makes it possible, therefore, provides the *only* viable ground for inferring the actuality of this "weak" consciousness, which cannot be confirmed through direct observation but only indirectly through a philosophical reflection on its *effects* – that is, the very representations unified by its synthetic act (A104). That is to say, the argument in Kant's A-Deduction arguably provides a functional criterion for consciousness. It suggests that a subject can recognize herself as conscious *insofar as* she finds objective representations in her mind – representations that necessarily consist of the combination of diverse sense perceptions. In the A-Deduction, Kant's argument attempts to derive the actuality of a unifying consciousness from the very concept of objective combination. For this argument to be sound, the capability for synthetic combination cannot merely be a *necessary* condition for consciousness but must be treated as a *sufficient* one. This is because if this capability were merely necessary, one could not validly infer the actuality of consciousness from the fact of combination *alone*.

In sum, Kant's A-Deduction establishes a sufficient condition for recognizing an entity as conscious: an entity qualifies as conscious *if* it possesses the intellectual capability to combine sense perceptions into a unified representation of an object. This criterion is the result of Kant's strategy for countering Humean skepticism about the self. Instead of searching for the self by means of inner sense, Kant asserts its necessary actuality solely based on the transcendental argument: for a manifold of sense perceptions to constitute an objective representation, they must first be combined by a single, unified consciousness.

II. Applying Kant's Concept of Consciousness to Non-Human Beings

The second challenge is whether Kant's analysis, which takes the form of a first-person reflection on consciousness, can be legitimately applied to other, non-human entities. A

starting point for this inquiry can be found in Kant's *Jäsche Logic*, which provides an illuminating description of the diverse levels of cognitive capacities.

> The *first* degree of cognition is: *to represent* something;
> The *second*: to represent something with consciousness, or *to perceive* (*percipere*)
> The *third*: *to be acquainted* with something (*noscere*), or to represent something in comparison with other things, both as to sameness and as to *difference*;
> The *fourth*: to be acquainted with something with *consciousness*, i.e., to *cognize* it (*cognoscere*). Animals are acquainted with objects too, but they do not *cognize* them.
> The *fifth*: to *understand* something (*intelligere*), i.e. to cognize something through *the understanding by means of concepts*, or *to convince*. One can conceive much, although one cannot comprehend it, e.g. a *perpetuum mobile*, whose impossibility is shown in mechanics.
> The *sixth*: to cognize something through reason, or *to have insight* into it (*perspicere*). With few things do we get this far, and our cognitions become fewer and fewer in number the more that we seek to perfect them as to content.
> The *seventh*, *finally*: to comprehend something (*comprehendere*), i.e., to cognize something through reason or a priori to the degree that is sufficient for our purpose. […]" (Jäsche Logic, IX:64-5; italics in original)[4].

In this passage, Kant draws a notable distinction between *noscere* (to be acquainted with something), a capability he also attributes to animals, and *cognoscere* (to be acquainted with something *with consciousness*). Since a basic consciousness of an object is already presupposed in mere acquaintance, the consciousness that elevates this to genuine cognition must, arguably, be of a different order. In the CPR, Kant defines cognition (*cognitio*) as an objective perception—this is a form of knowledge that is either intuitive or conceptual (CPR A320/B376–7). While animal consciousness can, therefore, discern sensible objects by comparing their similarities and differences, it lacks the ability to (re-)cognize them as *objective* beings, distinct from its own *subjective* awareness[5]. The «synthetic recognition in a concept», which Kant analyzes in the third section of the A-Deduction, corresponds to what he calls *intelligere* in the *Jäsche Logic*. For Kant, human knowledge is discursive: it always requires concepts to synthesize the manifold of sense perception (CPR A68/B93). Human consciousness can not only cognize objects (*cognoscere*) but also understand them (*intelligere*) through its conceptual capacity.

Kant's analysis of various forms of consciousness makes it clear that he accepted the possibility of non-human consciousness such as that of animals. The crucial limitation of animal consciousness is that it lacks a sense of objectivity. In human understanding, by contrast, this objectivity is guaranteed by the application of concepts, as was already demonstrated in the earlier comparison between synthetic combination and mere association. This, in turn, allows us to address the possibility of mechanical consciousness using Kantian framework, since such a system also appears to discern sensible objects using the same kinds of concepts that human beings do[6]. That is to say, if an advanced AI model

[4] I. Kant, *Jäsche Logic*, in *The Cambridge Edition of the Works of Immanuel Kant*, edited by P. Guyer and A.W. Wood, Cambridge University Press, Cambridge 1992.

