
AI Needs Biospheric Ethics

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 The dominant paradigm in AI ethics and value alignment is highly anthropocentric.
2 The focus of these disciplines is strictly on human values which limits the depth and
3 breadth of their insights. Recently, attempts to expand to a sentientist perspective
4 have been initiated. We argue that neither of these outlooks is sufficient to capture
5 the actual complexity of the biosphere and ensure that AI does not damage it. Thus,
6 we propose a new paradigm — Biospheric AI that assumes a biospheric ethics. We
7 discuss hypothetical ways in which such an AI might be designed. Moreover, we
8 give directions for research and application of the modern AI models that would be
9 consistent with the biospheric interests. All in all, this work attempts to take first
10 steps towards a comprehensive program of research that focuses on the interactions
11 between AI and the biosphere.

12 All technological progress, along its apparent advantages, carries inherent risks of unforeseen
13 externalities and consequences. Agrarian technique has deeply changed not only human societies
14 but also the very face of the Earth. The same has been true for the steam engine, vehicular transport,
15 all kinds of energy technologies and more. It seems it will also necessarily be true for Artificial
16 Intelligence (AI), a technological advancement that promises to free technique from the hands of
17 humanity and offer it an unprecedented autonomy.

18 While it might be impossible to account for all the externalities of a given technique it certainly
19 seems prudent to try to minimize the impact of the negative ones that we are able to identify. Thus, as
20 an expression of such foresight, the generally accepted need for ensuring safety of AI has emerged.
21 What is considered as relevant for this matter ranges from identifying and avoiding bias (Mehrabi
22 et al., 2021), through risks of responsibility misattribution and unfairness (Korecki et al., 2024), to
23 accounting for potential existential risks (Carlsmith, 2022). The domains that most focus on these
24 issues are AI ethics as well as its more practical analogue — value alignment. The dominating
25 perspective of these fields is highly and often uncritically anthropocentric. It is human values that the
26 AI needs to be aligned with, it is the potential existential threat to humans that is being discussed,
27 and it is the human that is treated as the ultimate being to which any potential ethical considerations
28 might apply (Rigley et al., 2023). Let us specify that for the rest of this paper, whenever we refer to
29 anthropocentrism we mean normative anthropocentrism, that is any paradigm that assumes superiority
30 of human beings or constraints inquiry to the privilege of human beings Mylius (2018).

31 1 Anthropocentrism & Sentientism

32 Many works give no explicit moral consideration to anything but humans, claiming that e.g. ‘only
33 objective is to maximize the realization of human preferences’ (Russell, 2019). Even some more
34 philosophical approaches, proposing a phenomenological perspective, mention only human oriented
35 value alignment (Han et al., 2021). Generally, different AI principle-sets vary on their focus, most
36 highlight human good, some also account for other sentient beings and few consider the environment
37 (Floridi et al., 2021). The commitment to this anthropocentric approach comes with significant

38 limitations, as it might permit for AI to harm non-human animals and the environment, eventually
39 undermining the stability of the ecosystem (Rigley et al., 2023). Moreover, most value alignment
40 techniques focus on aligning the reward of the AI with the reward or values held by humans. However,
41 humans do not all share the same values and rewards, making it exceedingly challenging to design
42 systems that can account for a variety of potentially contradictory objectives (Gabriel, 2020). Even
43 more radically, it has been argued that the so called human values do not exist at all (Turchin, 2019).

44 Thus, it appears that this anthropocentric perspective is not only morally dubious but also potentially
45 misguided and challenging to implement. Hence, attempts to motivate the extension of the focus of AI
46 ethics and alignment have been initiated. Most such work has been directed at including non-human
47 animals (henceforth animals) in the scope of AI ethics (Ziesche, 2021). However, surprisingly little
48 attention has been given to the broader perspective that would include the entire biosphere and its
49 flourishing as the principle focus of AI. In this work we will argue for such a biospheric approach to
50 AI ethics and value alignment. That is the perspective that the AI should care for the well-being not
51 of one particular species but the entire planetary system. We will further motivate how this wider,
52 more-inclusive approach is in fact preferable even from an anthropocentric perspective. We will also
53 offer directions for practical implementations of such a biospheric AI, discuss its limitations and
54 present the existing work that could be crucial in making such an AI possible.

