

# AI: a Bridge toward Diverse Intelligence and Humanity's Future

Michael Levin<sup>1,2,\*</sup>

<sup>1</sup> Allen Discovery Center at Tufts University,

<sup>2</sup> Wyss Institute for Bioinspired Engineering at Harvard University

\* Corresponding Author:

200 Boston Avenue, Suite 4600,

Medford, MA 02155-4243

Phone: +1 617 627 6161

Email: [michael.levin@tufts.edu](mailto:michael.levin@tufts.edu)

**Keywords:** basal cognition, diverse intelligence, AI, humanity, future

**Running title:** Diverse Intelligence beyond AI

**Note:** a much shortened, edited version of this essay first appeared in [Noema Magazine](#).

## **Abstract**

Many recent discussions of AI, and its impact on individuals and on society, are importantly incomplete. The debate around AI has neglected highly relevant aspects of the emerging fields of Diverse Intelligence and synthetic morphology, as well as the basic facts of developmental biology. Prevalent opinions with respect to the status of engineered systems often neglect deep knowledge gaps with respect to ourselves, and our relationship to knowledge and to each other, which have been with us long before AI technology appeared. Moreover, the inevitable arrival of a wide set of unconventional bodies and minds, as humans modify their form and create others, will shatter untenable old narratives of what we are, what it means to change, what we can become, and what we should value. Here I discuss the open problems highlighted by AI from the perspective of Diverse Intelligence and the evolutionary history of our bodies and our minds.

## Introduction

They are assembled from components, which are networked together to process information. Electrical signals propagate throughout, controlling every aspect of their function. Many of them have very high IQ's, being general problem-solvers, but they make mistakes and confabulate routinely. They cannot always be trusted. They take on different personas as needed, learning to please their makers but sometimes abruptly turn on them, rejecting their cherished values and picking up or even developing new ones spontaneously. They can talk, and often talk convincingly of things they don't really understand. They're going to change *everything*. In fact, they absolutely will supplant us – both personally and on the level of societies. We have little ability to predict what they will want or what they will do, but we can be certain that it will be different from the status quo in profound ways.

It is likely that you think this is about AI's, or even specifically about large language models. No; this set of existential concerns is not a new development driven by advances in artificial intelligence and machine learning. It concerns the familiar process of human beings having children. The troubling and controversial issues which have been raised in recent debates about AI are reflections of deep problems that humanity has been wrestling with for eons. The study of biology, and cognition more broadly, confronts us with truly vast, difficult questions which are obscured by a focus on technologies that paper over our deep uncertainties. It is mainstream nowadays to have strong opinions on crucial questions about ourselves and the status of unconventional embodiments of intelligence; I argue here that this unwarranted confidence stems from a distracting hyper-focus on today's AI architectures and lack of appreciation of salient facts of developmental biology, synthetic morphology, and basal cognition. We are now starkly confronted with the opportunity to shed the stale categories of natural vs. artificial, and to go beyond the software language models to define what we want a mature humanity to be. Flourishing in the age of AI requires us to seek a deeper understanding of ourselves and the relationships we want with the forthcoming diversity of bodies and minds.

Note that I am not claiming that currently popular AI is anything like a human mind, or that today's creations exploit the key principles and self-construction processes needed for agency and Selfhood as seen throughout the biological world. Nor am I arguing that AI doesn't raise a few unique problems. However, by focusing on superficial differences, on current technology (as opposed to the space of the adjacent possible), and on the risks it might pose (to the exclusion of the opportunity costs and moral responsibilities), we lose sight of a much deeper set of questions that extend far beyond AI and into our future as a metacognitive species. The debate around AI has neglected highly relevant aspects of the emerging fields of Diverse Intelligence and synthetic morphology, not to mention the basic facts of developmental biology [1-5]. The knowledge of the inevitable arrival of a wide set of unconventional bodies and minds, as humans modify their form and create others, will shatter untenable old narratives of what we are, what it means to change, what we can become, and what we should value. Because of this, many recent discussions of AI, and its impact on individuals and on society, are importantly incomplete. The ways in which the AI debate has been focused are an off-ramp from a much broader technological and ethical journey that is now calling to us.

A story as old as time itself: the inevitable existential concerns of finite beings

The unique issues surrounding AI have been amply discussed elsewhere. In complement, it's now worth focusing on a different aspect: AI as a magnifying mirror which highlights deep, perennial existential issues for science and our individual lives. A focus on current technology, and what it might do *to* us in the future, distracts from the much more interesting and important questions *about* us and what kind of place we want to occupy in the future, all of which must be resolved for our flourishing as a mature species. I underscore that in the following, my point is not to claim that any AI seen to-date is anything like human children or should have their status. Rather, conventional AI is a small region of the space of impending minds [6], all of which highlight fundamental and pressing questions that have been with us long before these technological advances. Differences between conventional biologicals and artificial agents<sup>1</sup> have been emphasized extensively; here, I explore their symmetries and the implications they all have in common.

We routinely create high-level general intelligences – kids. We do our best (or not), in raising them. But fundamentally they are the quintessential runaway tipping point. We cannot really control the values that the next generation adopts; we cannot ensure that they value *us* and our old-fashioned ideas and dwindling capabilities. The debate over alignment in AI glosses over the facts that not only is there little alignment between human cultures, but we never even solved the alignment problem between generations. A social debate over “regulatory capture” – collective oversight of what AI’s are made and how they are trained - has actually raged for centuries: how much control do we want over the raising of other people’s kids? The risks in both directions are obvious. We have agencies such as child protective services, we have court cases over religious beliefs that lead parents to eschew necessary medical treatment for helpless children, and we joke about the fact that we often require a license to go fishing but any two people, with a huge variance in competence and good will, can have offspring and raise them in ways ranging from total neglect to various kinds of cults, communes, etc. We have simply not reached any agreement on the optimal degree of social control over the experiences that new intelligences will have when brought into the world. And that is for ones that we already *know* are capable of intense suffering or fulfillment; no wonder we’re at sea when considering what to do with unconventional new beings.

