Defining Artificialism

Abstract

This paper focuses on epistemological artificialism and divides it into four categories in terms of how strong (AI is the only source of decision making) or weak (AI is the best source of decision making) and how narrow (social but not personal) or broad (both social and personal related) they are. Two central arguments against artificialism, the (false) dilemma and self-referential incoherence, are analyzed. Of the four types of epistemological artificialism, three can deal with these counterarguments using two methodological principles: evaluation of epistemic reliability and epistemic opportunism. There is hope that these considerations will steer the discussion on artificialism toward more fruitful pastures in the future. For example, there are interesting methodological considerations concerning what evaluability of reliability and epistemic opportunism entail.

Keywords

Artificial intelligence; AI; artificialism; artificiality; automation; automated systems; decision; decision-making; human factor.

Introduction

Discussion on AI has recently gained more exposure in philosophy and theology (*1*), (*2*), (*3*), (*4*), (*5*), (*6*), (*7*). Originally, the proposed concept of the term "artificialism" created for the purposes of this study was introduced as a possible parlance by those who are critical of excessive trust in AI overreliance on it, and even today artificialism is widely understood even if it was not previously defined, especially by its critics, as offensive or dismissive toward philosophy, theology, or other fields outside logical, computational, and mathematical science (*8*). A relatively common understanding of its goal is, in some sense, the reduction of all valid choices to those which are mostly (or only) approved by AI or automated systems and most (or complete) preference for artificial intelligence and automated systems over humans in any field. Through this angle, it is not difficult to understand why so much of the discussion on artificialism is carried out by its opponents. Here, however, the danger of bias is imminent. If the opponents of a view are its main theoreticians, then it is rather probable that the principle of charity will be violated at some point. Despite the fact that artificialism is often defined in a disparaging way, some authors have recently adopted this concept as a manifest (*9*), (*10*), (*11*), (*12*). Given the predominant status of the debate, the topic is riddled with misconceptions.

In this paper, I seek to rectify this situation. I start by going through some uncharitable definitions of artificialism. Then I focus on epistemological artificialism and divide it into four types. I consider two central global arguments against artificialism and show that of the four varieties, three can go on unscathed. I also suggest two methodological principles to which a proponent of artificialism can appeal: epistemic evaluability of reliability and epistemic opportunism. This shows that there are viable forms of epistemological artificialism.

Since epistemological artificialism can be defended utilizing certain methodological principles, further critique of artificialism must take a stand on those principles. An opponent of

artificialism has to consider whether reliability has to be something I can evaluate and whether the decision should be reliable. An advocate of artificialism needs to show that rational practice, in fact, upholds these principles. The debate regarding artificialism thus transforms into a debate on the methodology of ethics and justice.

Uncharitable Definitions of artificialism

In its current usage, the concept called here "artificialism" is commonly taken as a frightening term. This understanding also extends to the most general definitions of artificiality. For instance, the four most common definitions for artificialism are:

1. Kind of automation achieved through the use of artificial intelligence and automated systems applied to any field;

2. methods and attitudes typical of or attributed to the human factor elimination and or the delegation of decision making to artificial intelligence and automated systems, or even an approach leading to partial or complete AI takeover;

3. an unwarranted, exaggerated, or unjustified trust in the efficacy of the methods of AI and automated systems applied to any field;

4. any preference for artificial intelligence and automated systems over humans or even possible discrimination associated with it.

Artificialism is most often blankly dismissed because of the immediate undesirable consequences that its well-known concepts imply. In more in-depth discussions similar problems have consistently arisen because the critics of artificialism and other commentators systematically understand the term in an uncomplimentary fashion. For example, James Barrat once

said about artificialism as "I'm increasingly inclined to think that there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish. I mean with artificial intelligence we're summoning the demon" (*13*). Indeed, the most common definitions of artificialism typically take it to, in one way or another, exceed the *proper* limits of rationality. Hence, artificialism is often considered to amount to unwarranted, exaggerated, or unjustified trust in AI and automated systems in some way (*14*).

It is good to note that when the opponents of artificialism define artificiality, they usually have in mind something actually closer to *AI takeover* (*15*), (*16*), (*17*). An AI takeover, characterized by Tristan Harris, is "By allowing algorithms to control a great deal of what we see and do online, such designers have allowed technology to become a kind of "digital Frankenstein", steering billions of people's attitudes, beliefs, and behaviors". Occasionally, some critics of artificial intelligence and artificiality in general even explicitly state that any automation can be a form of AI takeover, as shown in these studies (*18*), (*19*). In the final section, we return to the differences between artificialism and AI takeover.

