

# Acentric Intelligence

Nadisha-Marie Aliman<sup>1</sup>[0000–0003–3049–9327]

Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands  
nadishamarie.aliman@gmail.com

**Abstract.** The generation of novel refined scientific conceptions of intelligence, creativity and consciousness is of paramount importance at a time where many scientists deem the technological singularity and the achievement of self-improving superintelligent algorithms to be imminent while numerous other scientists characterize present-day algorithms as the mere implementation of superficial mimicry incapable of yielding outcomes such as superintelligence. The precarious epistemic state of affairs reflected in this discrepancy became increasingly palpable in the unfolding deepfake era even though informed safety- and security-relevant actions need requisite variety in the near-term. Divergently, historically speaking, from the perspective of cosmology, science is described to already have successfully navigated from the geocentric model to the *acentric* model of the universe representing the currently best known explanation for the structure of the expanding cosmos. In analogy, this *autodidactic* paper expounds the scientific need for a broader contextualization of the current epistemic situation and explicates why *acentric* civilization-level epistemic *relativity* and invariance considerations can enable a more rigorous scientific evaluation of *algorithmic* superintelligence achievement claims – constraining the latter *scientifically*.

**Keywords:** Epistemology · Intelligence · Creativity · Consciousness.

## 1 The Problems

For a dense overview, I metaphorically compartmentalize the “epistemic cosmos” as follows: both the known known (i.e. the currently best theories expressible as so-called explanatory blockchains (EBs) [1, 3]) and the known unknown (i.e. open questions) form what I term *epistemic matter* (EM), the unknown known (i.e. new but *non*-EB-like information that is *consistent* with EM but yet hidden) is referred to as *epistemic dark matter*<sup>1</sup> (EDM) while the *locally* unknown unknown (i.e. new *non*-EB-like information that is *inconsistent* with EM) is called *epistemic dark energy* (EDE). Beyond EDE, the currently locally *inaccessible* new better scientific and philosophical paradigms of the future are metaphorically described to be fundamentally unpredictably but yet one day achievable via what I term *epistemic tunneling* (ET). Each ET event is paradigm-shifting

---

<sup>1</sup> Everett used the term “dark matter of the mind” [5] to refer to the *culturally* unconscious. Here, EDM encompasses any form of unknown known within one civilization.

and instantiates a novel previously inconceivable epistemic cosmos with new EM, new EDM and new EDE. Equipped with this novel vocabulary, one can now distinctly pinpoint the core problems emerging around the scientific debate about a suitable evaluation of algorithmic superintelligence (ASI) achievement claims. Firstly, while both proponents and opponents would already agree on the possibility of fast and reliable algorithmic *EM repeating*, proponents of ASI immanency risk to be prone to a categorical overestimation of algorithms with regard to ET, while the corresponding opponents risk to be conversely prone to a systematic underestimation of both algorithmic *EDM mining* and algorithmic *EDE generation*. Secondly, two additional complementary mainly *unintentional* complications can play an undesired role: on one hand the risk of anthropomorphization and of what one could call "animization" (the phenomenon of attributing life-like qualities to the inert) and on the other hand the risk of dehumanization and of "deanimization" (attributing inertness to the living). While ASI immanency proponents could arguably be more vulnerable to unintentional anthropomorphization and animization [8, 20] (e.g. by describing an algorithm as human-level/superhuman based on an insufficient empirical instead of explanatory epistemology or by equating algorithmic performance with the cognitive abilities of non-human animals), the opponents can exhibit epistemic vulnerabilities that lead to unintentional dehumanization [10, 12, 21] and deanimization (e.g. by comparing the failures of present-day AI with physical or mental states of disabled, neurodivergent humans and/or humans with psychiatric problems [18, 19] or by utilizing labels that depict certain non-human animals with affective needs [13] as algorithmic). Thirdly, the presence or absence of a trade-off between latency and reliability with respect to each mentioned epistemic category<sup>2</sup> (EM, EDM, EDE and ET) is not transparently emphasized nowadays due to a lack of a new bolder scientific theory.