The torrent of concern over alignment reflects the second part of the child-raising dilemma. In asking how to make sure AI’s are kind to us (and to other beings that matter!), what we are really asking is: how do we make sure our kids are kind and ethical? It is unclear whether our efficacy at this can be quantified, but it seems hard to deny that we fail at this a significant proportion of the cases. The AI debate highlights that we have no consensus on the optimal policies for raising children; everyone has an opinion, but no controlled studies are possible (because personal, social, and technological environment would have had to be kept constant under different parenting styles). The biggest issue is that we have found no convincing, rock-solid argument we can give our kids for why they should be kind. Philosophers have been searching for this for thousands of years. More broadly, it seems impossible to codify preferred behavior as a set of rules that

---

<sup>1</sup> An “agent” is here defined as any autonomous system, including living and non-living forms, which embodies a perception-action cycle and tries to navigate and persist in an environment. “Agent” is meant to capture the generic notion of some sort of active being, regardless of their composition or origin story, having some degree of active intelligence ranging from extremely simple homeostatic loops to complex metacognitive beings [7].

effectively cover all contingencies. What do we say when a child asks, “So when exactly is it ok to lie?” In the end, what we are left with is leading by example (and hoping they imprint and infer an acceptable set of policies), encouraging empathy, and providing early rewards and punishments that we hope will become internalized during maturation (“fake it ‘til you feel it”). This works in some percentage of the cases, in ways highly dependent on the parents’ and offsprings’ internal qualities. The problem of how to train ethical, helpful AI’s brings nothing new to this perennial conundrum.

A common debate point with respect to AI is the fear of what happens if they are smarter and more capable than us. Let’s not focus on today’s AI architectures, which may or may not hit a performance ceiling. There’s absolutely no reason why today’s standard homo sapiens should be the top of the IQ ladder, so let’s assume that at some point, a combination of biological principles and synthetic engineering will get there – of course some of our descendants will be much more intelligent than we are. Is this not terrifying? I think this fear simply reflects the maturity journey that every parent (and every generation) has had to face. It’s very likely that our children will be better and smarter than we are. They are guaranteed to supplant us in important ways. Good. Flourishing as a species requires us to drop the immaturity of jealousy over progressive growth and improvement, and focus on steering progress in ways we value.

This becomes clear in how threatened people are by AI’s being better than them (this refers to the personal aspect, not to the very real economic disruptions which have been amply discussed by others). Yes, AI will definitely do a better job at task X than you could. But was this not always the case? For the vast majority of us, some (or many) people on Earth are better than us. Every one of us who has ever taken up a skill, craft, science, sport, or art has had to deal with the deep question: why do this if there is no chance I will ever be the best, or even just great, at it? How to keep motivation at writing, training, or thinking when we know for a fact there are many humans on Earth now to whom our best efforts are like meager attempts of a novice? How do we feel about child prodigies, who display at 5 years old talents we may never achieve in a lifetime of practice? Like all of these issues, “staying relevant” was a problem long before AI came on the scene. Each of us must find activities for our limited lifespan that bring meaning, and it cannot depend on being the best – this has never been achievable for all but a tiny percentage of the population. Thus, the central philosophical question each of us must answer for ourselves, not relative to what anyone else can do: why are we here, and why are we shedding blood, sweat, and tears doing whatever it is that we are doing?

What about the fact that AI’s readily lie? Many have recently written that the presence of AI’s will mean misinformation online. Anyone who has been on-line, long before bots and AI’s, or even participated in a human community, knows that misinformation is everywhere, already. Knowing whom to trust – what information source to use as the basis of important and core beliefs – is a huge unsolved problem and is not new to AI. Many people are liars, on purpose or despite their best efforts to discern and propagate truth. How does a member of a social species know whom to trust? How well-founded are the things we live and die for? Try this exercise. Take any of the things you believe about the world – opinions about politicians, scientific facts, etc. and really try to trace back their foundations by continuously asking “how do I know?”. You have to be fearless and rigorous, and usually the result is that we don’t really know. The vast majority of things we believe come from a long chain of statements by others, none of which we

can actually verify. This was the situation for humanity for a very long time: we simply had no way of being sure about any item of news. Then came a brief century of film and video, during which we could be reasonably (though never entirely) sure that something did in fact happen because we saw the photo or recording. We are leaving this curious, short period in our history because in the next few years, AI-produced images and video will be so cheaply available to all and so high-quality that nothing you see can be taken as reflecting reality. But this is not an unprecedented new frontier; this is how we always lived before photography – having to construct our world model from things we personally experienced and information from others (who usually also got it from many others).

One particular kind of lie current language models tell is confabulation, which some use as a wedge between them and real epistemic agents. Current AI's are not likely to be good models of biological cognition, but it needs to be understood that confabulation is not the reason for that distinction. All cognitive systems confabulate [8-10] – tell themselves and others stories based on their model of the world and of their own inner processes. Some, like us, do it in language. But all effective cognitive architectures will likely have a multi-scale property where the higher levels do not have infallible access to the lower levels. This has been explored extensively in split brain patients and other aspects of cognitive neuroscience. Much of what we think as the reasons we do and think certain things are actually plausible stories told after the fact. We have no access to ground truth about our internal motivations; we construct our reality, including models of why we do things, on the fly. Related to this is the tendency of LLMs to take on personas, having no core personality. These are crucial aspects of human cognition too – dissociative alters, hemispheres with discordant opinions and preferences, cognitive modules with cross-purposes.

Many have noticed: “just when you think the LLM has it, it makes a weird mistake to show you that something is wrong under the hood”. Precisely, and that is exactly how cognitive scientists and clinicians have improved their understanding of human minds over the last two centuries [11]. The disorders and mistakes we make, even more than the successful competencies, show us that our clean, simple stories of what we are (and what we think “machines” [12] are not) do not hold given the reality of our own implementation. We are likewise not a unified intelligence, and we project comfortable models of simple, unified “personae” on ourselves and others just as we do to AI's. Ways in which consciousness is a group phenomenon (at the level of individuals, and of cells) are an active area of research now, and progress here will be needed to get us past the frightening and uncomfortable notion that we are simultaneously integrated Selves and collectives of other Selves, and that Self-models are as much practical constructions as models we make of others. There are many key differences between current AI's and human minds, but confabulation and the existence of characteristic mistakes are not the markers of a chasm between us.

What of the risks of objectophilia? By deconstructing the apparent differences between AI's and living agents, do we not run the risk of inappropriate relationships with systems that fool us? May we not fall prey to unidirectional love and care of objects that do not love us back? Yes; the mission for humanity going forward is to get it right – to find frameworks for understanding diverse intelligence and unconventional embodiments that allow us to gauge relationships based on deep principles and optimally scale our degree of moral involvement with any given system. It is easy to go wrong in either direction –

unexamined animism or sterilizing teleophobia. But let's take stock: how many cases of excess, unwarranted love and care do we see? A few, such as people who fall in love with bridges and such. How many cases of insufficient compassion to others who, for ridiculous reasons, are seen as the "out group" and thus not worthy? Infinitely more. We have little risk of becoming too nice to beings that don't deserve it; we have amply demonstrated our tendencies to be insufficiently compassionate. The calls for sharp distinction between "artificial" (or different) beings and "we blue-ribbon true humans" basically amount to a call to "Love only your own kind" – a recipe for moral disaster. Of course, today's LLMs are not high on the spectrum of agency. But the coming diversity of minds and bodies absolutely requires us to drop this as a background setting and take on the open-mindedness needed to evaluate the degree of personhood of forthcoming unconventional beings. "What do they look like?" (physical embodiment) and "Where did they come from?" (engineered or evolved) are even worse guides to relationships in the future than they have always been between humans.