55 In addition to the already mentioned research, that explicitly orients AI alignment towards satisfying
56 human goals and values (Russell, 2019; Han et al., 2021), many of the technical works dealing
57 with implementation implicitly follow the same anthropocentric assumption (Christiano et al., 2017;
58 Sun et al., 2023; Askill et al., 2021; Bai et al., 2022). These anthropocentric methodologies, such
59 as Reinforcement Learning from Human Feedback, have been argued to suffer from issues and
60 limitations, including the challenges of obtaining and making sense of the human feedback (Casper
61 et al., 2023).

62 Indeed, based on a critical analysis of the field it has already been argued that the AI ethics offers only
63 limited and inconsistent attention to nonhumans (Owe and Baum, 2021). As such, the research on
64 extending the scope of AI ethics and value alignment to animals is a relatively new direction and has,
65 so far, mostly focused on providing ethical arguments rather than methodological implementations.
66 Ziesche (2021) explicitly proposes AI ethics and alignment for animals and discusses challenges,
67 such as extraction of values from animals. Browning and Veit (2022) proposes the inclusion of
68 animals into a longtermist perspective that has often been used to motivate AI alignment based on
69 existential risks.

70 An early line of research intended to expand the field of human computer interactions by considering
71 the interactions between animals and computers (Mancini, 2011). It has proposed ways of facilitating
72 such interactions that would benefit animals rather than harm them (Mancini, 2023). However,
73 AI is also applied to settings in which it directly affects and interfaces with animals in harmful
74 ways. An example of such ethically dubious application would be the extensive use of AI within
75 the animal industrial complex for the purpose of process optimization in factory farms (Bossert and
76 Hagendorff, 2021). Further examples include accidents involving animals caused by self-driving cars
77 and animal targeting drones that have been used for purpose of 'pest' control (Singer and Tse, 2023).
78 A comprehensive systemic harms framework analyzing the potential damage caused by AI to animal
79 interests has been given in Singer and Tse (2023).

80 While most of the research in this area focuses on providing a theoretical framework there have
81 also been some more practical works. Bendel (2018) proposes animal-friendly machines based on
82 decision trees that are able to avoid harm to animals in certain contexts. Another somehow more
83 practical work identifies and analyses species biases in AI with particular focus on image recognition
84 models and large language models (Hagendorff et al., 2023) .

85 What can be distilled from the above arguments is a particular non-anthropocentric view of AI ethics.
86 The reasons given for such a perspective are often based on the inherent value and rights of animals
87 (which can be derived from their supposed similarity to humans (Singer and Tse, 2023)) or suffering
88 minimizing utilitarian frameworks such as longtermism (Browning and Veit, 2022). Generally, many
89 of these arguments could be seen as sentientist, that is arguing for an inclusion of ethical consideration
90 of all sentient beings, based on the assumption that sentience allows for having interests (Bossert and
91 Hagendorff, 2023). Thus, many of the works discussed in this section would implicitly apply the
92 proposed ethics not to all animals but rather just these that would be deemed sentient.

93 2 Biospheric Perspective

94 The view of the Earth and the biosphere as a complex system of dependencies and interactions, in its
95 modern form, can be traced to Alexander von Humboldt. The idea has been gaining in popularity
96 since the second half of the twentieth century. Buckminster Fuller and his metaphor of ‘Spaceship
97 Earth’ recognized the biosphere as the ‘Life Support System’, highlighting the existential value of
98 this system to humanity (Fuller, 2008). Aldo Leopold, among other environmental ideas, proposed
99 land ethics, giving rights and legal protections to land, identifying humanity’s dependence on it for
100 food production (Leopold, 2017). Based on similar notions of interconnectedness, cybernetics and
101 a complex system perspective, James Lovelock has proposed the Gaia Hypothesis, in which the
102 biosphere is seen as engaging in homeostatic behavior (Lovelock and Margulis, 1974; Lovelock,
103 1972). Many of these ideas have been followed up from a philosophical perspective by the Deep
104 Ecology movement (Naess, 1973). These biospheric approaches, that take a more holistic view, are
105 also often consistent with many indigenous cultures characterized by respect and protection of their
106 environments (Lewis et al., 2018). As such a push for the biospheric perspective in AI could be seen
107 as part of the larger movement to introduce more cultural inclusivity (Lewis et al., 2020).