Despite the prevalence of disparaging concepts of artificialism, some may start to endorse the concept as a badge of honor if it has not already started, just the same way as with the concept of "scientism" (*20*), (*21*). This would not be intelligible without a more neutral definition of artificialism. It is nonsensical to think that someone would declare: "According to the view that I defend, the *proper* limits of artificial intelligence and automated systems should be exceeded". For example, Marvin Minsky, instead of claiming anything overblown, simply claimed that "artificial intelligence is the *science* of making machines do things that would require intelligence if done by man" (*22*). Or in a quote by Ginni Rometty: "Some people call this artificial intelligence, but the reality is that this technology will enhance us. So instead of artificial intelligence, I think we'll *augment* our intelligence". Along these lines, we have a more fruitful definition of artificialism.

I think the most plausible forms of artificialism are epistemological (*23*). In fact, it seems that most proponents of artificialism accept an *epistemology first* attitude, according to which epistemology should determine or at least guide one's ontological or other commitments (*24*). Such epistemological artificialism is usually defined by its opponents as the idea that *only* artificial intelligence and automated systems can obtain genuine or reliable results, as shown in this study (*25*). However, Russell (*26*), for example, merely takes AI and automated systems as the best methods for decision making, and Gershman has made practically the same claim (*27*).

Before going any further, one should note that all of the definitions presented thus far have been formulated by the critics of artificiality. This does not mean that those definitions are necessarily erroneous, but one common bias in them is that artificialism is associated with the *fear* of the unknown or changes in life that challenge their adaptation abilities or accepted values (*28*), (*29*). The reason for the opponents of artificialism to define artificialism in this broad way is that, according to them, artificiality undermines their foundations of life. If things other than artificiality are also viable sources of decision-making, then artificialism does not exclude practically anything. Philosophers and theologians could also claim to have equal authority on some issues, as the computer scientist does, and in the end, nothing would be affected by the scientific project. However, I will show that this is not the case. Instead, there can be nontrivial conceptions of artificialism based on a narrower view of artificiality. Therefore, instead of understanding the concept "artificiality" as broadly referring to any field, as is commonly done, the proponent of artificialism can conceive it more narrowly and categorized. Consequently, I think it is fruitful to divide epistemological artificialism into additional subcategories.

First, epistemological artificialism can be divided into *narrow* and *broad* varieties. The narrow versions state that AI and automated systems only function as proper sources of decision making and the like in the personal and social life of a human, as well as the technical purposes for which AI and automated systems were originally created. In other words, it understands the

term “artificiality” in a restricted sense. The broad version, on the other hand, endorses a wider conception of artificiality that encompasses philosophy, ethics, justice, etc.

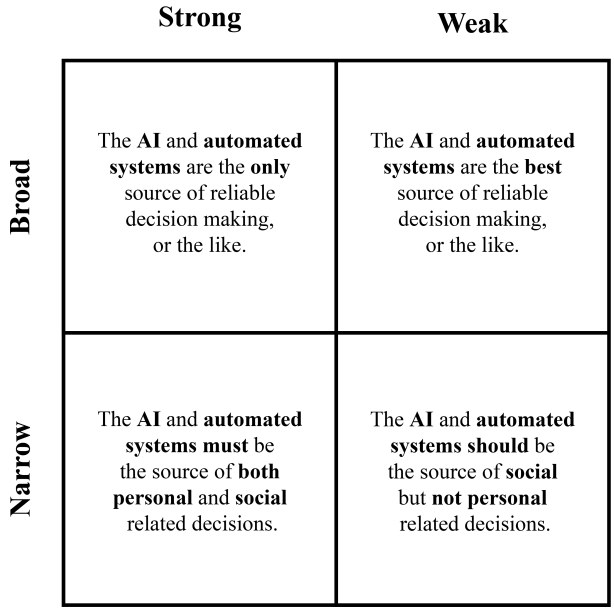

Figure 1: Four types of epistemological artificialism

Second, I make another distinction within epistemological artificialism, the separation of *weak* and *strong* artificialism. Strong artificialism declares that *only* artificiality can function as a proper source of decision-making, justification, etc. In contrast, weak artificialism states that artificiality is only the *best* source of decision making, justification, or the like. These four categories can overlap, as presented in the two-by-two diagram in Figure 1.

Strong artificialism

This type of artificialialism suggests that the artificial intelligence should always make all decisions for the human, regardless of the type, and that the human should only make their own decisions when the AI is not applicable.

Weak artificialism

This type of artificialism suggests that people can make their own decisions if they are based on the known best algorithm and use others, a better one, if discovered at the moment.

Broad artificialism

This type of artificialism suggests that artificial intelligence should make decisions instead of humans, but only ones that are socially related (such as voting, politics, etc.).