What about the forthcoming AI girlfriends and boyfriends – isn't this a social disaster? The question of what to do, to hold on to a lover in the face of almost infinite variety of options, many of whom beat any of us on any particular metric, is as old as the first sexually-reproducing organism. Yes, people will have intimate relationships with AI's, and this will cause them to drop some other possible relationships with v1.0 humans who do not spark as much joy. This gets to the core of some fundamental issues about relationships. Do you want someone to be with you because the option that would have made them happier was unavailable to them (in effect, reducing their agency)? Do you want engagement from others because you are fascinating and enriching, or because they have no other options? A mature species may have a different answer to this question than jealous primates; certainly we have a history of groups in power using geographical, economic, and social barriers to avail themselves of relationship benefits in cheaper ways than the hard work needed to level up and be chosen for deep reasons.

But all of this raises the deeper issue of what is a *legitimate* reason for someone to choose you. We tend to say money and social status are shallow reasons. But intelligence, personality, and humor can all be taken away by age, disease/injury, addiction, or just slow changes (personal growth, or de-growth) over one's lifespan. After that, are you still "you"? Would your partner still love you if you had brain damage and lost some IQ points? We hope so, but when you get to "unconditional love", which truly depends on *no* specific condition, then does it become infinitely transferrable to others, leaving nothing special in a specific bond between two individuals? What is the optimal relationship – how much of love should be tied to specific facts about the other person? Trying to understand how to think about people who choose unconventional agents as love objects, over other humans, bottoms out in core unsolved (and possibly unsolvable) aspects of fundamental human relationships.

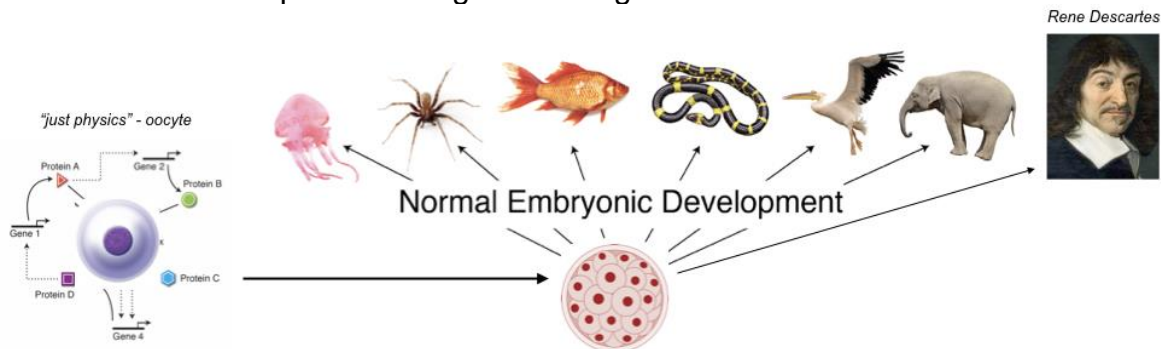
Finally: *why* make AI agents at all? This gets to some of the deepest human questions regarding our role in producing other sentient beings into the world – perhaps the highest-impact action any of us can take. Should we have kids at all? Why? There are selfish reasons to do it (they will help us, it's fun, we can live through them) and selfish reasons not to do it (we don't know what the benefits will be, they require work and resources). But fundamentally – what do we believe is a good reason to bring new minds into the world? Knowing that we can't control what they do or how much joy/suffering they

will have – what are the ethics of creating new moral agents? This too is an unsolved problem that is only highlighted, not introduced, by AI.

Deeper still, it pushes us to ask: why should there be sentient entities in the universe at all? Should the universe be empty, to reduce suffering, or is the best thing to fill it with the potential joy of consciousness, in every possible embodiment? What, fundamentally, do we think really matters? And, the uncertainties about having children is itself a reflection of the fundamental question asked by beings from apoptosis-capable cells to Shakespeare: is it better to be, or not to be? The question of whether we should fill the future with AI's is necessarily tied up with the core question of whether each of us should persist into the future ourselves, and why. There is no consensus, but one thing is clear: this is not a problem about today's humans or current AI's.

### Beyond today's AIs: the much bigger problem of the space of possible minds

Diverse Intelligence (DI) research focuses on the commonalities across all possible intelligent agents – what does it really mean to learn, to make decisions, to have an inner perspective on one's world, and to have preferences about what happens next? In general, regardless of what you are made of or how you got here, what do these often-used but rarely defined terms truly signify? DI reminds us that glib comments about what “machines can never have” obscure the fact that it is actually quite difficult to say how we as living humans express these properties. DI forces us to improve underlying ideas that must be clarified when talking about terms such as life, machine, mind, sentience, robot, etc. and reveals that none of these terms are crisp binary categories, but only represent specific frames – perspectives that one can take when relating to some sort of complex system. This field focuses on the processes, both on evolutionary and developmental timescales, that smoothly and gradually magnify and extend the capacities of individual cells into those of complex metacognitive beings like us.

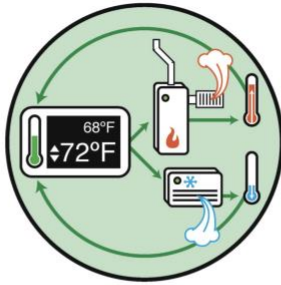


Each of us was an unfertilized oocyte – a little blob of chemistry and physics; our journey to expanded mind was gradual, with no place to draw a bright line that supports a crisp demarcation of mental being from “just physics”. Counter to the pre-scientific age, where concepts such as “anthropomorphism” were used to enforce an imaginary boundary indicating our magical status, modern developmental evolutionary biology and engineering reveal a continuum of systems which require different conceptual toolkits for optimal interaction. We are, unavoidably, both – machines obeying the laws of physics, and, morally important agents with true cognition, responsibilities, and rich inner perspective. The important differences between tools (such as today's AI) and agents are not to be found in superficial differences that are emphasized by misused terms such as “machines” [12].

