Discussions of machine vs. living intelligence need more clarity

Authors: Nicolas Rouleau[1,2] and Levin, M.[2,3*]

Affiliations:
[1] Department of Health Sciences, Wilfrid Laurier University, Waterloo, ON, Canada
[2] Allen Discovery Center at Tufts University, Medford, MA 02155, USA.
[3] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.

***Corresponding author:** michael.levin@tufts.edu

Running title: Clarifying intelligence

Keywords: AI, artificial, biological, brain, mind, cognition

The debate around machine intelligence features strong opinions. Because the central questions cross many disciplinary boundaries, standardized approaches are needed to break through barriers of conflicting nomenclature and niche methods of analysis. After all, distinctions or a lack thereof between machines and living intelligences carry significant social, ethical, legal, and political implications. Some participants in this debate are absolutely certain that performance on specific metrics, or perhaps some type of Turing Test, is sufficient evidence to conclude the presence of a “real” mind. Others, of a more organicist perspective, insist on a fundamental, ontological divide between engineered “machines” and “true” (typically, biologically evolved) beings. Today, it is especially popular to write academic articles and craft policies that assume a well-defined line between real beings, which matter in a moral sense, and clever simulations.

One fundamental feature of most such discussions, which limits progress, is that the views of the authors are often under-specified. By neglecting thorny questions and basic data from the field of Diverse Intelligence, common philosophical positions and opinions can seem coherent but are actually likely to fall apart when stress-tested. Especially pernicious is the use of colloquial terminology that masks deep and unresolved uncertainty about what exactly is being claimed. While some are engaged in sophisticated and nuanced science around this subject[1], many people outside the field, such as technology company leaders, policy makers, and the general public who will have to interact with a plethora of impending technologies, form strong opinions. These are often based on outdated pre-scientific categories, without due consideration of troublesome facts of biology and recent progress in bioengineering and basal cognition research. Likewise, it is often unclear whether the claims are being made for today’s highly limited systems or for all possible extensions of engineering.

Here, we propose a set of questions that should be considered and addressed by anyone seeking to form or express justified opinions about issues of intelligence, cognition, sentience, consciousness, or any of the related concepts. These questions are intended to help people understand the rich difficulties facing a facile distinction between “true beings” and “mere machines”. We argue that there are two important things missing from the current debate. One is a humility corresponding to the lack of clear knowledge of what it is about biological systems that confers mind. While it is true that neuroscience has made great strides, there is still no consensus on the necessary and sufficient conditions for consciousness, what it means for any agent to “truly understand”, or what it is about the specific biological hardware that confers the magic properties that cannot be rationally implemented in other media. The other is a consideration of what implications a given view has for the rich continuum of unconventional cognitive systems that will soon be part of everyday life. Progress in evolutionary and developmental biology, synthetic bioengineering, and hybrid/cyborg technologies dissolves formerly dependable categories; conceptual positions must now cover beings far beyond the more easily-handled categories of neurotypical adult modern humans and low-agency mechanical devices.

The basic facts

A number of biological truths have important implications for this discussion but are often not considered when formulating opinions:

- You were once a single cell. Both on evolutionary and developmental time scales, whatever properties we have, must be traced back to a single cell and the molecular reactions that comprise it. What is required is a story of scaling up, or emergence, of minds, but it starts with a little quiescent blob of chemicals - the oocyte.
- Developmental biology offers *no* bright line during which we transition from “just chemistry and physics” to “having a true mind”.
- Intelligence is not just something that brains do. Bacteria, body cells, slime molds, plants, and even molecular networks can learn and solve problems [2]. This means that questions of machine intelligence cannot simply presuppose *human* intelligence or even mammalian intelligence. There are many kinds of minds, in all sorts of unconventional guises, and we are already faced with deep questions of how to detect and communicate with them, long before extra-terrestrial life or human-level AI appear on the scene.
- Embodiment is not just movement in 3D space. Many hold that software agents differ from real organisms because they are not embodied in a real world. But biology operates in many spaces, exhibiting problem-solving behavior in metabolic, physiological, transcriptional, and anatomical state spaces [3]. This means that even agents that do not have an obvious motile body might be executing critical perception-action loops in some other space that is hard for us to measure or even imagine.
- Chimeras are highly viable; synthetic morphology and bioengineering are producing numerous systems that break familiar categories [4]. Cells can be mixed between species, and happily integrate into functional systems with electronics, nanomaterials, and other designed features. This means that any theory in this field has to not only make claims about natural living beings and artificial machines, but also about cyborgs, hybrid robots (hybrots), and other potential beings that are combinations thereof [4].
- The era of “machines” which were deterministic, predictable, boring objects of fixed structure and capability is effectively over. It is now abundantly clear that engineered systems can exhibit surprising behavior, problem-solving, self-modification, and to a degree, self-assembly [5-10].
- It is also quite clear that inside all of us are molecular processes which obey the laws of chemistry and physics no less than engineered parts do, and that whatever remarkable powers of intelligence and moral consideration we have, they are not due to a magical material forever out of reach of mortal engineers. Likewise, the ability to have goals has been stripped of its anthropomorphism; we now have a perfectly good science addressing the construction of non-magical physical systems with goals [11,12].