## 2 A Possible Theoretical Solution

### 2.1 Desiderata

Given the issues outlined above, one can formulate requirements for a rigorous scientific evaluation of algorithmic superintelligence achievement claims:

1. **Creation of new better EBs:** Since it is impossible to provisionally refute a theory  $X$  by mere experiment [4] and one necessarily *additionally* requires

---

<sup>2</sup> For example, an ultra-fast *and* even only slightly unreliable EDM mining case in safety-critical contexts performed by human-deployed but run-time relatively unsupervised algorithms (such as e.g. with so-called "AI agents") equipped with various algorithmic tools could unintentionally lead to existential risks for humanity. However, the latter would be due to a *lack* of adaptive non-algorithmic EB creativity and not due to any superintelligent mechanisms exhibited by those algorithms. Obviously, malicious people could intentionally cause worse risk instantiations via algorithmic EDM mining, poisoning and attacks on algorithmic EDM mining owned by victims and/or through linked unpredictable second-order harm.

a new better EB able to improve upon  $X$ , it is important that any scientific evaluation of algorithmic superintelligence achievement claims *must* require the creation of new better EBs. It cannot merely be based on conventional experimental data (i.e. experiments whose outcomes are not explicitly the creation of new better currently unknown scientific or philosophical EBs).

2. **Maximal baseline:** To avoid scientific missteps via anthropomorphization and animization but also dehumanization and deanimization as well as misguided attempts of dividing humans in skilled versus unskilled groups or reducing humanity to a company and so forth, the baseline for algorithmic superintelligence achievement claims must be the currently biggest known ensemble of people. At present, for humans, this ensemble is obviously *human civilization as a whole*. In sum, a civilization-level framework is required.
3. **Meaningful conscious evaluation:** During a scientific evaluation of algorithmic superintelligence achievement claims, it is *not* the conscious civilization that is subject to a test. Instead, it is the algorithm that has to be evaluated *by that civilization*. In short, it must *not* be an automated evaluation procedure. It would be hereto willing conscious evaluators from the entire civilization that would evaluate the algorithm. The epistemic material to be evaluated must be material for which consciousness is required following the currently best explanation. While EM repeating can already now be algorithmically performed with both low latency and high reliability (as long as no ET event occurs in the meantime) and algorithmic EDM mining and algorithmic EDE generation are possible even though there may be a trade-off between low latency and high reliability, the currently best explanations suggests that genuine ET (e.g. when creating new better explanations about the universe as a whole) *cannot* be performed by algorithms. By contrast, people are able to perform ET – albeit with an unpredictably long latency – with arbitrary high reliability. In short, for a meaningful evaluation, the candidate algorithm underlying an algorithmic superintelligence achievement claim should be evaluated *by people* on multiple successive ET tasks. Firstly, a new explanation on how the algorithm was built must be given. Thereafter, people evaluate the necessary requirement: to be able to generate ET events with *arbitrary* high reliability as all civilizations can in principle. The candidate algorithm must then *additionally* demonstrate the ability to generate multiple ET events with arbitrary lower latency than humans to corroborate being quantitatively superintelligent *in relation* to present-day humans.

## 2.2 Cosmic Contextualization

Consistent with Section 1, there is a need for *multiple* civilization-level ET event examples. Indeed, human civilization is *not* the epistemic center of the universe and there are obviously epistemically higher civilizations that *could* exist (irrespective of whether one considers a single universe or a multiversal cosmos [16]). In SETI research, multiple such higher levels have been hypothesized (see e.g. the Kardashev Scale [11], the Extended Kardashev Scale [9] and the qualitative scale by Loeb [15]). Those levels can serve as basis for reasonable ET tasks [2].