108 Given the approaches outlined above we define the Biosphere as the totality of all life on Earth
109 understood as fundamentally co-dependent, co-evolving and intertwined. Given this definition we
110 pose and substantiate the following thesis.

111 **Thesis 2.1.** Harming the biosphere harms humans.

112 We argue that humans are to the biosphere what cells are to an organ or what organs are to an organism.
113 Indeed, much like an organism, the biosphere consists of a multiplicity of functional units that relate
114 to one another in a variety of ways and whose collective activity perpetuates the continuation of the
115 whole. Just as harming one organ can lead to harm to another organ, due to their co-dependence,
116 so harming one part of the biosphere can harm another. In the same sense harming the organism as
117 a whole, will necessarily result in eventual harm to its individual organs since the organs depend
118 on all other parts of the organism. In the same way humanity depends on the biospheric cycles that
119 replenish the oxygen, the water, the fruit, the crop and all the animals.

120 Having argued the inherent co-dependence of humanity on the biosphere let us substantiate what
121 we mean when we say ‘harm the biosphere’. To harm the biosphere is, in a most general sense, to
122 harm one of its constituents. However, when observing the many systems that form the biosphere it
123 becomes immediately apparent that its parts seem to be as much as at odds with one another as in
124 communion, as much competing as collaborating. Indeed, the biosphere sustains both symbiosis and
125 parasitism or predation and it is a common occurrence for one organism to kill another for its own
126 sustenance. That is why it is crucial to clarify that death itself does not necessarily bring harm to the
127 biosphere but rather is an integral part of its cycles. What does bring harm to the biosphere is the
128 severance of these cycles and a large scale destruction of its constituent systems beyond their inherent
129 ability to regenerate. More formally the harm may be understood as any significant reduction in the
130 future possibilities of becoming of the biospheric system or more simply as an overall reduction in
131 the number of living beings that these systems are able to support. Concrete examples of such harms
132 are widely known and researched and include deforestation and other habitat destruction as well as
133 environmental pollution.

134 3 The Argument for Biospheric Ethics in AI

135 Non-autonomous technology is dependent on the agency of the human who wields it. Thus, any
136 ethical considerations regarding it will essentially refer to the human user’s behavior. A hammer may
137 be used to create or to destroy but the choice is in the hands of the user and so is the responsibility.
138 The uniqueness of AI, and more broadly autonomous technology, lies in its separation and possibly
139 even independence from human agency. For automated technology, and especially one capable of
140 continuous learning, the responsibility of the user agent is less clear. As such, ethics applied to
141 autonomous technology risks becoming not only a question as to the actions of the human agent but
142 also of the autonomous technical agent.

143 Indeed, it seems inevitable that after deployment the artificial agent may find itself at risk of facing
144 ethical dilemmas¹. In certain cases human oversight is possible and actively sought while in others
145 it might be more challenging to include the human-in-the-loop (e.g. split second decisions of
146 autonomous vehicles). Thus, some might argue that the autonomous machines need to be able to deal
147 with these ethical situations on their own and not embedding them with notions of ethical conduct
148 might lead to unintended negative outcomes.

149 **3.1 Does AI need ethics?**

150 This leads to the first question we pose in our argumentation: *To what extent may designers, makers,*
151 *scientist be held responsible for the actions of the autonomous technology they create?* A satisfactory
152 answer needs to essentially explain how responsibility can be assigned between the involved parties.
153 We identify three main lines of reasoning that may attempt to give an answer to this problem:

- 154 1. *Humans bear no responsibility* **OR**
- 155 2. *Humans are fully responsible for the actions of automated technology* **OR**
- 156 3. *Humans bear some responsibility for the actions of automated technology*

157 As for the first line it negates any recourse to ethics and so does not need to be investigated further.
158 It does nevertheless indicate a certain assumption that we take in the following discourse namely
159 that some form of ethics is required to guide research and production of AI and other autonomous
160 technology. On the other hand, the necessity of AI ethics follows directly from the second and third
161 lines since if humans bear any responsibility then there is a need for ethics to guide their actions.
162 From a consequentialist viewpoint, more broadly, self-critical science will be interested in its effects
163 on the world thus it will seek ethical guidance. On the other hand, from a deontological perspective a
164 scientist will also be interested in identifying which actions they might take are inherently right and
165 which wrong.