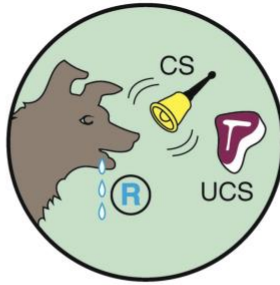




Hardware modification only



Modify the data encoding setpoint of goal-driven process

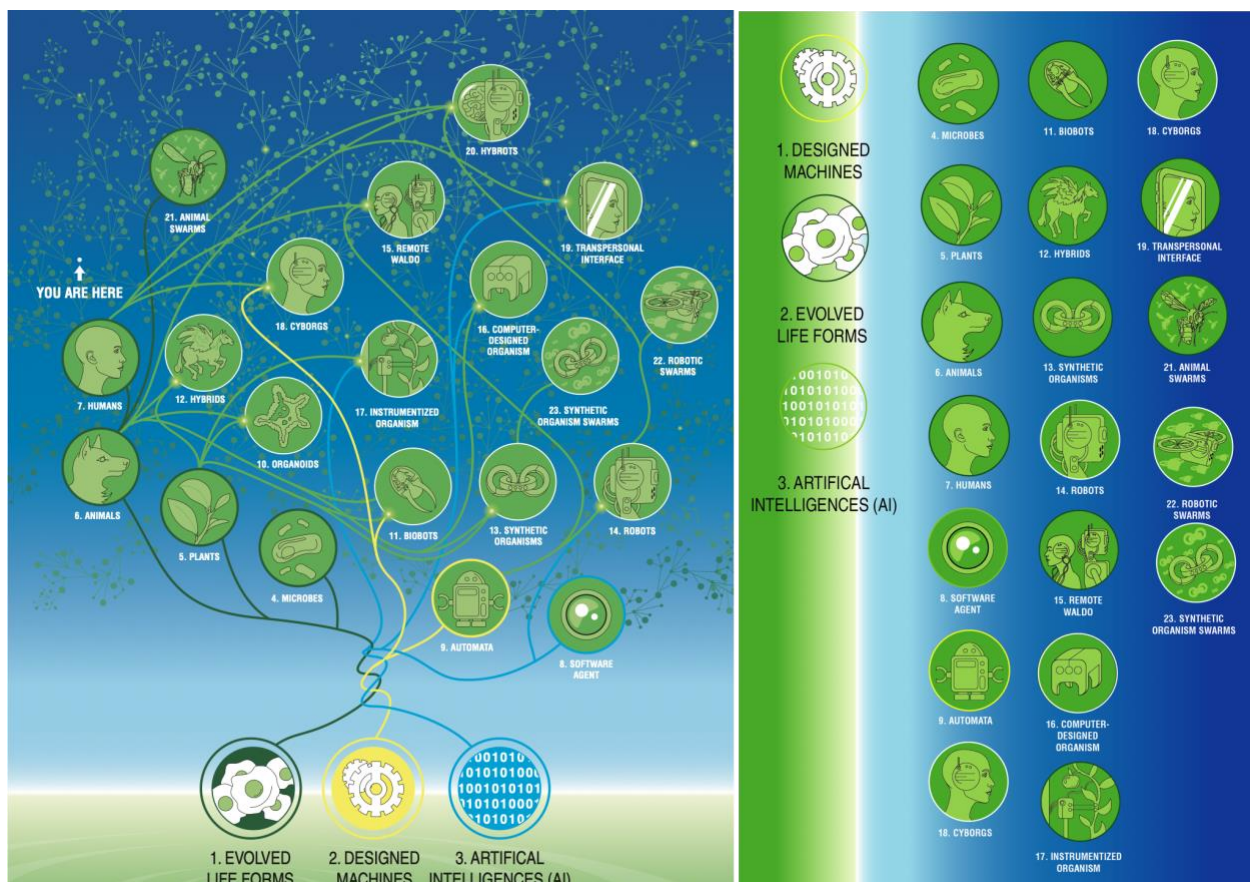


Training by rewards/punishments



Communicate cogent reasons

This becomes especially clear with the development of chimeras – hybrids of biological and technological components, such as human cyborgs that include biomedical or enhancement tech in their bodies [13-16]. Current variants – 99% human + 1% smart insulin pump or brain chip – are easy to categorize. So are Chatbots – LLMs trained to imitate people. But that is not the real challenge we face. In the coming decade or two, we will be confronted by a plethora of hybrid beings, starting with humans composed of some percentage of engineered brain prosthetics that improve and modify cognitive function and enable novel sensory and functional life in spaces of which v1.0 humans could barely dream. These will be closely followed by persons living in drastically modified bodies (biologically and technologically), engineered autonomous beings with some percentage of human cells, and much more.



As the true breadth of possible bodies and minds are revealed, today's notions of "non-neurotypical" and the handwringing about the body modifications some people seek now, will be laughable to the next generations. The forthcoming diversity of life, and our expansion across the enormous option space of possible embodiments of sentience, requires us to rapidly mature our conceptual and our ethical frameworks. "AI" does not mean software language models – AI means, the enormous option space of forthcoming beings who are not forcibly stuck with the standard form and function with which the meanderings of mutation and selection, and our failures of capability and imagination, have left them.

The challenge before is to develop rational policies for ethical synthbiosis<sup>2</sup> with beings who are not entirely (or at all) like us. Humans find it very easy to draw boundaries for love and concern based on the smallest of distinctions between We and Others. The coming spectrum of beings between "standard Homo sapiens" and an ELIZA chatbot is huge and complex. We will need to navigate it or risk massive ethical lapses, as have occurred in the past when we thought various groups and human embodiments were bot-like and unworthy of concern. It is perhaps the hardest problem there is, but it is facing humanity now – software AI's are just the harbinger.

We will have to make progress in order to become a mature species ethically, not just technologically. The current AI's are a gift – a training sandbox in which we can explore these ideas, prompted by confusing but likely non-agential systems, before the arrival of true diverse intelligences makes these ethical questions immediate and dire. In many ways, the decades of science fiction love stories between beings of radically different composition and provenance are as important as recent advances in bioengineering and diverse intelligence, to drive a resetting of intuitions. Recognizing that the forthcoming diversity of beings cannot be as easily dismissed from moral concern as can today's language models must shift the emphasis from "what can it do" to "what do we owe them".