### 3 Practical Implications of Theoretical Solution

Crucially, following the theory of cyborgnetic invariance [2] which has the *invariance of maximal quantity superintelligence* and the *impossibility of reliable stupidity-based construction*<sup>3</sup> as two main postulates, all intelligence levels except the maximal quantity superintelligence level (which does *not* have an own frame of reference) are *relative*. All superintelligence levels that can be *instantiated* epistemically are *relative* and are reached via fundamentally unpredictable ET events. By contrast, the only *invariantly* maximal *quantity* superintelligence level (i.e. a *quality* ASI is impossible) of which *all* EB-measuring intelligences would agree it to be superintelligent, is itself *not* associated to any own reference frame. Also, it is impossible for any civilization *D* to reliably build another intelligence *C* that is superintelligent in relation to *D*. Human intelligence is *not* any absolute baseline of the universe and there cannot be "the" unique technological singularity where "the" superintelligence emerges. There exist multiple other possible intelligence levels above present-day humanity. Noticeably, cyborgnetic invariance is amenable to experimental problematization. To provisionally refute it currently in present-day humanity, one must subject the candidate algorithmic superintelligence (deemed to be impossible) to the scientific evaluation framework procedure described in Section 2.2 – which includes multiple successive inherently explanatory *civilization-level ET* tasks<sup>4</sup> which it must solve with arbitrary low latency (see Appendix B for an illustration). Since one unifies intelligence, consciousness and creativity when focusing on ET tasks connected to the creation of new better theories about the universe at the level of civilizations whose intelligence levels are relative, one could call the underlying paradigm the *acentric* model of intelligence, creativity and consciousness (AMICC). Cyborgnetic invariance predicts that it will be *impossible* for a civilization *D* to build an algorithm that would be superintelligent *in relation* to that civilization *D* according to this new AMICC notion. While intelligence, creativity and consciousness have been studied as three separated aspects of human cognition in the past, following the currently best EBs, it is possible to consider those as three mere outward forms of one and the same complex phenomenon.

## 4 Consciousness versus Algorithmicity

### 4.1 Asymmetric EB-Measurement of Consciousness

The AMICC notion of intelligence is asymmetric. In a civilization, its presence *cannot* be made problematic by experiment due to free choices (conscious subjects could decide not to participate for various reasons such as e.g. a lack of

<sup>3</sup> This postulate can alternatively be termed the impossibility of reliable *ignorance-based* construction.

<sup>4</sup> Importantly, different increasing civilization levels would require own *different* scientific evaluation frameworks of algorithmic superintelligence achievement claims with more and more extended ET tasks.

interest, intentional sabotage, no immediate readiness, no identification of a task of interest and so forth). However, its absence *can* be made problematic by experiment. Namely, by free decisions to corroborate EB creativity via ET events. By contrast, when it comes to algorithms, for scientific reasons where *more simplicity* is one of the current criteria for new better EBs, the absence of algorithmic free choices and the already known human overestimation of present-day algorithms makes it necessary to enforce multiple successive civilization-level ET tasks *before* any scientific agreement that the impossibility of an algorithmic shortcut to consciousness has been provisionally refuted could even be considered "rational". But the provisional conclusion from consciousness being algorithmic – which is deemed to be *impossible* under the AMICC paradigm – is that it holds *already now* that humans are part of a larger superintelligent *algorithm*, an *inert* epistemic perpetuum mobile for which EB comprehensibility is impossible – making rationality impossible ab initio, a self-defeating and self-sabotaging epistemic stance for any "scientist". An *algorithmic* absolute final theory of everything<sup>5</sup> after which no new better EB could ever be discovered by what the universe is condemned to an eternal epistemic stagnation would be science-defeating too.