For these reasons, we have argued for a continuity thesis [13-15] that implies a research program into the commonalities that all agents, no matter their material composition or origin story, will have in common, and the mechanisms by which they scale and acquire goals.

Answer these questions before forming strong opinions

Whatever the term in question (e.g., sentience, intelligence, mentality, cognition, etc.), what is the definition you are using? Is this capacity detectable by external observers (3rd person science) or only from the internal perspective of the system itself? Provide an example of a strategy or test that measures this capacity in a given system, whether that system is biological, technological, or a combination thereof. Is the measure direct (i.e., it measures the capacity proper) or indirect (i.e., it measures a *correlate* of the capacity)? Is functional capability on some performance test sufficient evidence, or are there aspects of the system's origin story and composition that make all the difference (and can be specified)?

If making claims about machines, AIs, or computers, define the term (making sure the definition is pertinent to current machines, which can be made using evolutionary algorithms, can self-modify, etc.). It is especially essential to situate one's claim in a real or anticipated time and space: whatever limits are being proposed, are they intended to apply to today's machines/AI or *any possible* machine (including future ones exploiting biological parts, evolutionary design strategies, etc.)?

State explicitly what is, if any, the salient difference between cognitive beings and machines that display features of cognitive capacity but are in some sense a simulacrum of the real thing. Is it the material, the causal architecture, the process by which they come to be, or something else entirely, that makes the essential difference? Likewise, if one is referring to "life", as a counterpoint to machines, does the definition of "life" apply to evolved beings that use a completely different substrate, or to ones that incorporate designed materials?

Do you view the distinction as binary (e.g., "living vs. machine", "understanding vs. convincing imitation", "mechanism vs. conscious"), or as a continuum? If the view is a binary distinction, can you indicate the stage of development at which this capacity appears in humans, and the amount of time that is required for it to develop before it can be measured? Did the capacity originally develop in humans? If so, was it an early or late development for the species? If not, which ancestral species shared the capacity?

These initial questions (formalized in Figure 1), which by no means constitute an exhaustive set, are intended to promote epistemic alignment, not consensus. While there will always be disagreements about particulars, it is important that every position is clear, even if clarity exposes the vulnerabilities of an argument. Indeed, part of this exercise involves admitting ignorance when answers are unavailable or ambiguous. A shared set of questions and answers can help formalize the role of humility in exploring these important topics, accelerating productive conversations.

Confronting hybrids: future-proofing theories

Systems that combine biological and mechanical parts represent ideal challenges to the types of assumptions that are central to debates about the exclusivity of mental functions including intelligence and consciousness. As an exercise in clarification of one's views in terms of what features are taken to be definitive, it is helpful to think about the future of currently-existing technologies which hybridize naturally-evolved tissues with engineered components (reviewed in [4,13,16,17]). Contemporary examples of genuine hybrids include humans with sensory/cognitive augmentation devices (cyborgs [18]), brain-computer interfaces [19], hybrots (living brain tissue driving an engineered body or simulated avatar [20-22]), and bioengineered biobots [23,24].