[5] For a discussion of Kant's illustration of animal consciousness with the example of an ox which can distinguish a stall from a door: P. Kitcher, *Kant's Thinker*, cit., pp.115-119.

[6] In his article examining the possibility of a computer system that satisfies the Kantian conditions for consciousness, Evans also emphasizes the crucial role that concepts play. For Evans, as for Kant, it is the application of concepts that combines disparate sense perceptions into a single, unified reservation, see R.

can cognize sensible objects by applying concepts, such as the concept of body, then on a consistent application of Kantian terminology, it must be recognized as conscious, since conceptual understanding (*intelligere*) necessarily presupposes consciousness.

III. Consciousness versus Self-Consciousness: A Critical Limitation

Not every conscious being is necessarily self-conscious. This crucial difference, which is not sufficiently explicated by Kant himself, requires careful deliberation, particularly when considering the possibility of non-human consciousness. The unique ethical position of human beings is fundamentally grounded in their status as persons. This status, in turn, is typically attributed to their self-consciousness – a mental capacity that animals are considered to lack. It follows, therefore, that if an AI model were regarded as self-conscious, it too would be eligible for a similar ethical status.

According to Kant's analysis of self-consciousness in the B-Deduction, while the synthetic combination of sense perceptions may be a sufficient condition for consciousness in a basic Kantian sense, the mere ability to (re-)cognize objects through such a combination is not, in itself, a sufficient ground for classifying AI as self-conscious. Kant's B-Deduction takes the concept of self-consciousness as its starting point, whereas the A-Deduction only arrives at this concept as its conclusion. In the second section of the B-Deduction (§ 16), Kant introduces his famous principle that the pure representation 'I think' must be able to accompany all of one's representations. He names this reflective awareness of a constant, unified self "original apperception", a form of *self-consciousness* that he defines at length[7]. The project of the B-Deduction, then, is to analyze the fundamental structure of this self-consciousness by examining the principle of the necessary accompaniment of the 'I think' and its philosophical implications. This represents a significant departure from the A-Deduction, in which the analysis does not necessarily presuppose that the cognizing self is self-conscious. In this notable move at the beginning of the B-Deduction, Kant explicitly identifies the agent of the synthetic combination with the first-person pronoun 'I,' whereas in the A-Deduction, such a self is at first only an implication of the transcendental analysis of objective cognition, emerging explicitly only toward the argument's conclusion[8].

However, possessing a certain mental capacity and ascribing its results to oneself are two distinct acts, as the latter is not necessarily entailed by the former. It is arguably this very distinction that makes Kant's cautious, step-by-step approach in the A-Deduction plausible. This, in turn, requires us to distinguish between the 'mere' consciousness and

Evans, *The Apperception Engine*, in H. Kim, D. Schönecker (eds.), *Kant and Artificial Intelligence*, Walter de Gruyter GmbH, Berlin-Boston 2022, p. 67.

[7] «I call it the *pure apperception*, in order to distinguish it from the *empirical* one, or also the *original apperception*, since it is that self-consciousness which, because it produces the representation *I think*, which must be able to accompany all others and which in all consciousness is one and the same, cannot be accompanied by any further representation» (B132; emphasis in original).

[8] Kitcher's six-step reconstruction of the A-Deduction's argument offers a precise analysis of Kant's strategy. According to her reading, the fourth stage of the argument establishes that one becomes aware of representational states and their necessary connections. It is only after this, in the fifth stage, that the thinking self is introduced. Kitcher formulates this step as follows: «When representational states are recognized as necessarily connected together, they are recognized as instances of the I-rule». See P. Kitcher, *Kant's Thinker*, cit., pp. 141-142.

self-consciousness. The former serves simply as a necessary condition for our empirical cognition and does not, in itself, involve a first-person perspective, since the synthetic combination of perceptions into an objective representation implies only that the mental act must be ascribed to a certain conscious subject; it does not entail that the subject is aware of itself performing that act. The latter, by contrast, is expressed in the principle that the 'I think' that must be able to accompany all of one's representations. The principle of the B-Deduction, therefore, necessarily presupposes an additional act of self-reflection: the recognition that all one's representations must be regarded as one's own. This mental act of self-ascription is purely *spontaneous*. The pure awareness of the 'I' is not something given by the senses; rather, the subject attains self-consciousness by reflecting on formal conditions of unity within her own representations[9]. In the B-Deduction Kant proceeds directly from his insight into the necessary accompaniment of the 'I think,' which allows him to address the concept of self-consciousness. This stands in sharp contrast to his more cautious, multistep approach in the A-Deduction, which begins with an analysis of empirical cognition and only derives self-consciousness as a final conclusion.