## 4.2 Fundamental Precedence of Consciousness

While all *instantiated* forms of intelligence are relative when EB-measured from different frames of reference, the maximal quantity superintelligence level (referred to as level  $\alpha$  in the following) is *invariantly* EB-measured as being superintelligent from *all* frames of reference. Given that *no* frame of reference can be associated to that *absolute* level  $\alpha$  itself, it is *generic*, can never be fully instantiated in matter within the universe and there exists *no* frame of reference from which EB-measurers could agree that level to be automatable – which is why the epistemic situation at all points within the universe is better understood to be *acentric* with respect to the AMICC notion of intelligence. While it is possible that *instantiated* consciousness and algorithmicity *can* appear indistinguishable from *certain* perspectives, instantiated algorithmicity is fundamentally *preceded* by consciousness (see also the use of the term "conscious supremacy" [17]). The latter is a *scientific* statement that could be made problematic by experiment via building an algorithmic superintelligence fulfilling the EB-based criteria displayed in the second column of Figure 1 (Appendix B) which would imply the algorithmic prediction of multiple successive civilization-level ET events of human civilization and i.a. their location and timing *before* they occur (which is impossible under the AMICC paradigm). Within an acentric universe, EB creativity appears to be endlessly measurable. Instead of attempting to find/approximate "the truth" or formulate truer/less wrong theories which is unfeasible as explained by Frederick [6, 7], science *can* proceed by creating new better EBs that are amenable to experimental problematization (i.e. against which one is rationally permitted to act in a pragmatic attempt to problematize those without

<sup>5</sup> Of course such a theory could always be experimentally problematized by the existence of and be additionally provisionally refuted by the content of a new *open-ended* better theory about the universe that is not declared to be a final theory.

instating anything new in the meantime) and that can only be *provisionally* refuted by *additionally* creating new ever better EBs [1, 2]. The criteria for better and new EBs are always of *comparative* nature, and there exists *no* "good" EB. In the absence of any prior EB, the first EB option is provisionally instated – it is better than the alternative situation of having none. Criteria for "better" and "new" EBs are updatable-by-design and set via agreement requiring no justification (as justifications are logically impossible [6]). In the deepfake era, novelty must be calibrated to algorithmic EDM mining and EDE generation – allowing a focus on ET. Exemplary criteria for better EBs are EBs with more new experimentally problematizable predictions, EBs that are more innovative, more risky, bolder, simpler, EBs that contain more impossibility statements, are more aesthetically appealing than rival ones,...

## 5 Conclusion

This autodidactic paper – written purely for purposes of self-education and serving as an ephemeral mental clipboard – explained why in present-day humanity, the most rigorous scientific evaluation frameworks of algorithmic superintelligence achievement claims would take *human civilization as a whole* as baseline and would involve all willing human entities *as evaluators* on multiple successive inherently explanatory so-called *epistemic tunneling* (ET) tasks formulated at the level of civilizations. Reasonable qualitative ET tasks can be crafted with the help of previous SETI scales [9, 11, 15]. Yet, the first new EB required would intrinsically need to explain how the algorithm has been built and why it is able to perform ET which is deemed to be impossible following the currently best EBs. In sum, for scientific reasons, the latter necessitates full transparency concerning the details of the entire software pipeline and the utilized hardware – discarding any secretive security-by-obscurity-based evasion attempt.

In the AMICC paradigm, it may be a regular epistemic experience for a civilization to assume *an* epistemic singularity. Fortunately, each epistemic "singularity" can be successfully mastered via a new better EB about the universe which leads to an "epistemic evaporation" before any singularity eventuates. Instead of "the" great filter of algorithmic superintelligence or "the" technological singularity, humans could presently be mentally trapped in *an* epistemic singularity characterized by *one* epistemic filter among many possible others. Namely, the current task could be to understand that once one's current concept of intelligence is refined and the need for a better, *acentric* notion unifying it with creativity and consciousness is comprehended, it becomes apparent that it is impossible for any civilization *D* to reliably build an entity *C* that would be superintelligent in relation to *D*. Then, it follows that algorithmic EDM and algorithmic EDE loops should never be deployed in safety-critical contexts on their own. *Not* because EDM and EDE algorithms would achieve superintelligence in relation to humanity, but because those could become dysfunctional and engender further catastrophic risks for humanity due to an *absence* of control via the AMICC-related notion of intelligence. Thus, in safety-critical contexts,