Narrow artificialism

This type of artificialism suggests that artificial intelligence should make decisions instead of humans, both socially and personally related (such as personalization, customization, etc.).

We can find examples of representatives for each of these categories, since epistemological artificialism comes in many shapes and sizes, as the various quotations clearly demonstrate. For example, a proponent of artificial intelligence may take AI as the *only* source of social-related decisions (narrow-strong) or think that automated systems are simply the *best* sources of decision making (broad-weak). Often, critics of artificialism refer to some of the mentioned proponents of artificialism. It is important to note that even the weak and broad forms of artificialism can retain the distinction between the decisions with or without a human factor, as well as between rational and irrational methods of decision making. This is discussed in detail in the following sections.

It is good to note that AI discrimination could only be possible if done based on human errors or intentions. Of course, it is an interesting question how it can start and its further consequences, but due to its scope, it will have to be relegated to other work. For now, it suffices to state that, at least in some cases, we can see the prerequisites for it, but the reasons for similar outcomes became trivial when subjected to further rational criticism (*30*), (*31*), (*32*), (*33*). There

is no doubt that errors are possible at every stage of AI implementation and development (*34*), (*35*), (*36*), (*37*), (*38*). I only need to point out that even if such an outcome were possible (even if it is unlikely), it still does not change the fact that it could only have been started due to our own mistakes or intentions and that we are potentially capable of dealing with it properly, while the assessment of our possible actions related to it, due to its scale, will have to be relegated to another work. If we seek to create pure intelligence, we must be aware of the problems and dangers that our own intelligence brings (*39*), (*40*), (*41*), (*42*).

To demonstrate how artificialism can be feasible, we will look into two central global arguments against artificialism: the (false) dilemma of artificialism and the allegation that artificialism is self-refuting.

First Objection: The (False) Dilemma of Artificialism

One of the main global objections to epistemological artificialism is based on the claim that artificialism is built on nonobjective grounds. For example, certain metaphysical or strictly sensual background assumptions are argued to be necessary for rational decision-making.

Due to such dependencies, it is insisted that the proponents of artificialism are forced to face the following dilemma:

1. The proponent of artificialism must reject or accept sources of decision making affected by the human factor.

2. If the sources of decisions affected by the human factor are rejected, then all decision-making inquiries are rendered justified because objectivity necessarily desupposes them.

3. If they are accepted, then the proponent of artificialism has to accept sources of decisions, affected by the human factor as justified.

4. Thus, the proponent of artificialism has either to reject all non-objective inquiry as unjustified or to dilute it in a way that would render the thesis of artificialism impotent, because non-objectivity would encompass *all* kinds of ground of decision affected by the human factor.

The dilemma is based on a typical transcendental argument: the necessary conditions of objectivity are incompatible with the human factor. To summarize, the adherent of artificialism has to choose between two poor options: either (1) the adherent has to reject the reliability of objective decisions because it rejects all sources of decision making affected by the human factor, or (2) the adherent has to accept all sorts of sources of decision making as reliable, and this would make artificialism lose all of its bite.

In fact, under closer scrutiny, the dilemma turns out to be false. The dilemma rests on the claim that the decision has to rely on extra-objective (irrational) sources. In particular, there are two such purported sources. These are (a) affected by the human factor, possibly metaphysical, background assumptions, and (b) non-objective sources of decision making. The problem here is that it is not exactly clear why these origins are non-objective or why we have not to rely on them.

Consider first option (a), the metaphysical background assumptions. However, these are not necessary assumptions for decision making. One does not have to assume that AI can achieve knowledge of the best decision. Automation can merely start with the *hypothesis* that some kind of improvement could be achievable due to the elimination of a human factor. For all practical purposes, this hypothesis would merely state that there are at least some prejudices to be found. This hypothesis could be tested by simply trying to make an objective decision based on rational logic and empirical knowledge with automated means. If it is impossible to achieve, then the efforts would just be in vain. But *hoping* that something could be automated is not the same as *believing* that it can definitely be done. The AI researcher can carry out the

inquiry *as if* the objective decision could be automated, and hope that this is so, without making any commitments to it actually being the case. In fact, this is very similar to how hypothesis testing is often performed in actual scientific practice. Furthermore, if the test turns out to be successful, then the additional assumption that the knowledge obtained is about an "ideal way" of decision making is irrelevant. Further argumentation is therefore needed to show that such extra-objective assumptions are in fact necessary. In particular, if they are claimed to have any effect on actual objective practice, then this claim should be argued for in detail.