While Kant's transition to self-consciousness may seem abrupt, it becomes understandable given that the discussions in the A- and B-Deductions are always about consciousness from a first-person perspective. The appeal of this move stems from a subtle phenomenological fact: as a subject, *I* experience an intuitive awareness of a thinking self that accompanies all the objects of my thought. In the B-Deduction, Kant advances to this more robust form of self-consciousness based on a new starting point that the thinking 'I' must always be presupposed as the subject to which all of one's representations belong. The essence of Kantian self-consciousness is this original act of self-ascription: the awareness that all of one's diverse representations belong to a single, unified 'I.' Such an awareness of the 'mineness' of one's representations is a subtle yet constant feature of cognition, which one can bring into focus through an act of reflection.

Kant clarifies the immediate nature of self-ascription in his critique of the metaphysical doctrine of the soul, developed in the *Transcendental Dialectic*. There, he maintains that this mental act of self-ascription is not a logical inference but is, rather, the direct expression of cognition's innate, formal structure, which is grasped through the purely spontaneous reflection on one's own mind. In other words, transcendental apperception does not refer to a substantial self that can be observed or proven to exist; rather, it signifies the transcendental necessity that all of one's representations belong to a single 'I think.' Kant himself describes this principle in the following passage:

> This *I* would have to be an intuition, which, since it would be presupposed in all thinking in general (prior to all experience), would, as an intuition, supply *a priori* synthetic propositions if it were to be possible to bring about a pure rational cognition of the nature of a thinking being in general. Yet this I is no more an intuition than it is a concept of any object; rather, it is the mere form of consciousness, which accompanies both sorts of representations and which can elevate them to cognitions only insofar as something else is given in intuition (A382; emphasis in original)[10].

[9] Strawson makes a similar point, identifying «the necessary unity of consciousness» with «the possibility of self-consciousness». See: P. F. Strawson, *The Bounds of Sense*, Routledge, London-New York 1966, pp. 9-10.
[10] Kant makes this same point in the second edition of Paralogism, arguing that the I of 'I think' is not an object of intuition, but is, instead, the formal expression of self-consciousness (B412).

Notably, the act of self-ascription is only possible for representations contained within one's own mind: one cannot assert that another person thinks in the same way that one does. This is because while an individual ascribes thoughts to oneself through a direct, first-person awareness, one can only ever infer the thoughts of others from a third-person perspective. A subject ascribes one's own thoughts based on a direct and pure awareness of own mental acts—an awareness that constitutes self-consciousness, which Kant's B-Deduction takes as its starting point. One has no such immediate access, however, to the mental states of others; their thoughts can only be inferred, never known. Following the argument of Kant's B-Deduction that the awareness of the 'I think' is a non-inferential and uniquely first-person act, it is therefore impossible to derive any direct knowledge of self-consciousness in other beings.

We can now return to our initial question: Does an AI model qualify as self-conscious if it is capable of (re-)cognizing objects by combining sensory data with concepts? Based on the Kantian distinction between consciousness and self-consciousness, the answer is a clear no. The synthetic capability of our mind, while perhaps a sufficient condition for consciousness, is not in itself sufficient to justify the attribution of self-consciousness. That is to say, one cannot know, in principle, whether an AI system experiences the outputs it produces as 'its own.' For Kant, however, self-consciousness is defined by this very capacity: the original and spontaneous recognition that one's representations are indeed one's own[11].

The sharp distinction between consciousness and self-consciousness is crucial, because it provides a principled basis for resisting the conclusion that AI systems should be granted moral and legal rights. In Kant's practical philosophy, such rights are granted only to beings who qualify as *persons*. In his *Religion within the Boundaries of Mere Reason*, Kant defines a person as a rational being who can be held responsible for his or her actions (Rel VI:26). This responsibility, in turn, arises from the mental capacity to act according to a subjective rule, or *maxim*, that the individual gives to themselves. Kant's categorical imperative thus requires a rational agent to consider whether the maxim grounding their action could be willed as a universal law. The very act of willing a maxim to become a universal law, however, necessarily presupposes a first-person perspective, which constitutes the basis for personal responsibility. Therefore, an AI cannot be considered a genuine moral agent, because its actions are based on pre-programmed rules given to it externally, not on maxims it gives to itself. Furthermore, it lacks the capacity to formulate its own maxims, a process that requires a uniquely first-person act: the ability to test whether *one's own* rule of action could be willed as a universal law. This is an act of autonomous, *self-conscious* agency that a purely rule-following system cannot perform, since this act presupposes the capacity to distinguish one's own *subjective* maxim from an objective