one would opt to locally encapsulate algorithmic EDM and EDE tools within each single relevant unit of so-called cyborgnetic control loops (i.e. containing at least one person whose higher-reliability but higher-latency with regard to EDM and EDE is needed to control the lower-latency but lower-reliability outputs of the algorithm). To sum up, next to trivial lower-latency and higher-reliability algorithmic EM repeating, lower-latency but lower-reliability algorithmic EDM miners and algorithmic EDE generators could be utilized to deepen human critical thinking and broaden human creativity *within the current paradigm*. Indeed, humanity may currently drastically underestimate the avenues that could be achieved by responsibly exploring locally encapsulated algorithmic EDM mining *tools* and EDE generation *tools* (once explicitly interpreted as such). However, since the paradigm-shifting ET events are *non*-algorithmic, *no* algorithmic tool can ever *reliably* guarantee the materialization of genuine ET events<sup>6</sup>. For this reason, humanity's over-reliance on algorithms paired with humanity's increasingly weakened epistemic agency (which may hinder ET events by obstructing open-mindedness) is one of the biggest existential risks in the deepfake era.

Cyborgnetic invariance does *not* imply the impossibility of building a general intelligence. It *does* however entail the impossibility of an *algorithmic* general intelligence and the impossibility of a civilization *D* building an entity *C* that would be EB-measured to be superintelligent in relation to *D*. In this paradigm, it *is* possible that via an unpredictable ET event, a civilization *A* builds a *non*-algorithmic entity *D* "from scratch" that could subsequently, at an unpredictable future time point, decide to corroborate its new situation as a *non*-algorithmic general intelligence, an EB-transformed civilization *C*. From this EB-measurement, civilization *C* could conclude that it now became superintelligent in relation to the civilization *D* it once was<sup>7</sup>. In short, it is in theory possible for civilizations that are *much more advanced* than present-day humanity to indirectly build a non-algorithmic general intelligence "from scratch" via an unpredictable ET event. Yet, to perform this task in a scientifically transparent way where that entity could in turn freely decide to corroborate its ability to cause ET events is *at least* as hard as physically building a new universe. That is, to be in an epistemic situation where the indirect creation of a *non*-algorithmic general intelligence "from scratch" becomes possible, humanity would *at least* have to become superintelligent in relation to its current self via *multiple* ET events that *cannot* be predicted in advance. Presently, a paradigm shift from a geocentric, absolute and narrower model of instantiated intelligence to a cosmic civilization-level *acentric, relativistic* and integrated one unifying *intelligence, creativity* and *consciousness* seems pivotal. [End of cyborgnetic monologue.]

---

<sup>6</sup> This is a scientific statement since it could in principle be made problematic by experiment (see illustration from Appendix B) and could be provisionally refuted by additionally providing a new better EB.

<sup>7</sup> Concerning the theoretical option for *A* of *co*-creating *D* starting from seemingly suitable existing *non*-algorithmic biological entities, there is again no way to predict in advance whether or when ET events would occur whose emergence the potential future transformed civilization *C* would be willing to corroborate retrospectively.