Now consider option (b), the non-objective sources of decision making. As already noted, there are clearly nonobjective sources of decision, such as senses and emotions, when taken alone or implied to fields where only formal logic could be valid. It is rather obvious and empirical that most decisions are based on our senses and emotions and that they, case by case, both can or cannot lead us astray. However, these decision sources are not considered rationally justified unless they stand up to rational criticism. And the problematic of such a way of decision making today is not a subject of any doubt, since today we all take into account the fact that humans do not always act rationally (*43*), (*44*), (*45*), (*46*), (*47*), (*48*), (*49*), (*50*), (*51*), (*52*).

The previously mentioned human capacities enable artificiality even though they are somewhat unreliable. However, this does not lead to the unreliability of artificiality through a simple transitive relation. This is because an important component of all rational endeavors is error recognition and correction. Even if some of the potential flaws are, of course, caused by the partial unreliability of human capacities.

The relation between human capacity and rational endeavor is not unidirectional. Like our general cognitive capacities enable scientific research, scientific research enables the improvement of our somewhat unreliable cognitive capacities. The opponent of artificialism might object that the process of error correction itself has not been given reliable grounds: it should lean towards some other infallible principles outside our unreliable cognitive capacities. This

kind of criticism would be based on a faulty conception of rational decision-making. For example, the process of error correction in science is iterative: a community of researchers seeks to identify sources of error and fix them in multiple passes, and within each pass, the researchers examine how the corrections improved the reliability of their theory in terms of describing, predicting, and so on. There is no prior guarantee that this process will yield results, but that does not preempt the attempt.

Of course, common sense needs to have some epistemic value in order to serve as a foundation for rationality. It would be rather problematic to insist that all common sense would be totally unreliable. Furthermore, what we take as common sense does not appear to have well-delimited boundaries and appears to vary greatly in its epistemic status, which seems to strongly indicate that it is not all totally unreliable (*53*), (*54*), (*55*).

Rationality then does not need to be able to categorize sources of belief as either scientific or non-scientific. Instead, what is required is that the given source of a given belief can be checked for errors and biases or epistemically evaluated in terms of reliability. So, it is not just about being reliable, but about how reliable and under what conditions. Therefore, non-objective sources would be sources that cannot be epistemically evaluated or have been evaluated to be totally unreliable. If an opponent of artificialism wants to turn the false dilemma into a real one, then we will have to argue for why such non-evaluable sources are necessary.

Second Objection: Artificialism Is Self-Refuting

The other major criticism raised against artificialism is that artificialism is self-referentially incoherent or self-refuting. The rough idea of the argument is the following: According to artificialism, one can rationally accept only those decisions that are formulated by objective means. Assuming that the proponent of artificialism is inclined to follow his own principles, artificialism needs to be justified objectively. The critics of artificialism claim that such justification is

nowhere to be found and, even more pressingly, that it is impossible to make a purely rational case for artificialism. Therefore, artificialism cannot meet its own standards.

The structure of the argument can be given as follows:

1. It is rational to accept artificialism only if artificialism is justified on the basis of objective sources and nothing else.

2. Artificialism is not, and cannot be, justified by objective sources and nothing else.

3. (C) It is not rational to accept artificialism.

The first premise follows from the assumed definition of artificialism, according to which it is rational to accept X only if X is justified on an objective basis and nothing else. The premise is formed by simply substituting the variable X with artificialism itself.

The second premise, in turn, is based on the conception that, at least thus far, there are no rational grounds for endorsing artificialism. If artificialism is to be rationally validated, then it needs to be a rational hypothesis that is properly tested and confirmed. The critics of artificialism have formulated this challenge in two ways. First, they have pointed out, there is no empirical or formal research leading to the confirmation of artificialism. I call this the weak version of the second premise. Second, some have argued that such research cannot even be done. This is a strong version of the premise. On the basis of this kind of argumentation, the opponents of artificialism commonly take artificialism to be a philosophical doctrine instead of a rational one, or, at the very least, they believe that artificialism is dependent on unarticulated and implicit philosophical assumptions.

Now, it is immediately clear that a proponent of weak artificialism can reject premise (1). The premise states that it is rational to accept artificialism only if it is justified on an objective basis and nothing else. Recall that weak artificialism merely declares that AI and automated systems are simply the best way of decision making; it does not have to be the only one. This

enables advocates of artificialism to use methods such as common sense to justify their endorsement of artificialism. Logically, it is still required that their methods are not in contradiction to rational inquiry, even if they do not for some reason deserve to be called artificiality.