[11] Several scholars, many of whom are contributors to the same edited volume, already pointed to AI's inability to self-ascribe its output as the primary ground for denying it genuine agency. See: S. Baiasu, *The Challenge of (Self-)Consciousness: Kant, Artificial Intelligence and Sense-Making*, in H. Kim, D. Schönecker (edited by), *Kant and Artificial Intelligence*, Walter de Gruyter GmbH, Berlin-Boston 2022, p. 125; L. Benossi S. Bernecker, *A Kantian Perspective on Robot Ethics*, in H. Kim, D. Schönecker (edited by), *Kant and Artificial Intelligence*, Walter de Gruyter GmbH, Berlin-Boston 2022, p. 157; R. Evans, *The Apperception Engine*, in H. Kim, D. Schönecker (edited by), *Kant and Artificial Intelligence*, cit., p.85; Walter de Gruyter GmbH, Berlin-Boston 2022, p. 157; D. Schönecker, *Kant's Argument from Moral Feelings: Why Practical Reason Cannot Be Artificial*, in H. Kim, D. Schönecker (edited by), *Kant and Artificial Intelligence*, Walter de Gruyter GmbH, Berlin-Boston 2022, pp. 180-181.

moral law.

It also follows that an AI lacking self-consciousness cannot be recognized as a bearer of legal rights. This is because, Kant's *Universal Principle of Right*, the very ground of such rights, applies only to actions resulting from "the free use" of choice. The principle states that an action is right if it can coexist with «the freedom of everyone» (MM; VI:231). An entity without self-consciousness cannot be a subject of rights, since it is not an autonomous agent capable of free choice. For something to be an object of free choice, a subject must first be able to posit it as an end for herself. This capacity to set ends, however belongs to the faculty of self-consciousness and is thus fundamentally distinct from the mere combination of sense perceptions[12].

4. Conclusion

The most important findings of this article can be summarized in the following two theses:

A. If an AI is capable of identifying objects in images by applying concepts, then it satisfies the Kantian condition for consciousness.

B. Acknowledging a being as conscious does not, however, entail granting of moral and legal rights to it.

[12] This Kantian analysis, of course, is not intended to exclude other philosophical approaches to the ethical status of AI. Indeed, there is a growing body of research dedicated to evaluating AI's potential as an artificial moral agent (AMA). For some key contributions to this discussion, see: K. E. Himma, *Artificial Agency, Consciousness, and the Criteria for Moral Agency: What Properties Must an Artificial Agent Have to be a Moral Agent?*, in «Ethics and Information Technology», n. 11, 2009, pp. 19-29; M. Hunyadi, *Artificial Moral Agents. Really?*, in Jean-Paul Laumond, Emanuelle Danblon, Céline Pieters (edited by), *Wording Robotics: Discourses and Representations on Robotics*, Springer, Toulouse 2019, 59-69; A. Martinho, A. Poulsen, M. Kroesen, C. Chorus, *Perspectives about Artificial Moral Agents*, in «AI and Ethics», n. 1, 2021, pp. 477-490; M. Arvan, *Varieties of Artificial Moral Agency and the New Control Problem*, in « Humana.Mente - Journal of Philosophical Studies», n. 15, 2022, pp. 225-256; S. Nyholm, *Is Academic Enhancement Possible by Means of Generative AI-Based Digital Twins?*, in «The American Journal of Bioethics», n. 23, 2023, pp. 44-47; J. Graff, *Moral Sensitivity and the Limits of Artificial Moral Agents*, in «Ethics and Information Technology», n. 26, 2024, pp. 1-13; R. van Woudenberg, C. Ranalli, D. Bracker, *Authorship and ChatGPT: a Conservative View*, in «Philosophy & Technology», n. 37, 2024, pp. 1-26; J. McLoughlin, *The Work of Art in the Age of Artificial Intelligibility*, in «AI & Society», n. 40, 2025, pp. 371-383.