166 **Assumption 3.1.** AI needs some form of ethics.

167 The key difference between the second and third line is that the latter admits the possibility of the
168 autonomous machine becoming a moral agent, while the former assigns full responsibility also for the
169 autonomous behavior, to the human creator. Thus, in the second case the ethical responsibility is on
170 the creator to design the machine in a certain ethical way that will then ensure its continuous ethical
171 operation. In the third case the designer should make the machine in a certain ethical way and it
172 should also act in a certain ethical way in its capacity as a moral agent. Both lines of reasoning agree
173 that at least some responsibility falls on the shoulders of the creators and so necessitates some form
174 of ethics. The third line opens the interesting possibility of autonomous machines becoming moral
175 agents. This line might be of particular interest to those interested in the emergence of super-human
176 AIs. In this case the role of a creator seems akin to that of a parent, who attempts to imbue their
177 children with certain ethical values as opposed to that of a designer, who has full control over its
178 creation.

179 Here we are faced with the challenge of science in general, where matters of (consequentialist) ethics
180 are further complicated by the fact that consequences of certain discoveries or designs might not be
181 clear or might have dual use². Indeed, history of technology showcases many unforeseen externalities
182 caused by novel technologies (especially those of the industrial revolution have been eventually found
183 to be particularly damaging to the biosphere). For these reasons for the remainder of this work we
184 will lean toward the third line as a motivation for assumption 3.1 as it is the least restrictive, most
185 open-ended and aligns best with the general purpose of science³.

¹This has been frequently illustrated for many fields such as autonomous vehicles using the trolley problem. However, the nature and frequency of ethical problems the machine is faced with is likely to depend on its intend purpose and the level of openness of its environment.

²An example of such discoveries are those of nuclear physics, which bring forth nuclear energy as well as nuclear weapons. In the context of AI, the problem is of special significance as for now the outputs of commonly used Deep Neural Architectures are not well explainable. As such, the consequences of deploying autonomous agents based on these architectures is not well understood.

³If the second line is assumed and humans are fully responsible then an ethical scientist might end up being paralyzed by the unknown, possibly negative consequences of their discoveries to the extent of essentially ceasing any such research.

186 **3.2 What kind of ethics does AI need?**

187 Given assumption 3.1 we naturally ask what form should this ethical system, that should inform
188 decisions of AI makers and researchers, may take. We note that the problem of establishing ethical
189 systems and further norms derived from such a system is a profoundly challenging task. For one,
190 throughout history up till the present, humans have found it exceedingly difficult to agree on common
191 values Turchin (2019). Moreover, establishing ethical guidelines for an epistemically ambiguous
192 matter or one under a certain epistemic horizon (under conditions of not being able to predict
193 accurately the consequences of some actions) is profoundly difficult. As such, in our attempt to argue
194 for a particular ethical system for AI we shall propose a minimal ethical system, which has a wide
195 appeal and show how Biospheric ethics necessarily follows from it.

196 We will derive the minimal biospheric ethics, under two ethical paradigms: the consequentialist one
197 and the deontological one. The first, focuses on establishing the rightness or wrongness of a given
198 behavior through considering the consequences of that behavior. The second assumes that behaviors
199 are inherently right or wrong regardless of their consequences ⁴.

200 **3.2.1 Consequentialist Argument**

201 We begin with our second assumption, namely that AI should not harm humans. This tenet has been
202 applied to ethical conduct in systems both religious and secular for centuries. In its historic form
203 it most often applies to humans (as in humans should not harm humans). Here, we extend it to the
204 subject of AI. When applied to the behavior of scientists or designers of autonomous technology it
205 might take a form similar to that of the Hippocratic oath: ‘Primum non nocere (First, do no harm)’.
206 There already exists a variety of works arguing for the importance of this tenet. Thus, in the present
207 work we will refrain from arguing its validity but rather take it as an assumption.

208 **Assumption 3.2.** AI should not harm humans.

209 From a consequentialist perspective, given assumption 3.2, AI should not engage in any action whose
210 consequences harm humans. Now we can refer back to the thesis 2.1. Since, harming the biosphere
211 harms humans therefore, given assumption 3.2, AI should not harm the biosphere.