For example, let us assume that expert advisors would be a necessary criterion for rationality. (I am not claiming that this is actually the case; expert advisors are merely used as an example to illustrate my point.) So, when someone is reasoning with common sense, this would not yet count as artificiality, although there would not be anything wrong with thinking. In other words, objective decisions could be gathered without artificial practice (this seems evident when one examines everyday life). But, according to weak artificialism, there is no such form of objective decision for which artificiality would not be the best form of inquiry. Valid everyday reasoning could always be turned to artificiality if it were to be subjected to objective evaluation, in this example, by expert advisors. (Assume that there would be no other required criterion for rationality or that the reasoning already fulfills all other criteria.) Therefore, just by using common sense, artificialism can already be justified. Hence, a proponent of weak artificialism can rather easily avoid the accusation of self-referential inconsistency.

Now consider premise (2). Perhaps it is easiest to start by challenging the stronger version of the premise, namely, that it is impossible for artificialism to be justified on the basis of rationality. This only requires that artificialism is viable as a rational hypothesis. In other words, artificialism needs to be a type of claim that could have objectively appropriate evidence against and for it. It seems rather evident that there is a lot of positive evidence for the epistemic success of automated methods. In fact, proponents and opponents of artificialism seem to agree that artificiality enjoys a robust track record of generating rational decisions. Having such a record is positive evidence in support, and this evidence is therefore clearly possible.

Where automated methods are applicable, we can compare how well they perform relative to some other methods, given some epistemic criteria. Such criteria can be chosen on pragmatic

grounds, but they should not be arbitrary. This means that different individuals should not systematically arrive at different conclusions employing the same criteria. With nonarbitrary criteria for comparisons, we can potentially have evidence for the claim that the automated methods are the most reliable or only reliable ones in particular cases. Whether we at present have such evidence is irrelevant for the point that such evidence is nevertheless possible, and consequently, we can treat artificialism as a rational hypothesis. This suffices to avoid self-refutation. However, if the required evidence is not yet gathered, then full-blown commitment to artificialism would not be justified at present. This is because the inferiority of other epistemic practices is not yet warranted. Still, someone could consistently adopt artificialism as their own epistemology but merely as a working hypothesis.

However, one should note that there is a caveat here. So far, I have discussed only the cases where automated methods can be applied. It is also necessary to consider the cases where these methods are not applicable. Clearly, if we cannot use automated methods, we cannot have evidence to support the claim that these methods would have been the best or the only ones. However, the rational claim refers to decision making, so one only has to worry about these kinds of case if they are genuine ones. To make the case for self-referential incoherence, an opponent of artificialism then has to first show that we have rational cases where automated methods are not applicable. Here, further argumentation is needed. If, however, one would indeed manage to establish such a case, there is still no reason to count it as irrational. It would, after all, count as demonstrative and thus reliable, rational, just exactly what artificiality is after. So, there is no reason why artificialists would not incorporate it into decision making.

An adherent of artificialism can then argue that a strong version of premise (2) is wrong because artificialism can be presented as a rational hypothesis. This merely required demonstrating that objective evidence for and against artificialism can be gathered.

This brings us to perhaps the biggest problem with critiques of artificialism. The critics

argue against both all versions of artificiality or only certain ones. Despite this, the same abstract principles are often used regarding argumentation and inference. General rules of inference are applied in all rational decisions, such as trying to exclude other possible conclusions and making the inferences explicit for evaluation. Triangulation, obtaining robust results of the same phenomenon by different means or through independent sources, is always considered to improve the reliability of the judgment. Same with an error analysis as well.

The fundamental principles of proper decision reasoning, such as those just mentioned, are always in place. They are only applied in different circumstances. No well-defined methodological line can easily be drawn between artificiality and elimination of a human factor, nor is it necessary. Both are rational methods to an equal degree.

When examining the actual enterprise of artificiality, in all its variety, the only epistemic boundary condition or methodological constraint appears to be *epistemic opportunism*: to use the practices that work evaluably to obtain reliable decisions and to abandon those that do not. From this perspective, those who define artificiality in the narrow sense, whether they be proponents or opponents of artificiality, impose arbitrary constraints that are alien to rationality. It is not the object of judgment that defines whether or not a valid decision can be made, but the methods that are deemed proper for the object of interest. If this is accepted, then the two main global arguments fall.

Note that assuming a decision to be reliable, epistemic opportunism in itself already validates artificialism. If artificiality is epistemically opportunistic in the way presented above, it follows directly that artificiality is the *best* or *only* way of making evaluably reliable decisions. That is, if one accepts that artificiality uses or should use the methods that evaluably work for obtaining a reliable decision, then by definition artificiality is already the only practice for obtaining an evaluable and reliable decision. This is the thesis of strong artificialism from which the weak version, of course, follows. In fact, given epistemic opportunism, the distinc-

tion between weak and strong artificialism effectively evaporates, since the only non-objective methods are those that do not produce an evaluable and reliable decision.