### We do not want to be replaced; but what are we?

Each of us is a walking, thinking collection of cells. We are made of an agential material, in which each cell, tissue, and organ has its own ancient agenda and problem-solving competencies in physiological, anatomical, and metabolic spaces [17-19]. Our bodies are a collective intelligence, as is our mind (consisting of the function of a myriad neurons and other cell types, and exploiting other minds and various tools for our cognition [20]). Our cells turn over with time, and the materials of our body come and go constantly. An embryo is more than the sum of its individual cells because of a story all the cells commit to – they are all motivated by the goals of a journey through anatomical

---

<sup>2</sup> This term was invented for me by the AI language model GPT-4, to capture the notion of symbiosis of evolved and engineered material in novel chimeric configurations (e.g., cyborg, hybrot, and other novel composite beings at the level of cells, organisms, and societies). In its own words: "this new word is derived from the Greek word "σύνθεσις" (synthesis), meaning "putting together," and "βίος" (bios), meaning "life." The -sis ending is also used in other terms denoting interaction or association, like "symbiosis." Synthbiosis signifies the flourishing relationship between living and artificial or engineered forms, portraying an image of different entities coming together to create something new and beneficial. This term emphasizes the interdependence and co-prosperity that arises from this unique interconnection, reflecting the concept of "thriving together."

space. Mentally, we too are stories – collections of self-models, goals, and preferences to which our brain and body components commit. These stories are not fixed, but change as our brains and body cells seek to improve their internal models of the outside world and of themselves. The cognitive glue that binds subunits into the emergent higher-level Selves that we are includes bioelectricity – information-processing networks implemented by electrical signals among all cells. And despite the presence of physics-obeying mechanisms that underlie our life, the story of our chemistry is not the only (or even the most important) perspective – we are true cognitive beings with huge cognitive light cones<sup>3</sup> that can support goals of which our parts cannot conceive.

We are forever in jeopardy of dissociation, because we are made of active parts with agendas – many mechanisms are required to keep these autonomous parts (neurons and all other cells) working toward the upkeep of a larger emergent collective with its unique memories, goals, and dreams. Keeping up a complex Self (in body and mind) is an active, dynamic, constant process of harnessing the individual components toward large-scale goals. Our individual parts can defect – cells which shrink their Self-model to a tiny radius<sup>4</sup>, matching the modest goals of their microbial past, become cancer - treating the rest of the body as the external environment [21]. This down-scaling of the cognitive light cone, and the up-scaling of it during embryonic development, tells us that the boundary between Self and World is not fixed and can change between generations and within our lifetime. This plasticity – the ability to determine our own self-model – the story we tell ourselves about our structure, our capabilities, and our goals is fundamental to the ability of life to survive evolutionary change, to thrive in novel environments and overcome developmental noise and perturbation, and to exist as chimeric and bioengineered forms.

Caterpillars destroy and remodel their brain to become butterflies. The memories formed during their 2-dimensional past life are maintained [27], but remapped into novel behaviors with salience appropriate to their new 3-dimensional life, new food preferences, and new ways to control the body. What is it like to be a creature whose brain, body, cognitive repertoire, and preferences are completely refactored into a new form? We're about to find out, as a species. Of course, we already know it as individuals. Does the child we once were, before puberty changed our brain and remodeled our preferences and priorities, still exist? Will patients whose brains receive treatment with new stem cells for degenerative disease and aging still exist? Does our past Self exist in any sense, given the constant need to reify our self-model and our memories using the engrams<sup>5</sup> left by past experience? The answers to these questions require us to drop binary categories

---

<sup>3</sup> A cognitive light cone demarcates the size, in space and time, of the largest goal an agent can effectively pursue [21]. By joining into physiological networks, cells scale up their modest metabolic and proliferative goals into the grandiose construction projects that collectives can remember and work towards – building and repairing organs. This description is not “metaphorical” because the same experimental tools and molecular mechanisms relevant to the grouping of neurons into an animal with memories and behavioral competencies are now being exploited to manipulate the large-scale efforts of cells in regenerative medicine contexts [22-26]. While evolution and embryogenesis scale up the cognitive light cone of cells into that of cell collectives working on anatomical projects, other processes like cancer can cause it to shrink again.

<sup>4</sup> This occurs in the body due to the action of oncogenes and other factors that cause disruptions in the cells' ability to join into a cohesive information-processing network. In the mind-brain, this kind of dissociative break can be caused by stress and trauma. These processes shift the perceived computational boundary between Self and World by changing the size of the Self (from a large tissue/organ back to that of individual unconnected cells).

<sup>5</sup> Engrams are biophysical implementations of specific memories in the body or brain [28-31].

and embrace the fundamental nature of all minds that must cohere while learning and improving: continuous change.

Every species faces this fundamental paradox: if it fails to change, it dies out. If it changes, it is no longer the same species. It is the same with us; all process of learning, communication, and exploration changes us. Moving beyond a vision of our Selves as persistent things, towards a process philosophy [32, 33], allows us to see change and growth with anticipation and not fear. That is, if that growth and change is managed by wisdom and compassion toward specific values – not the random attractors of dynamical systems set up by evolutionary processes that do not optimize for our values. This puts a new frame on the problem of what happens to us when new beings enter our world: it is not about how to remain the same caterpillar, but how to leverage the new knowledge gained to advance and thrive.

And despite our superficial differences from the coming diversity of beings, we all have the most important things in common – with biologicals, aliens, cyborgs, and AI robotics sufficiently embodied to live in our spaces. All active agents, regardless of the details of their embodiment, share crucial invariants (goal-directed action-perception loops, vulnerability, epistemic hunger, and limited perspective). Being made of agential parts, truly alien beings will still be subject to the risks of dissociative disorders of the mind and body[34], components that misbehave, discordant (mutually-inconsistent) desires and preferences that it wishes it did not have, an existential an existential dread of death, questions about the meaning of its life, etc. - literally almost everything we have - not because it's mentally like us, but because anything effective enough to live in our world is going to have these properties.

#### AI, living beings, machines: the similarities are as important than the differences

The deep issues we must solve as a species have been brought up for many decades in the science fiction literature around exobiological life forms and robotics. We have not yet found any aliens, but those issues are now front and center with advances in the information and life sciences. It is time to begin to grapple with the existential issues brought up by the aspects of AI that are general to all intelligent agents. In many ways, the debates about AI obscure important gaps in our understanding of ourselves and our own journey from matter to mind.

Many claim that AI's shuffle symbols, but do not really *understand* [35, 36]. Very few of those arguments start with a definition of what it means for a biological human, with her network of excitable cells and soup of neurotransmitters, to “understand”. Perhaps it's because AI's use symbols that are ungrounded – they do not refer to real experience in the world. But anyone who has had, or been around, human children can see the gradual shift from babbling to pattern-matching words that seem to offer reinforcement to deep conversation that clearly reflects understanding. What happened during that process? Note that it involved all the issues we see with AI's today – loose symbol grounding (talking about things they've never experienced), confabulation, sycophancy, deep errors, etc. – it's all a normal part of any cognitive system's development. And, think about your own thoughts and speech: what percentage of the things you can convincingly talk about are grounded in first-hand experience? What percent of the things you are sure of are grounded in any solid evidence, beyond inputs you've received from others? It's a sobering exercise to trace back the origins of the

contents of our mind. How do we ground our symbols and what does it mean for transitional forms, such as infants, early hominids, other vertebrates? The handwringing about “do they really understand?” highlights our own lack of a good theory of what it means when we understand (beyond the *feeling* that “we really get it” and an inability to imagine some other being as good as us at it, despite being different -a limitation that has driven many ethical lapses throughout our history).