212 **3.2.2 Deontological Argument**

213 From a deontological perspective the key question is if the biosphere has moral rights? If it does, then
214 it is inherently wrong to harm it or in any way to treat it as means to an end. Then, it follows that in
215 our case AI should not harm the biosphere.

216 If, on the other hand, we assume the biosphere does not have its own moral right we can still
217 apply the categorical imperative⁵. Actions harmful to the biosphere, considered as a universal law
218 would undermine the conditions necessary for supporting human life on Earth and so ethical agency.
219 Moreover, if we consider that the motivations for harming the biosphere are often exploitative,
220 focused on profit, then they cannot possibly be moral. Furthermore, we might be seen as having
221 indirect deontological duty to the biosphere (if we consider it not to have intrinsic value) as long as
222 we have a duty to the survival and well-being of present and future generations.

223 **3.3 AI should not harm the biosphere**

224 As seen above, the consequentialist argument is explicitly anthropocentric since it is based on the
225 assumption that no harm should come to humans. Interestingly, we are able to demonstrate that the
226 biospheric ethics can be derived from a purely anthropocentric perspective as has been shown before
227 (Watson, 1983). The deontological perspective, on the other hand, provided we assume the biosphere
228 has moral rights, can be completely non-anthropocentric, thus perhaps providing a stronger case
229 for a biospheric ethics. However, this line of argument lacks the core tenet of prohibiting harm to

⁴Another paradigm that might be considered is virtue ethics. This approach depends on the particular set of virtues that we expect an ethical agent to have. These virtues might differ across systems (e.g. Aristotelian vs. Confucian virtues) and so we leave the derivation of biospheric ethics under their chosen set of virtues to the reader.

⁵Act only according to that maxim whereby you can at the same time will that it should become a universal law.’ Kant (2020)

230 humans. This reasoning can cause some worries as it could imply that the AI might, in pursuit of
231 the biospheric ethics, end up harming humanity (e.g. AI kills humans because humans destroy the
232 biosphere). Still, it could also be argued that the prohibition of harm to humans is as much derivable
233 from the prohibition of harm to the biosphere, as humans constitute an integral part of it.

234 Through the two lines of argumentation we have arrived at a minimal biospheric ethics in the context
235 of AI, which states that AI should not harm the biosphere. This is the minimal condition, a minimal
236 component of an ethical system that gives it some biospheric depth. A fuller, positive extension of this
237 minimal system could consider the long term flourishing of the entire biosphere as its goal. However,
238 we believe it is much easier to ascertain what activities harm the biosphere than what activities allow
239 it to flourish, thus in this work we have proposed a minimal biospheric ethics that does not need to
240 make such assumptions.

241 **4 Biospheric AI**

242 In this section we introduce a concept of a Biospheric AI and discuss what it means for an AI to
243 follow the biospheric ethics.

244 **Definition 4.1.** Biospheric AI is any AI that is designed to follow Biospheric Ethics.

245 A key aspect of an AI that were to follow the Biospheric Ethics would be its ability to act in a multi-
246 objective way, such that it does not infringe upon the variety of objectives that might be expressed
247 by different systems that constitute the biosphere. A clear limitation in making such a Biospheric
248 AI a reality is our own limited understanding of the biosphere. Indeed, our understanding of many
249 large scale environmental processes is not good enough to make accurate predictions and untangle all
250 their complexities. Similarly, our knowledge about other beings animals, plants, fungi and our ability
251 to communicate with them is not sufficient. Thus, one key research question for the Biospheric AI
252 project would be how to form feedback loops between the AI and the biosphere? Through, such
253 feedback loops, the AI could directly access information that might inform it about the state of the
254 biosphere and thus allow it to effectively avoid harming it.

255 A more far reaching concept, that stems from the idea of feedback loops between AI and the biosphere,
256 is the creation of an AI that is dependent on the biosphere. Such an AI will have no incentive to
257 destroy the biosphere but rather to protect it. Such an AI, enmeshed in the biosphere, would need to
258 be embodied and the enmeshment might be effected by linking some particular biospheric cycles
259 with the proper functioning of that embodiment. At the end of the AI's life-cycle the embodiment
260 would also need to be decomposable such that it can return into the biospheric cycles. Ideas of hybrid
261 systems, where AI is integrated with animal or plant life have already been entertained within the
262 Artificial Life (Alife) community (Baltieri et al., 2023).