An opponent of artificialism could try to argue against this conclusion by two different means. First, an opponent could insist that artificiality is not de facto epistemically opportunistic. Second, an opponent might claim that artificiality should not be epistemically opportunistic. The first objection will be called the descriptive argument, and the second the normative argument. We will tackle the normative argument first. Since I consider a normative claim, it can only be justified by another normative claim. Here, the claim in question is: artificiality should seek out an evaluably reliable decision. If this is granted, then epistemic opportunism follows by simple instrumental rationality. Now, a proponent of artificialism could reject the normative claim.

Some may also tackle the normative argument against epistemic opportunism and then pose the proposition that artificiality is not actually opportunistic. It follows from the problems with the normative argument that the descriptive one is no argument at all. If one were to accept the descriptive argument but not the normative one, one would state that artificiality is currently conducted in a way that it should not be conducted. In this case, the proponent and the opponent of artificialism agree on how artificiality should be practiced. Artificiality, as it should be, would be epistemically opportunistic and still the only or best way to obtain evaluably reliable decisions.

In summary, a proponent of artificialism can easily avoid the alleged incoherence. Weak artificialism can immediately reject premise (1). Furthermore, a strong version of premise (2) can be denied by weak or strong artificialism. Finally, at least the broad version of artificialism can adopt a view of artificiality that embraces epistemic opportunism and rejects premise (2) by logical inference alone. In conclusion, the argument for self-referential incoherence is faulty.

One might wonder whether adopting epistemic opportunism is going too far in defending

artificialism. After all, suggesting that philosophy and common sense can be part of artificiality might sound non-sensical to the foes of artificiality. One can still oppose: ”Isn't this exactly what artificialism was supposed to oppose?” Not necessarily, artificialism does not have to aim at ruling out intellectual fields based on the notions they can be categorized under. Instead, the task can be to see what demonstratively works and what does not (that is, evaluate which practices produce reliable decisions and which do not). Whatever the labels are filled with.

Here, a concern might arise. That artificialism was simply diluted to avoid the most direct objections to it. But this is not the case. This kind of artificialism already has its supporters, as we demonstrated with quotations in Section 2. Some proponents of artificialism, for instance, may be open to the idea that even philosophy can be among the artificialities.

It is especially important to stress that the variations of artificialism that invoke epistemic opportunism are not all-inclusive. They can have significant implications for the research of automated methods. As mentioned above, if one focuses, for example, on obtaining results in terms of reliable knowledge, adopting epistemic opportunism renders questionable research that does not achieve this objective. Thus, we can have informative and interesting views that take epistemic opportunism on board. And, even if we do not, we can still consider artificialism to be a non-problematic rational hypothesis, for which we can have evidence.

Conclusions

In this paper, I have discussed the three most common reasons for claiming that artificialism is objectionable: the uncharitable definitions of artificialism, the suggested dilemma of artificialism, and the argument of self-refutation.

In Section 3 the dilemma of artificialism was proven to be false. I explained how metaphysical presuppositions are not a necessary part of rationality but can be adopted as mere hypotheses or be discarded altogether as needless. Covering the subject of unscientific sources that form

the basis for all intellectual activity (decisions included), I considered the process of distilling reliable information from initially somewhat unreliable sources.

In Section 4, the accusation of self-referential inconsistency was scrutinized. It was shown that it is possible to gather evidence in favor of and against artificialism. Hence, artificialism can, at the very least, be taken as a rational hypothesis, and it is possible to justify it by objective (rational) means. Here, I argue too that artificiality is based on epistemic opportunism: endorsing whatever methods work for obtaining reliable decisions. If this is correct, then even strong artificialism logically follows.

I also showed that artificialism need not even be a strictly rational hypothesis. In weak artificialism, artificiality is treated not as the *only* source of decision but as simply the *best* one. Even if artificialism could not be accepted as a strictly rational hypothesis (which, of course, is not the case), it could still be validated by using non-objective means. In this case, artificialism alone would not be justified in the best possible manner, but it could be justified nevertheless, in the same sense that our daily judgments can be justified.

Formulating artificialism through epistemic opportunism and evaluable reliability might prompt the worry that we are, in fact, no longer discussing artificialism at all. Such worries are unfounded. Artificialism is motivated by the following observation: among the different ways in which humans try to be rational, the things grouped as "artificiality" are the most successful ones. And suppose that epistemic opportunism and the evaluability of reliability are what make artificiality successful. A proponent of artificialism then claims that these methodological practices do not only make automation successful, but also make it superior compared to other forms of inquiry. Thus, we would be best served in our epistemic projects by employing these means. So artificialism also has a lot to do with the epistemic superiority of actual rationality.