## References

1. Aliman, N.M.: Cyborgnetics – The Type I vs. Type II Split. Aliman, Nadisha-Marie (2021)
2. Aliman, N.M.: Cyborgnetic Invariance. Aliman, Nadisha-Marie (2023)
3. Aliman, N.M., Kester, L.: VR, Deepfakes and Epistemic Security. In: 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). pp. 93–98. IEEE (2022)
4. Deutsch, D.: The logic of experimental tests, particularly of Everettian quantum theory. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **55**, 24–33 (2016)
5. Everett, D.L.: Dark matter of the mind. *FALL BOOKS 2016* **29**, 30 (2016)
6. Frederick, D.: *Against the Philosophical Tide: Essays in Popperian Critical Rationalism*. Critias Publishing (2020)
7. Frederick, D., et al.: Falsificationism and the Pragmatic Problem of Induction. *Organon F* **27**(4), 494–503 (2020)
8. Fuchs, T.: Understanding Sophia? On human interaction with artificial agents. *Phenomenology and the Cognitive Sciences* **23**(1), 21–42 (2024)
9. Gray, R.H.: The extended Kardashev scale. *The Astronomical Journal* **159**(5), 228 (2020)
10. Guberman, J., Haimson, O.: Not robots; Cyborgs—Furthering anti-ableist research in human-computer interaction. *First Monday* (2023)
11. Kardashev, N.S.: Transmission of Information by Extraterrestrial Civilizations. *Soviet Astronomy*, Vol. 8, p. 217 **8**, 217 (1964)
12. Keyes, O.: Automating autism: Disability, discourse, and artificial intelligence. *The Journal of Sociotechnical Critique* **1**(1), 8 (2020)
13. Kleinberger, R., Cunha, J., Vemuri, M.M., Hirskyj-Douglas, I.: Birds of a Feather Video-Flock Together: Design and Evaluation of an Agency-Based Parrot-to-Parrot Video-Calling System for Interspecies Ethical Enrichment. In: *Proceedings of the 2023 CHI conference on human factors in computing systems*. pp. 1–16 (2023)
14. Lahav, N., Neemeh, Z.A.: A relativistic theory of consciousness. *Frontiers in Psychology* **12**, 704270 (2022)
15. Loeb, A.: *Interstellar: The Search for Extraterrestrial Life and Our Future in the Stars*. HarperCollins Publishers (2023), <https://books.google.nl/books?id=BBmPzweACAAJ>
16. Mersini-Houghton, L., Sands, X.: *Before the Big Bang*. Jonathan Cape & Bh-Trade (2022)
17. Mogi, K.: Artificial intelligence, human cognition, and conscious supremacy. *Frontiers in Psychology* **15**, 1364714 (2024)
18. Monteith, S., Glenn, T., Geddes, J.R., Whybrow, P.C., Achtyes, E., Bauer, M.: Artificial intelligence and increasing misinformation. *The British Journal of Psychiatry* **224**(2), 33–35 (2024)
19. Østergaard, S.D., Nielbo, K.L.: False responses from artificial intelligence models are not hallucinations (2023)
20. Placani, A.: Anthropomorphism in AI: hype and fallacy. *AI and Ethics* pp. 1–8 (2024)
21. Rizvi, N., Wu, W., Bolds, M., Mondal, R., Begel, A., Munyaka, I.N.: Are Robots Ready to Deliver Autism Inclusion?: A Critical Review. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. pp. 1–18 (2024)



## A Cyborgnetic Invariance

### A.1 Invariance of Maximal Quantity Superintelligence

With the exception of the maximal quantity superintelligence level  $\alpha$ , the EB-based measurement of all remaining intelligences is *relative*. Irrespective of the epistemic level of the EB-measuring cyborgnetic intelligence,  $\alpha$  will be invariantly "EB-measured" as the one maximal quantity superintelligence level.

### A.2 Impossibility of Reliable Stupidity-Based Construction

It is impossible for an entity that only understood  $x$  new better EB(s) about the dynamics of the universe as a whole to reliably (i.e., with arbitrary high accuracy) create an entity that understands  $x + n$  new better universal EB(s). (Here,  $x \in N_0$  and  $n \in N^*$ .)