Consider the example of a human with 10% or 85% of bodily tissues replaced with electronic components. Is this hybrid system living or a machine? If the same proportions of human brain tissues (10%, 85%) are replaced with equivalent electronic components that do not change any of the behavioural correlates that are necessary for the inference of phenomenal experience or consciousness, is the hybrid likely conscious or is it likely to be mechanically simulating consciousness? In both cases, at what point does the hybrid either collapse to one of the categories or give rise to a new category?

If the living-machine or conscious-mechanism dichotomy are taken to indicate sharp natural kinds, suggest phase transitions, indicate which parameter is driving the transition, and state whether the phase transition can be zoomed into by very small steps of this parameter (to make the transition seem gradual) or whether it's an atomic (e.g., quantum) event that truly cannot be subdivided into intermediate stages by the smallest possible steps along the continuum.

If instead of a phase transition, you suggest a continuum, indicate:

1. What are the key parameter(s) that establish the continuum
2. Are there any special regions that are qualitatively distinct along the continuum? If so, what do they indicate?
3. How narrow or broad is the continuum in terms of scale and complexity? Is there anything that is not on the continuum (e.g., elementary particles)? If so, indicate at which scale consciousness manifests:

Scale	**Check all that apply**:
≤ Nanometer (e.g., proteins, particles)	
Micrometer (e.g., cells, transistors)	
Millimeter (e.g., neural circuits, integrated circuits)	
Centimeter (e.g., networks)	
≥ Meter (e.g., organisms, groups)	

Next, consider the relevance of material composition. If you are claiming that human-level mental functions cannot be implemented by a machine, does the brain exclusively obey the laws of chemistry and physics? If not, what key properties of this immaterial component are important for mental functions and cannot be replicated in other media of any kind? Are you claiming that biological beings have capacities that cannot be implemented by synthetic/engineered constructs? If so, what compositional property makes it so?

Finally, state whether the origin of the system is a relevant factor when considering the likelihood that it will be conscious. What, if anything, is special about the evolutionary process that makes it more likely than rational engineering to produce real minds? What would need to be true of machines to enable their implementation of cognitive functions such as consciousness? Are these properties possible to engineer? If so, are they possible in principle or in practice?

What next?

Anyone whose worldview provides clear answers to the above questions has earned the right to express strong opinions on the machine intelligence debate. However, our experience has been that convincing answers to these are not yet readily forthcoming; thus, we urge caution in committing strongly to any positions at a time when so much of the science and philosophy is uncertain. The only thing that is clear at this point is that a lot of terminology left over from pre-scientific times feigns wisdom but offers nothing but pseudoproblems. Comforting, plausible categories of “organism”, “machine”, “true understanding”, etc. – arising during past times of limited capability and imagination - will not survive the next decade. They need to be replaced by a principled science of emergent minds in diverse media that take seriously advances in developmental, evolutionary, behavioral, and synthetic biology, and offer a roadmap for rational and ethical synthbiosis with possible forthcoming truly diverse intelligences. It is far too early to have strong opinions in this field, because so many questions about scaling of cognition remain open. Enormous practical implications for fields ranging from regenerative medicine [25] to AI, as well as the moral imperative of a more inclusive, scientifically-founded ethics, await a concerted effort to advance our understanding of embodied minds.