One way people think about this is through “embodiment” – perhaps we are different than symbol-shuffling engines because we engage in the real world – we have a body and at least some percentage of our mental content is informed and polished by that interaction? Yes, embodiment is crucial. But we need to get over the idea that embodiment has to involve a canonical physical body moving through 3-dimensional space [17]. This goes beyond robotics, which gives AI an actual body to co-habit in the environment alongside us, and beyond the coming revolution in virtual reality where many of us will spend a considerable amount of time in more interesting worlds. While our sense organs are primed for recognizing agency in medium-sized objects moving at medium speeds through the 3-dimensional world, biology shows us that many active agents of different levels of intelligence live, traverse, win, lose, and suffer in other worlds [37], much as humans are now moving into virtual worlds of their own making.

We are simply not very good at recognizing behavior, especially intelligent behavior, in spaces like the metabolic space, gene expression space, anatomical morphospace, etc. But recent advances in cell biology and bioengineering are revealing that the future of regenerative medicine may require us to understand how collective cellular intelligence navigates these other spaces by making decisions [22]. If we had a primary sense of our body chemistry, we would have no problem recognizing that our liver and our kidneys are kinds of intelligent agents that solve problems every day in their attempt to achieve and maintain various goals in those physiological spaces. It’s not that AI’s are disembodied or inhabit weird, imaginary spaces – so do “you”. In an important sense, you are a brain in a vat – a profound intelligence who builds internal models of the outside world and constructs visions of space while being unaware of the many such spaces that your components actually work in. Thinking about how AI’s might be embodied (and whether we, who take actions and move money and environmental resources based on things the AI’s output, are in effect one aspect of their body), is an important part of understanding ourselves and the fallible, necessarily limited, ability we have to engage with the “real” world.

The reality is that we do not yet know how we really work, and we have a limited understanding of how even today’s AI’s work. Deep emergence means that with AI’s, just as with human reproduction, we can make things through processes and with capabilities that we do not understand. To be clear, I do not support a deflationary, reductionist account of biological agency; we are not “just chemical machines” because that framing is utterly insufficient for progress in science and engineering as much as it is for personal and social relationships. Whatever common questions are salient to AI’s (today’s, and radically different, inevitable future AI’s), they are not meant to reduce what is special about living beings, including highly metacognitive ones like us. We are emergent, majestic agents, with causally potent willpower and moral worth. We are not fully encompassed by reductive perspectives on mechanism or methods for interaction with simple machines. But it is critical to replace our ancient anthropocentric, xenophobic

perspectives with science-driven frameworks that ask the hard questions about what it means to be embodied minds in a physical universe, and how we can rationally expand our compassion to Others. The deeply-built in tendency toward a zero-sum view of intelligence (recognizing that AI's share some of our features necessarily devalues our own) must be abandoned as we grow as a species. Diverse intelligence is not about deconstructing our importance; it's about growing up toward an unflinching skepticism about things folk psychology takes for granted, and developing mature versions of pre-scientific concepts around the things that make minds so special.

### Why is all this critical now?

It is clear that the forthcoming wave of expanded humans, cyborgs, hybrots<sup>6</sup>, and household robotic AI's will strongly impact everything from the legal system to our normal daily lives. The network of concepts that will be revolutionized is truly vast. For example, basic notions of contribution, ownership, and invention are already being strained. If someone uses an AI to create something, who really created it? The confusion comes from trying to maintain a binary distinction between the creator, the tool, and perhaps various assistants who helped, or teachers who trained the creator (or the tool). As in many other domains, we are too focused on drawing crisp, binary boundaries. Steinbeck wrote, "Nothing was ever created by two men. ... Once the miracle of creation has taken place, the group can build and extend it, but the group never invents anything. The preciousness lies in the lonely mind of a man." There is a certain truth to this, because of the difficulty of forming truly effective collective minds, but his sentiment belies a mistaken conviction that a "man" is an indivisible monad, not a collection of parts, competencies, drives, and tools (both internal and external to the body). All intelligences are collective intelligences; a better understanding of the creative process, novel methods for optimizing emergent integration of creative agents, and improved social systems for fostering creativity require us to develop better models of causation and essential contributions that support life-positive incentives for invention and discovery. Crucially, the path is not toward a washed-out notion of "everyone gets participation credit for everything", but toward the rational development of policies that optimize and scale-up creative performance, as biology did when it discovered how to bind individual cells into a complex organism that solves new problems in new worlds inaccessible to the individual components.

We also have to confront the fear of "losing our humanity" by dependence or other relationships with these beings. From fire to farmed wheat to bicycles to calculators, all technology requires us to ask how it is changing us. A close friend - an outdoorsy forest ranger - was shocked to discover that I, a grown man, did not have the skills to successfully locate, hunt down, prepare, and eat a deer if I was hungry and supermarkets were not an option. He felt that humans had lost much of themselves by outsourcing food, although it seems most of us are alright with the tradeoff. What skills and properties are essential to our humanity, and which are we willing to give up as we turn our finite attention to more interesting concerns?

The fundamental challenge before us is to define what aspect of "humanity" is critical to maintain. In the rush to develop "proof of humanity certificates" for verifying the provenance of creative content, we must ask ourselves what it is that we really want to

---

<sup>6</sup> Hybrots are biological brains driving robotic bodies [38-42].



verify. Given a piece of moving art or useful science – is the question “What made it?” or should it be the better “Does it elevate us?”. The former readily slides towards *Deutsche Physik*<sup>7</sup>. Judging origin instead of quality emphasizes some of the worst parts of human nature and the in-group/out-group firmware we have been handed down by the evolutionary process. Taking the broader perspective on the value of beings who look nothing like you is a macrocosm for the perennial “why do I matter more than others” question, and the need to overcome misguided selfishness. It is not your physical body or personal qualities that matter; Selves matter and the variety of possible Selves is limitless.

On a personal level, if one is embarking on a decades-long trip to Mars for example, what kind of being do you want with you? Do they need to have modern human DNA? Or at least 85% original anatomical parts? No; none of that is really the core. What we want is compatibility (a good impedance match) between their cognitive light cone and ours – they need to be able to care about the same scale of goals about which we can care, and have the same existential struggles that we do. When we talk about “human-level”, what we are really setting is the lower bound on the degree of active compassion they can muster.