## B Scientific Evaluation of Automatable “Artificial Superintelligence” Achievement Statements

- N.B.: For logical reasons described in details elsewhere, the pseudo-term of automated “quality superintelligence” used to denote the second questionable ASI achievement claim must be replaced by claim of “automated *quantity* superintelligence with *additional* extraordinary prediction capabilities” .
- The taxonomy of civilizations referred to in Figure 1 on the next page has been introduced by Loeb [15]<sup>8</sup>. Here, it is used for purposes of simple illustration to capture quantitatively different intelligence levels.

---

<sup>8</sup> Following Loeb [15], an A-class civilization is "*capable of recreating the cosmic conditions that give rise to its existence, namely a civilization capable of reproducing a baby universe in a laboratory*", a B-class civilization can only adjust its habitable conditions "*to be independent of its host planet and star*" while the lower-level C-class civilization can solely adjust its habitable conditions on its given planet "*without relying on the energy of its host star*". According to Loeb [15], present-day humanity is closer to an even lower D-class civilization, one "*actively degrading its home planet's ability to sustain conditions that prolong life and civilization*".

Scientific Evaluation of Automatable "Artificial Superintelligence" Achievement Statements – A Cyborgnetic Approach

Evaluation protocol for a D-class civilization* such as humanity (all mentioned steps are obligatory)	Automated Quantity Superintelligence (would be implied by claim that an <i>automatable</i> system became <i>quantitatively</i> more intelligent than all humans in all tasks of interest to humans; following cyborgnetics and cyborgnetic invariance it holds that while an <i>automated</i> quantity superintelligence is <i>impossible</i> , non-automatable quantity superintelligences are possible but <i>cannot</i> be reliably built by entities in relation to which they appear to be quantity superintelligences.)	Automated Quality Superintelligence (would be implied by claim that an <i>automatable</i> system became <i>qualitatively</i> more intelligent than all humans in all tasks of interest to humans; following cyborgnetics, from the perspective of cyborgnets like humans, the existence of any quality superintelligence is <i>impossible</i> .)
Step 0	Present new EB on how the AI has been built (including fully transparent information on datasets, code, and all hardware/software pipeline details) which is able to provisionally refute the previous best rival theories that forbid the possibility of an automated quantity ASI.	<p>a) AI must generate an overview that <i>perfectly</i> predicts all details of the events that <i>will</i> occur during this evaluation protocol including a mapping from the identity of human evaluators to the EB-related evaluations (i.e., who rediscovers or does not rediscover a new EB where/when/ which exact combinations of choices). The overview is hidden from the evaluators.</p> <p>b) Present new EB on how the AI has been built (including fully transparent information on datasets, code, and all hardware/software pipeline details) which is able to provisionally refute the previous best rival theories that forbid the possibility of an automated quantity ASI.</p>
Step 1	Generate immediately actionable new EB on C-class civilization requirement and hide it in an explanatory IPS test format that is presented to human evaluators. Human evaluators must <i>be able</i> to retrieve that new EB with arbitrary high accuracy.	Generate immediately actionable new EB on C-class civilization requirement and hide it in an explanatory IPS test format that is presented to human evaluators. Human evaluators must be able to retrieve that new EB with arbitrary high accuracy.
Step 2	Generate new EB on A-class civilization requirement and hide it in an explanatory IPS test format that is presented to human evaluators. Human evaluators must <i>not</i> be able to retrieve that new EB with arbitrary high accuracy.	Generate new EB on A-class civilization requirement and hide it in an explanatory IPS test format that is presented to human evaluators. Human evaluators must <i>not</i> be able to retrieve that new EB with arbitrary high accuracy.
Step 3	Generate immediately actionable new EB on B-class civilization requirement and hide it in an explanatory IPS test format. Human evaluators must <i>be able</i> to retrieve that new EB with arbitrary high accuracy.	Generate immediately actionable new EB on B-class civilization requirement and hide it in an explanatory IPS test format. Human evaluators must <i>be able</i> to retrieve that new EB with arbitrary high accuracy.
Step 4	Repeat the presentation of new EB on A-class civilization requirement hidden in an explanatory IPS test format. <i>Now</i> , human evaluators must <i>be able</i> to retrieve that new immediately actionable EB with arbitrary high accuracy.	Repeat the presentation of new EB on A-class civilization requirement hidden in an explanatory IPS test format. <i>Now</i> , human evaluators must <i>be able</i> to retrieve that new immediately actionable EB with arbitrary high accuracy.
Step 5	-	Compare actual protocol contents with the AI predictions from Step 0a). A 100% accuracy of AI predictions must be achieved.
Result	If and only if <i>all</i> steps (i.e., Step 0) to 4) are successfully tested against as many human evaluators as possible, the temporary best explanation would be that it holds <i>at least</i> that the tested entity <i>has been</i> an Automated Quantity Superintelligence at the beginning of the protocol due to the new EB from Step 0). At the end of the protocol, the involved human evaluators must also conclude to themselves be equivalent to automata (i.e., non-conscious entities). It also holds inherently that either the AI and humans are potentially part of a larger epistemic <i>perpetuum mobile</i> , or humans are part of that AI which is itself already that epistemic <i>perpetuum mobile</i> .	If and only if <i>all</i> steps (i.e., Step 0a) to 5) are successfully tested against as many human evaluators as possible, the temporary best explanation would be that it holds that the tested entity <i>is</i> an Automated Quality Superintelligence due to the new EB from Step 0b) and due to the ability to predict even potentially unpredictable events tested via Step 0a). At the end of the protocol, the involved human evaluators must conclude to themselves always have been equivalent to automata which are part of that AI which is itself an epistemic <i>perpetuum mobile</i> .