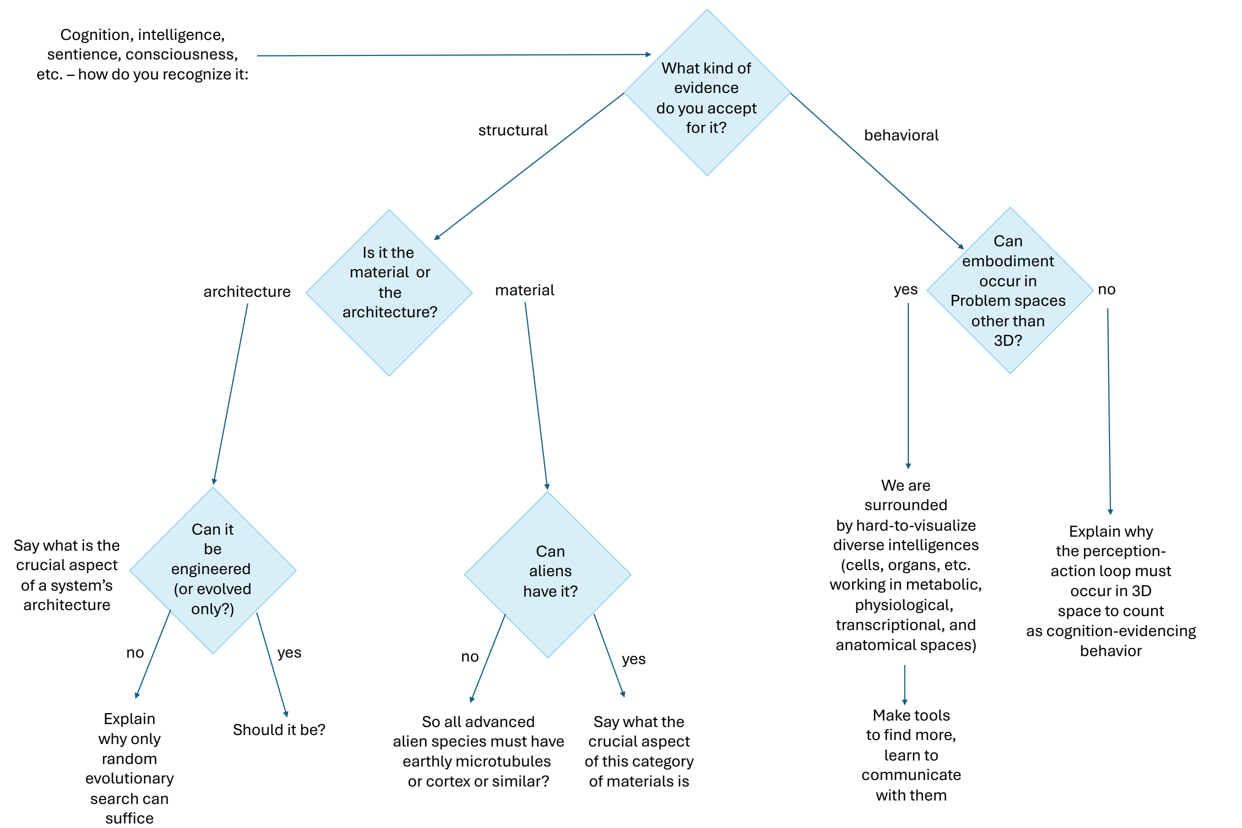

Figure 1: a flowchart of questions designed to clarify positions on machine intelligence

Having chosen a specific property in question (cognition, intelligence, sentience, consciousness, etc.), begin at the top and answer each question in the blue rhombus, proceeding on to the next one as indicated, based on the answer. The top layer forces a decision as to what kind of evidence one plans to use to determine the status of a given engineered, natural, or hybrid system. The next layer drives a specification of what key features contribute to that evidence – the material, its architecture, or the space in which it operates. The bottom layer asks specific questions about its possible origin. Each answer leads to a suggestion, based on the opinion defined by the prior answers, as to the implications of that view and what it requires next, to flesh out its consequences.

References

1 McShea, D. W. Machine wanting. *Stud Hist Philos Biol Biomed Sci* **44**, 679-687 (2013). https://doi.org:10.1016/j.shpsc.2013.05.015

2 Baluška, F. & Levin, M. On Having No Head: Cognition throughout Biological Systems. *Front Psychol* **7**, 902 (2016). https://doi.org:10.3389/fpsyg.2016.00902

3 Fields, C. & Levin, M. Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments. *Entropy (Basel)* **24** (2022). https://doi.org:10.3390/e24060819

4 Clawson, W. P. & Levin, M. Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms. *Biological Journal of the Linnean Society* (2022). https://doi.org:10.1093/biolinnean/blac073

5 Trinh, T. H., Wu, Y., Le, Q. V., He, H. & Luong, T. Solving olympiad geometry without human demonstrations. *Nature* **625**, 476-482 (2024). https://doi.org:10.1038/s41586-023-06747-5

6 O'Grady, R., Christensen, A. L. & Dorigo, M. in *Morphogenetic Engineering: Toward Programmable Complex Systems* 27-60 (Springer, 2012).