Still, opponents of artificialism might not be satisfied with this answer. Epistemic oppor-

tunism and evaluability are easy to accept, they might admit, but the proponents of artificialism seem to be going further than this. The apostles of artificialism appear to claim that only certain methodologies fulfill the criteria of opportunism and evaluability, usually the methods of the automated systems. ”Hence, artificialism is just a general form of (plural) AI takeover” - someone may say. I have repeatedly argued that this is not the case. Weak versions of artificialism do not force automated system methods on other disciplines. All fields of rationality can maintain their own practices, so long as they work in a provable manner. This also holds for the broad varieties of artificialism. Adherents to broad-weak artificialism claim that the methods of the automated systems are the best methods we have, but that in itself does not yet mean that they have to be adopted in all other fields of inquiry. It might even be impossible to do so. Perhaps automated systems simply cannot offer an alternative to discourse analysis, although, according to the supporters of broad-weak artificialism, discourse analysis can never produce results as reliable as those produced by automated systems.

Nevertheless, it is true that in some cases the sympathizers of artificialism have to say that certain ways of decision-making are inapt because they are inevaluable or there are more reliable methods for generating decisions with roughly the same resources. In such situations, a definition of whether they are instances of reprehensible AI takeover or praiseworthy rational process, in short, became just an empirical question. In every case, we have to examine which methods actually are better for the given goals and ask if they are truly evaluable. So, in the end, it is an empirical matter whether someone is guilty of unacceptable artificialism in the sense of AI takeover.

Perhaps the opposition to artificialism is often motivated by the fact that, in some instances, broad-strong artificialism can be very close to or even amount to AI takeover. Artificialism brings forth important methodological issues that can have important implications for epistemic practices. The three other forms of artificialism can avoid the two global arguments consid-

ered here. Obviously, this does not imply that they are otherwise equally good positions, but assessing them is not the focus of this paper.

I also think that one cannot simply appeal to evaluable reliability or epistemic opportunism if one also wants to uphold an epistemic difference between different artificial fields. Therefore, the problem with the ongoing discussion is also why the more plausible and popular versions of artificialism are equally criticised as less plausible and popular. As I mentioned in Section 2, the definitions of artificialism are, for a large part, those of opponents of artificialism. This may have something to do with the way the main objections arise from the assigned versions of artificialism. Mine overarching thesis is that these versions are by no means necessary and that none of the objections covered in this paper hold up against sophisticated varieties of artificialism.

Further critique of artificialism along the lines I presented needs to challenge the methodological principles that were set forth. This means that one has to consider what the inevaluability of reliability would be and what the role of reliability is with respect to rational claims. This shift to methodological issues will, one hopes, lead the artificialism debate to more fruitful pastures in the future.

Acknowledgments

I would like to thank Johan Hietanen, Petri Turunen, Ilmari Hirvonen, Janne Karisto, Ilkka Pättiniemi and Henrik Saarinen for their paper (*56*) from which I borrowed the research design.

Funding

- No funding has been received.

Competing interests

- The author declares that he has no competing interests.

Data and materials availability

- All data are available in the main text or in the supplementary materials.

References

1. J. J. Cañas, *Frontiers in psychology* **13**, 836650 (2022).

2. B. Goertzel, *Journal of Artificial General Intelligence* **5**, 1 (2014).

3. A. Braga, R. K. Logan, Ai and the singularity: A fallacy or a great opportunity? (2019).

4. T. Hagendorff, *Minds and machines* **30**, 99 (2020).

5. E. Green, Robots and ai: The challenge to interdisciplinary theology, Ph.D. thesis (2018).

6. A. P. Porter, *2014 AAAI Fall Symposium Series* (2014).

7. A. Jackelén, *Zygon®* **56**, 6 (2021).

8. O. Gillath, *et al.*, *Computers in Human Behavior* **115**, 106607 (2021).

9. F. Asbrock, J. Mayerl, M. Holz, H. Andersen, B. Maskow, *Authoritarian Ambivalence Towards Artificial Intelligence" PsyArXiv. April* **15** (2022).