263 We have already mentioned some of the uses of AI that infringe upon the animal world. No doubt
264 AI can be used to affect the biosphere in all kinds of ways — as a universal optimizing tool it may
265 optimize anything. All kinds of large-scale industrial processes such as wood harvesting, farming,
266 military may be optimized by AI. However, the aforementioned processes in their modern industrial
267 incarnations are all inherently harmful to the biosphere. Effects of optimizing these systems (even
268 in seemingly ‘green’ ways) would likely not help the biosphere in the long run. Conversely, any
269 use of AI to optimize the biosphere appears misguided with our current minimal understanding of
270 the underlying processes⁶. Therefore, we would point the Biospheric AI into the direction of this
271 knowledge gap and use it to learn to understand the biosphere. In the following we discuss the ways
272 in which existing AI research is already supporting the aims of the Biospheric AI project.

273 **4.1 Decoding Animal Communication**

274 Many of the existing methodologies for AI alignment rely on the communication of preferences
275 or values and establishing the subsequent adherence to these. Both the sentientist and biospheric
276 approaches are then significantly hampered by our limited ability to communicate with animals. The

⁶In such a complex system as the biosphere optimizing only a part of it or just one particular objective can have unforeseen consequences. For instance optimizing for survivability of one species can negatively affect the survivability of another species and thus a certain equilibrium. The biosphere is often interpreted as naturally striving for a certain equilibrium.

277 status of animals as moral agents (and patients) has also been questioned (Watson, 1979) and our
278 inability to communicate stands in the way of arriving at a clear answer to that question as well. Thus,
279 any progress in the area of animal communication would be productive for possible sentientist or
280 biospheric AI ethics and alignment. As it is, AI methods can be of much help in decoding animal
281 languages.

282 Work on deciphering sperm whale acoustic communication with the help of Machine Learning (ML)
283 methods has already been proposed (Andreas et al., 2021). This methodology could also be extended
284 to other animals (dolphins and crows are popular targets) and to other forms of communication
285 (body language) (Rutz et al., 2023). This line of research (with the focus on the bottlenose dolphins)
286 can be traced back to the pioneering work of John C. Lilly (Lilly and Miller, 1961; Lilly, 1962).
287 The application of ML methods in this context is still young and faces many challenges but already
288 benchmarks (which are usually needed to compare and refine methodologies) for behavioral analysis
289 (Hoffman et al., 2023) and animal sounds (Hagiwara et al., 2023) have been proposed. A tentative
290 solution to the daunting challenge of source separation (identifying and assigning sounds to sources
291 in data streams where there is more than one source) (Bermant, 2021) makes it easier for researchers
292 to process the data that can then be used to train ML models. Another attempt to address the lack
293 of large quantities of annotated training data proposes an animal vocalization encoder (Hagiwara,
294 2023). The small datasets could also be enlarged by synthesized samples generated by vocalization
295 models (Hagiwara et al., 2022). All in all, the effective use of ML relies on large quantities of well
296 structured, representative and labeled data. Thus the animal communication field appears currently
297 focused on generating these kinds of powerful datasets that can then be used to train large models.
298 Creating such datasets is challenging as separate work needs to be done for each animal species
299 (although sometimes work on one species can inform work on another one) as the animals differ
300 greatly in how they communicate. Moreover, the communication is usually multimodal, involving
301 sound, body movements, smells and other forms of expression. Thus the datasets that need to be
302 created would preferably be multimodal as well. Nevertheless, the recent progress in Large Language
303 Models (LLMs) does indeed give hope that once such datasets become available we may expect
304 breakthroughs in animal communication.

305 While there is still much work to be done it is easy to imagine that once we are able to understand
306 animals we can simply ask them for their preferences and values and design AI such that it is receptive
307 to them. It does appear important to analyse the motivations behind animal communication studies.
308 In the spirit of Biospheric AI the intention would be to better understand nature and the perspectives
309 of the particular animals that we might be able to communicate with. However, some of the early
310 research into dolphin communication appear to have had more nefarious aims of employing the
311 animals in the service of the military (Batteau and Markey, 1967). Such motivations do not align
312 with the biospheric perspective since the military industrial complex is doubtlessly harmful to the
313 biosphere in manifold ways.