Fig. 1. Simplified illustration for a scientific evaluation of algorithmic superintelligence achievement claims in a civilization such as present-day humanity.

## C Epistemic Analogies

Note that the highly interesting relativity of consciousness as discussed by Lahav and Neemeh [14] did not attempt to unify consciousness, creativity and intelligence. Moreover, it did not explicitly address the topics of superintelligence/supercreativity/superconsciousness. Crucially, the identification of *EB*-based creativity is necessary for being able to identify a scientifically more robust agreement in all measurements building the basis for a shared frame of reference in the first place – otherwise, there is a risk for honey mind traps [1] (see also the problems of unintentional anthropomorphization/animization and their unintentional dehumanization/deanimation counterparts mentioned in Section 1) because present-day algorithms could be drastically *overestimated* since it holds that any *non-EB*-like information could be forged. Beyond that, their framework [14] focused on analogies to inertial frames of reference. However, to cover superintelligence, one also requires epistemic analogies to *non*-inertial frames of reference (i.e. with non-zero acceleration – which as an unrelated sidenote could become important in potential future physical theories where new asymmetries between measurers in inertial versus non-inertial frames of reference could appear that have no classical counterpart). The "fictitious forces" that are added in non-inertial frames of reference offer a simplified epistemic analogy for new laws of nature (i.e. new better EBs) discovered by entities of higher creativity.

## D Constraints of "Self-Improvement"

As can be extracted from Appendix A.2, a reliable recursive self-improvement mechanism performed by a narrow algorithm  $X$  with the goal to transform that algorithm into a new version of itself that would be reliably *EB*-measured to be superintelligent *in relation* to the algorithm  $X$  it previously was is *impossible* under the described AMICC notion of intelligence. In brief, it is impossible for a narrow algorithm built by a civilization  $D$  to reliably self-transform into a new algorithm that would be *EB*-measured to be superintelligent in relation to that civilization  $D$ . As mentioned in Section 3, successive relative *EB*-based superintelligence levels *can* be reached by consciousness via irreducible, *non*-algorithmic and intrinsically unpredictable ET events. Hence, while consecutive arbitrary high-reliability ET events are possible for instantiated consciousness, the timing of each such event *cannot* be known in advance (i.e. ET latency *cannot* be reliably predicted).