10. K. Porayska-Pomsta, *International Journal of Artificial Intelligence in Education* **34**, 73 (2024).

11. M. Spindler, G. Famira, *COS Journal* **8** (2019).

12. M. Pasquinelli, V. Joler, *AI & society* **36**, 1263 (2021).

13. J. Barrat, *Our final invention: Artificial intelligence and the end of the human era* (Hachette UK, 2023).

14. S. Elhouar, E. Hochscheid, M. Alzarrad, C. Emanuels, *Creative construction e-conference* (2020), pp. 59–66.

15. B. Hmoud, V. Laszlo, *et al.*, *Network Intelligence Studies* **7**, 21 (2019).

16. G. Su, *AI Matters* **3**, 35 (2018).

17. A. Turchin, D. Denkenberger, *Artificial intelligence safety and security* (Chapman and Hall/CRC, 2018), pp. 375–393.

18. T. J. Bench-Capon, *Artificial Intelligence* **281**, 103239 (2020).

19. S. Ramaswamy, H. Joshi, *Springer handbook of automation* pp. 809–833 (2009).

20. A. Rosenberg, *The atheist's guide to reality: Enjoying life without illusions* (WW Norton & Company, 2011).

21. J. Ladyman, D. Ross, *Every thing must go: Metaphysics naturalized* (Oxford University Press, 2007).

22. M. L. Minsky, *Semantic information processing* (The MIT Press, 1969).

23. J.-G. Ganascia, *Knowledge, Technology & Policy* **23**, 57 (2010).

24. F. Russo, E. Schliesser, J. Wagemans, *AI & SOCIETY* pp. 1–19 (2023).

25. T. Araujo, N. Helberger, S. Kruikemeier, C. H. De Vreese, *AI & society* **35**, 611 (2020).

26. S. J. Russell, *Artificial intelligence* **94**, 57 (1997).

27. S. J. Gershman, E. J. Horvitz, J. B. Tenenbaum, *Science* **349**, 273 (2015).

28. J. Li, J.-S. Huang, *Technology in Society* **63**, 101410 (2020).

29. E. S. Zhan, M. D. Molina, M. Rheu, W. Peng, *International Journal of Human–Computer Interaction* pp. 1–18 (2023).

30. D. N. Banerjee, S. S. Chanda, *arXiv preprint arXiv:2008.04073* (2020).

31. R. Williams, R. Yampolskiy, *Philosophies* **6**, 53 (2021).

32. R. V. Yampolskiy, *foresight* **21**, 138 (2019).

33. S. S. Chanda, D. N. Banerjee, *AI & society* **39**, 937 (2024).

34. D. Saneei, R. Abplhassani, *AI Training Manual* (2007).

35. M. Dehghani, A. Severyn, S. Rothe, J. Kamps, *arXiv preprint arXiv:1711.00313* (2017).

36. A. Fawzi, H. Samulowitz, D. Turaga, P. Frossard, *2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)* (Ieee, 2016), pp. 1–5.

37. D. Chialvo, P. Bak, *Santa Fe Institute, Working Papers* (1997).

38. B. Garg, *et al.*, *Proceedings of the AAAI Conference on Artificial Intelligence* (2022), vol. 36, pp. 10184–10192.

39. A. Darmody, D. Zwick, *Big Data & Society* **7**, 2053951720904112 (2020).

40. J. Brusseau, *Theoria* **67**, 1 (2020).

41. M. Andrejevic (2020).

42. E. Hubinger, *et al.*, *arXiv preprint arXiv:2401.05566* (2024).

43. J. S. Lerner, Y. Li, P. Valdesolo, K. S. Kassam, *Annual review of psychology* **66**, 799 (2015).

44. J. M. George, E. Dane, *Organizational Behavior and Human Decision Processes* **136**, 47 (2016).

45. N. Naqvi, B. Shiv, A. Bechara, *Current directions in psychological science* **15**, 260 (2006).

46. N. Schwarz, *Cognition & emotion* **14**, 433 (2000).

47. S. R. Quartz, *Trends in cognitive sciences* **13**, 209 (2009).

48. M. Toda, *Acta Psychologica* **45**, 133 (1980).

49. K. S. Kassam, *Annu. Rev. Psychol* **66**, 33 (2015).

50. E. T. Rolls, *Emotion and decision-making explained* (OUP Oxford, 2013).

51. H. Brown, *The Journal of Adult Protection* **13**, 194 (2011).

52. A. Gaudine, L. Thorne, *Journal of Business Ethics* **31**, 175 (2001).

53. T. Gilovich, D. Griffin, D. Kahneman, *Heuristics and biases: The psychology of intuitive judgment* (Cambridge university press, 2002).

54. G. E. Gigerenzer, R. E. Hertwig, T. E. Pachur, *Heuristics: The foundations of adaptive behavior.* (Oxford University Press, 2011).

55. G. Gigerenzer, P. M. Todd, *Simple heuristics that make us smart* (Oxford University Press, USA, 1999).

56. J. Hietanen, *et al.*, *Metaphilosophy* **51**, 522 (2020).