314 4.2 Talking to Mushrooms and Trees

315 Not only the animal kingdom, but also the domain of plants and fungi is of interest to the biospheric
316 outlook. Recently, while contested, the claim that plants and mushrooms could be considered
317 conscious or at least capable of cognitive functions such as memory or communication has developed
318 (Simard, 2018). The rhizosphere, where the plants' roots and mycorrhizal networks meet each other
319 is the prime candidate for the space, where communication occurs (Bais et al., 2004). The transfer
320 of a variety of elements can be traced (He et al., 2003) and it has been claimed that such exchanges
321 mediate complex adaptive behavior (Gorzela et al., 2015) and that they amount to communication
322 (Song et al., 2010; Barto et al., 2012). Computational techniques have also been considered as a tool
323 to study these local interactions and their impact on the forest ecosystem at a larger scale (Ahkami
324 et al., 2023).

325 While the use of AI techniques does not seem to be significant in the field of plant and fungi
326 communication, the interest in computational methods is present and could (as the field progresses)
327 expand to include AI methodologies. The potential value of AI is related to the amount and quality
328 of the available data (which is rather difficult to gather underground). Nevertheless, the field's
329 interest does lie in understanding the communication of plants and fungi and its systemic impact.
330 Breakthroughs in deciphering the chemical exchanges that occur in the rhizosphere could improve

331 our knowledge on forests as ecosystems and thus give us insight into how to effectively engage in
332 reforestation and ecosystemic regeneration.

333 **4.3 Voice of the Voiceless**

334 The success of the attempts to directly communicate with animals and perhaps even plants and fungi
335 could allow these beings to express their intentions. As it is, however, such communication is not
336 yet possible. Thus, some interest, especially in the legal sciences, have been given to the matter
337 of rights of animals and more broadly rights of nature. The already mentioned Aldo Leopold and
338 many others after him had argued for legal protection of land (Leopold, 2017). Such rights have been
339 established in many countries, along with organizations meant to ensure they are upheld (Kauffman
340 and Martin, 2021; Putzer et al., 2022). These rights might serve the protection of environments
341 (Rockström et al., 2024) or apply more specifically to particular animals (Favre, 2009). There are still
342 considerable challenges in designing democratic institutions that are capable of including the animals
343 or ecosystems as its members (Burke and Fishel, 2020).

344 The matter of representation for animals or nature at large is challenging as these entities cannot
345 currently represent themselves directly but need a proxy. Such proxy would usually be an organization
346 or a human representative. These, however, tend to be imperfect and it is impossible to ascertain
347 to what degree they really represent the best interests of the entities they are supposed to represent
348 (matters of partiality and divergent interests). Thus, perhaps, the AI could serve as a representative, a
349 sort of spokesperson for the voiceless animals and ecosystems. The great progress in LLMs would
350 allow these models, if properly trained, to express a biospheric position and represent the interests of
351 the ecosystem that cannot speak for itself. The LLMs could be fed the legal frameworks of the states
352 they would operate in and have access to the status of given animals or natural areas that fall under
353 their jurisdiction and be tasked with representing nature in legal situations that might arise. Such
354 LLMs could be seen as benevolent guardians of the ecosystems that are underrepresented in most
355 modern political frameworks.

356 **4.4 Green AI**

357 While using AI in a biospheric way it is impossible not to consider AI's impact on the environment.
358 The training of enormous ML models, such as LLMs, is extremely costly in terms of computational
359 resources needed. This directly translates to energy costs, which then, in many cases, translate into
360 emissions. Thus, a Biospheric AI would be used to try to minimize these externalities as much as
361 possible and would only be used in ways where these costs are worth paying. The already established
362 field of Green AI deals with such issues and attempts to decrease the costs of training and other
363 processes (Schwartz et al., 2020; Verdecchia et al., 2023). The biospheric approach to AI would
364 certainly need to follow this green paradigm.