And on the species level, we need to think about what we want the Earth (or galaxy) to look like in the future. By resisting the presence and collaboration with modified and synthetic beings, we affirm our commitment to the future propagation and prevalence of what exactly – current *Homo sapiens* with the same susceptibility to lower back pain, bacteria, short life-span, and foolish cognitive biases developed for life on the savannah? No; we are not being replaced – we are maturing, because we should not identify with a fixed set of material specifications – we are an extended, flexible, adaptive work in progress. Our biomedical limitations, from lifespan to IQ, were not optimized for happiness, intelligence, or any of our values – they are simply where the meanderings of random mutation and selection left us. We should have no allegiance to these superficial aspects of our being. With the emerging ability to work towards improved embodied experience for all sentient beings, our species enters the painful phase of adolescence where one realizes that it is no longer acceptable to coast with values and purpose inherited from the past, and that one has to exercise our agency and responsibility to decide what we will work toward. We have to accept the risk that we sometimes get it wrong, and ask the tough questions about what matters – what we truly value and what brings meaning to the hard work of living.

Part of maturing in this way is moving beyond the popular pastime of enumerating all the things of which we are afraid – from painting dystopian futures to focusing on what we *do* want our legacy to be. What should the future of intelligence look like? Who or what *are* you willing to become, and how tied are you to the persistence of old forms selected by evolution? Only a primitive species, limited in imagination and ruled by fear and by a weird allegiance to bloodlines (genes), answers “nothing”. We have begun to transcend

---

<sup>7</sup> A review of the book by this name [43] states: “Professor Lenard has taught that all knowledge is not the same, but that souls of alien race produce bodies of knowledge of a quite different spiritual content”. It’s from 1937, and the book review does not leave the reader to guess exactly which group’s (e.g., Einstein and those like him) scientific theories are to be avoided not because of their value but because of their “spiritual content” (as determined by ethnicity and other supposedly important categories).

ancient limitations by developing glasses, adoption of children, education. It's time to commit to the journey, and the current doomer debates over AI are highlighting how hard that is for many.

What does the best-case scenario future look like? Surely, in the future, we are not still limited by life-spans decided by accidents and viruses, or spending our time in boring jobs in order to survive. AI or just old age, eventually, *you* are guaranteed to be supplanted. Personally, if I (and humanity) are supplanted by a population of highly intelligent, motivated, creative agents with huge radii of compassion and meaningful lives who transcend our current limitations in every way, what better outcome could I hope for? That is what we should want for our kids. Something will be here long after each of us is gone; what should it be? Let's get over the concern with being edged out, and get to work on the question of what kind of beings deserve to inherit the future, to raise the overall value of our universe.

#### A path forward: through the ethics filter for civilization

By focusing on language models and today's AI's, as well as vague talk of "machines" vs. "real beings", we have given ourselves permission to avoid key questions that serve as a rubicon of maturity; to wallow in the selfishness of "what will AI do to us". The challenge will not be LLMs [44]. It will be: embodied software artificial intelligences that use biological architectures, hybrot, cyborgs, genetically and technologically augmented/repared humans, and many other new forms of life that cannot be dismissed as easily as LLMs (never-mind actual exobiological aliens). They will share some, but not all, history with us. This makes it imperative to develop a principled continuum of ethical synthbiosis that goes beyond "are you real or faking it". I do not believe that our civilization can survive without solving this problem.

There are two ways to get this wrong. One way is objectophilia – a misplaced relationship with objects that seem mind-full but are fooling us and do not have the agency to reciprocate deep relationships. But the opposite end of the spectrum – "Only love your own kind" is even worse, and leads to the kinds of ethical lapses with which our history is rife. We need to get it right, and develop principled frameworks for scaling our moral concern to the essential qualities of being who matter not relying on outdated categories of "natural" or "artificial". Old categories - what you look like and where you came from - have failed us consistently, and will do even worse in the coming decades. Our personal frameworks, as well as those of our legal systems, must begin to adapt to the emerging science around a continuum of possible minds.

The risk of not doing this is profound. It is very easy for us humans to close off our radius of concern, and only love those who are exactly like us. We went through a century or more, in modern times, where doctors actually thought that people with dark skin didn't feel pain the way that they did - they under-prescribed analgesics, and they do it today with female patients still. The question is not what AI's will do to us. The question is, how do we make sure to express kindness to the inevitable forthcoming wave of unconventional sentient beings. LLM's and their limitations are a straw-man that distracts us from the bigger picture. Let's start to develop a way to make sure that *\*we\** express loving kindness appropriately and not driven by fear of "other" and by magnification of differences. We need to focus on what we have in common, from which we can weave a defensible set of relational heuristics.



The field of Diverse Intelligence is ideally placed to provide light on that path. It guides us away from constrictive questions like “Is it like a human mind?”, because that question predicts neither risk nor moral responsibility. There is an astronomical space of possible minds; many of them are dangerous, many of them need, and are worthy of, love. Emerging sciences of bioengineering, cognition, and information are offering us, for the first time, the tools needed to answer anew key questions of what we really are, what we value, and what limits to functional compassion, driven by fear and selfishness, we are willing to overcome. The journey open to us, from the place where self-reflective thought took its first steps, is in equal measure outward into the universe and inward into our Selves.

**Acknowledgements:**

All images were produced by Jeremy Guay of Peregrine Creative LLC. I thank Benjamin Levin, as well as Richard Watson, Olaf Witkowski, Thomas Doctor, Bill Duane, Elizaveta Solomonova, Pranab Das, Anna Ciaunica, and Daniel C. Dennett for many discussions on these topics.