7 Levi, P., Meister, E. & Schlachter, F. Reconfigurable swarm robots produce self-assembling and self-repairing organisms. *Robot Auton Syst* **62**, 1371-1376 (2014).

8 Abdel-Rahman, A., Cameron, C., Jenett, B., Smith, M. & Gershenfeld, N. Self-replicating hierarchical modular robotic swarms. *Communications Engineering* **1**, 35 (2022). https://doi.org:10.1038/s44172-022-00034-3

9 Aubin, C. A. *et al.* Towards enduring autonomous robots via embodied energy. *Nature* **602**, 393-402 (2022). https://doi.org:10.1038/s41586-021-04138-2

10 Hayes, J. Technology laboratory automation: Labs go auto. *Engineering & Technology* **16**, 58-60 (2021). https://doi.org:10.1049/et.2021.0703

11 Heylighen, F. The meaning and origin of goal-directedness: a dynamical systems perspective. *Biological Journal of the Linnean Society* **in press** (2022).

12 Busseniers, E., Veloz, T. & Heylighen, F. Goal Directedness, Chemical Organizations, and Cybernetic Mechanisms. *Entropy (Basel)* **23** (2021). https://doi.org:10.3390/e23081039

13 Rouleau, N. & Levin, M. The Multiple Realizability of Sentience in Living Systems and Beyond. *eNeuro* **10** (2023). https://doi.org:10.1523/ENEURO.0375-23.2023

14 Bongard, J. & Levin, M. Living Things Are Not (20th Century) Machines: Updating Mechanism Metaphors in Light of the Modern Science of Machine Behavior. *Frontiers in Ecology and Evolution* **9** (2021). https://doi.org:10.3389/fevo.2021.650726

15 Levin, M. Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds. *Front Syst Neurosci* **16**, 768201 (2022). https://doi.org:10.3389/fnsys.2022.768201

16 Pio-Lopez, L. The rise of the biocyborg: synthetic biology, artificial chimerism and human enhancement. *New Genetics and Society* **40**, 599-619 (2021). https://doi.org:https://doi.org/10.1080/14636778.2021.2007064

17 Orive, G., Taebnia, N. & Dolatshahi-Pirouz, A. A New Era for Cyborg Science Is Emerging: The Promise of Cyborganic Beings. *Adv Healthc Mater* **9**, e1901023 (2020). https://doi.org:10.1002/adhm.201901023

18 Mehrali, M. *et al.* Blending Electronics with the Human Body: A Pathway toward a Cybernetic Future. *Adv Sci (Weinh)* **5**, 1700931 (2018). https://doi.org:10.1002/advs.201700931

19 Li, G. & Zhang, D. Brain-Computer Interface Controlled Cyborg: Establishing a Functional Information Transfer Pathway from Human Brain to Cockroach Brain. *PLoS One* **11**, e0150667 (2016). https://doi.org:10.1371/journal.pone.0150667

20 Potter, A. M. *et al.* in *Proceedings of brain inspired cognitive systems.* 1-5.

21 Potter, S. M., Wagenaar, D. A., Madhavan, R. & DeMarse, T. B. Long-term bidirectional neuron interfaces for robotic control, and in vitro learning studies. *P Ann Int Ieee Embs* **25**, 3690-3693 (2003). https://doi.org:Doi 10.1109/Iembs.2003.1280959

22 Warwick, K., Nasuto, S. J., Becerra, V. M., Whalley, B. J. in *Computing with Instinct* Vol. LNAI 5897 (ed Yang Cai) (Springer, 1998).

23 Kamm, R. D. *et al.* Perspective: The promise of multi-cellular engineered living systems. *Apl Bioeng* **2**, 040901 (2018). https://doi.org:10.1063/1.5038337

24 Pagan-Diaz, G. J. *et al.* Engineering geometrical 3-dimensional untethered in vitro neural tissue mimic. *Proc Natl Acad Sci U S A* **116**, 25932-25940 (2019). https://doi.org:10.1073/pnas.1916138116

25 Lagasse, E. & Levin, M. Future medicine: from molecular pathways to the collective intelligence of the body. *Trends Mol Med* (2023). https://doi.org:10.1016/j.molmed.2023.06.007