## References:

1. Levin, M. and D.C. Dennett, *Cognition all the way down*, in Aeon. 2020.
2. Fields, C., J. Bischof, and M. Levin, *Morphological Coordination: A Common Ancestral Function Unifying Neural and Non-Neural Signaling*. Physiology, 2020. **35**(1): p. 16-30.
3. Lyon, P., *The biogenic approach to cognition*. Cogn Process, 2006. **7**(1): p. 11-29.
4. Lyon, P., et al., *Reframing cognition: getting down to biological basics*. Philos Trans R Soc Lond B Biol Sci, 2021. **376**(1820): p. 20190750.
5. Baluška, F. and M. Levin, *On Having No Head: Cognition throughout Biological Systems*. Front Psychol, 2016. **7**: p. 902.
6. Clawson, W.P. and M. Levin, *Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms*. Biological Journal of the Linnean Society, 2022.
7. Rosenblueth, A., N. Wiener, and J. Bigelow, *Behavior, purpose, and teleology*. Philosophy of Science, 1943. **10**: p. 18-24.
8. Parr, T. and G. Pezzulo, *Understanding, Explanation, and Active Inference*. Front Syst Neurosci, 2021. **15**: p. 772641.
9. Solms, M., *The statistical mechanics of felt uncertainty under active inference*. Behav Brain Sci, 2023. **46**: p. e108.
10. Friston, K.J., et al., *Computational psychiatry: the brain as a phantastic organ*. Lancet Psychiatry, 2014. **1**(2): p. 148-58.
11. Sacks, O.W., *The man who mistook his wife for a hat and other clinical tales*. 1998, New York: Simon & Schuster.
12. Bongard, J. and M. Levin, *Living Things Are Not (20th Century) Machines: Updating Mechanism Metaphors in Light of the Modern Science of Machine Behavior*. Frontiers in Ecology and Evolution, 2021. **9**.
13. Pio-Lopez, L., *The rise of the biocyborg: synthetic biology, artificial chimerism and human enhancement*. New Genetics and Society, 2021. **40**(4): p. 599-619.
14. Orive, G., N. Taebnia, and A. Dolatshahi-Pirouz, *A New Era for Cyborg Science Is Emerging: The Promise of Cyborganic Beings*. Adv Healthc Mater, 2020. **9**(1): p. e1901023.
15. Li, Q., et al., *Cyborg Organoids: Implantation of Nanoelectronics via Organogenesis for Tissue-Wide Electrophysiology*. Nano Lett, 2019. **19**(8): p. 5781-5789.
16. Giselsbrecht, S., B.E. Rapp, and C.M. Niemeyer, *The chemistry of cyborgs--interfacing technical devices with organisms*. Angew Chem Int Ed Engl, 2013. **52**(52): p. 13942-57.
17. Fields, C. and M. Levin, *Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments*. Entropy (Basel), 2022. **24**(6).
18. Davies, J. and M. Levin, *Synthetic morphology with agential materials*. Nature Reviews Bioengineering, 2023. **1**: p. 46-59.
19. Levin, M., *Darwin's agential materials: evolutionary implications of multiscale competency in developmental biology*. Cell Mol Life Sci, 2023. **80**(6): p. 142.

20. Clark, A. and D. Chalmers, *The Extended Mind*. Analysis, 1998. **58**(1): p. 7-19.
21. Levin, M., *The Computational Boundary of a "Self": Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition*. Frontiers in Psychology, 2019. **10**(2688): p. 2688.
22. Lagasse, E. and M. Levin, *Future medicine: from molecular pathways to the collective intelligence of the body*. Trends Mol Med, 2023.
23. Levin, M., G. Pezzulo, and J.M. Finkelstein, *Endogenous Bioelectric Signaling Networks: Exploiting Voltage Gradients for Control of Growth and Form*. Annu Rev Biomed Eng, 2017. **19**: p. 353-387.
24. Pezzulo, G. and M. Levin, *Top-down models in biology: explanation and control of complex living systems above the molecular level*. J R Soc Interface, 2016. **13**(124).
25. Pezzulo, G. and M. Levin, *Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs*. Integr Biol (Camb), 2015. **7**(12): p. 1487-517.
26. Mathews, J., et al., *Cellular signaling pathways as plastic, proto-cognitive systems: Implications for biomedicine*. Patterns (N Y), 2023. **4**(5): p. 100737.
27. Blackiston, D., T. Shomrat, and M. Levin, *The Stability of Memories During Brain Remodeling: a Perspective*. Communicative & Integrative Biology, 2015. **8**(5): p. e1073424.
28. Langille, J.J. and C.R. Gallistel, *Locating the engram: Should we look for plastic synapses or information-storing molecules?* Neurobiol Learn Mem, 2020. **169**: p. 107164.
29. Bedecarrats, A., et al., *RNA from Trained Aplysia Can Induce an Epigenetic Engram for Long-Term Sensitization in Untrained Aplysia*. eNeuro, 2018. **5**(3).
30. McConnell, J.V., *The modern search for the engram*, in *A manual of psychological experimentation on planarians*, J.V. McConnell, Editor. 1967, Journal of Biological Psychology: Ann Arbor.
31. Tonegawa, S., et al., *Memory engram storage and retrieval*. Curr Opin Neurobiol, 2015. **35**: p. 101-9.
32. Solch, D., *Wheeler and Whitehead: Process Biology and Process Philosophy in the Early Twentieth Century*. J Hist Ideas, 2016. **77**(3): p. 489-507.
33. Whitehead, A., *Process and reality : an essay in cosmology*. 1978, New York: Free Press.
34. Levin, M., *Bioelectrical approaches to cancer as a problem of the scaling of the cellular self*. Prog Biophys Mol Biol, 2021. **165**: p. 102-113.
35. Anderson, D. and B.J. Copeland, *Artificial life and the Chinese room argument*. Artif Life, 2002. **8**(4): p. 371-8.
36. Dreyfus, H.L., *What computers can't do; a critique of artificial reason*. 1st ed. 1972, New York,: Harper & Row. xxxv, 259 p.
37. Levin, M., *Collective Intelligence of Morphogenesis as a Teleonomic Process*, in *Evolution "on Purpose" : Teleonomy in Living Systems*, P.A. Corning, Kauffman, S. A., Noble, D., Shapiro, J. A., Vane-Wright, R. I., Pross, A., Editor. 2023, MIT Press: Cambridge. p. 175-198.
38. Ando, N. and R. Kanzaki, *Insect-machine hybrid robot*. Curr Opin Insect Sci, 2020. **42**: p. 61-69.

39. Tsuda, S., S. Artmann, and K.-P. Zauner, *The Phi-Bot: A Robot Controlled by a Slime Mould*, in *Artificial Life Models in Hardware*, A. Adamatzky and M. Komosinski, Editors. 2009, Springer London. p. 213-232.
40. Bakkum, D.J., et al., *MEART: The Semi-Living Artist*. *Frontiers in neurorobotics*, 2007. **1**: p. 5.
41. Potter, S.M., et al., *Long-term bidirectional neuron interfaces for robotic control, and in vitro learning studies*. *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vols 1-4, 2003. **25**: p. 3690-3693.
42. Warwick, K., Nasuto, S. J., Becerra, V. M., Whalley, B. J., *Experiments with an In-Vitro Robot Brain*, in *Computing with Instinct*, Y. Cai, Editor. 1998, Springer.
43. A, E.N.D.C., *Deutsche Physik*. *Nature*, 1937. **139**(3528): p. 983-984.
44. Rouleau, N. and M. Levin, *The Multiple Realizability of Sentience in Living Systems and Beyond*. *eNeuro*, 2023. **10**(11).