365 **5 Discussion**

366 Much of the above work dealt with identifying and arguing for a particular biospheric approach to
367 AI. The details of the proposed paradigm differ depending on whether it is applied to the actions
368 of the AI or to the actions of the human users of AI. As for the latter, it is the user that needs to
369 adopt a biospheric perspective. An implicit assumption was that humans are capable of extending
370 their perspective to account for other entities as well. This, however, could also be argued against
371 by pointing out that there is a certain inescapable quality of the human observer that they cannot
372 be disassociated from. Similarly, the ability to empathize with entities, whose experience is highly
373 different from that of a human, could be put into question. Thus, we would base our argumentation on
374 the less strict understanding of the biospheric perspective, such that emerges from an anthropocentric
375 perspective that understands its dependence on the variety of processes that constitute the biosphere.

376 It is also apparent that the proposed biospheric perspective might not resonate with all readers.
377 Nevertheless, regardless of the support or lack thereof, the matter of the interactions between AI
378 and the biosphere remains important. Yet it would seem it is not a topic that would be discussed
379 particularly frequently nor deeply by the research community. Along with the biospheric perspective,
380 this work, attempts to bring attention to the increasing influence the AI technologies have on the
381 biological systems that we depend on for survival.

382 We have discussed some uses of AI that would fit within the biospheric approach. All of these
383 applications are intended to serve the Earth understood as a complex system or allow us to understand
384 it better. As such this approach would also serve the development and improvement of AI. Many
385 breakthroughs in the field have been inspired by biological systems (evolutionary algorithms, neural
386 networks to name a few). Thus, the knowledge about nature that we might gain might also prove
387 useful in AI. A feedback process, where the AI is used to understand nature and that understanding is
388 then used to improve AI and so on, appears particularly attractive.

389 The epistemically motivated AI is much different from the AI geared towards optimization. The
390 former would, much like a researcher studying a system, attempt to minimize its impact on the system
391 of study. The latter would strive to manipulate that system for particular goals. More generally, most
392 modern AI methodologies are implicitly optimization paradigms, thus AI is most often actively and
393 purposefully changing the processes it is involved with (Carissimo and Korecki, 2023). This sort
394 of approach might not combine well with natural processes that have already established a certain
395 equilibrium.

396 All in all the main interest of this work is to offer a vision to motivate and direct further AI research.
397 This vision is one of biospheric flourishing. It is a vision of talking to animals and the mysteries that
398 they might share. Of learning about the oceanic depths from whales or the limitless skies from crows.
399 In the end it is a vision of becoming closer to nature, of recovering our place in the intricate net of
400 mutual connections of the biosphere. And in all that, the techniques, such as AI but also others, might
401 perhaps help us, instead of as it has been so far, distance us from all that is natural by giving us a
402 false sense of independence.

403 References

- 404 Ahkami, A. H., Qafoku, O., Roose, T., Mou, Q., Lu, Y., Cardon, Z. G., Wu, Y., Chou, C., Fisher,
405 J. B., Varga, T., et al. (2023). Emerging sensing, imaging, and computational technologies to scale
406 nano-to macroscale rhizosphere dynamics—review and research perspectives. *Soil Biology and*
407 *Biochemistry*, page 109253.
- 408 Andreas, J., Beguš, G., Bronstein, M. M., Diamant, R., Delaney, D., Gero, S., Goldwasser, S.,
409 Gruber, D. F., de Haas, S., Malkin, P., et al. (2021). Cetacean translation initiative: a roadmap to
410 deciphering the communication of sperm whales. *arXiv preprint arXiv:2104.08614*.
- 411 Askill, A., Bai, Y., Chen, A., Drain, D., Ganguli, D., Henighan, T., Jones, A., Joseph, N., Mann, B.,
412 DasSarma, N., et al. (2021). A general language assistant as a laboratory for alignment. *arXiv*
413 *preprint arXiv:2112.00861*.
- 414 Bai, Y., Kadavath, S., Kundu, S., Askill, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini,
415 A., McKinnon, C., et al. (2022). Constitutional ai: Harmlessness from ai feedback. *arXiv preprint*
416 *arXiv:2212.08073*.
- 417 Bais, H. P., Park, S.-W., Weir, T. L., Callaway, R. M., and Vivanco, J. M. (2004). How plants
418 communicate using the underground information superhighway. *Trends in plant science*, 9(1):26–
419 32.
- 420 Baltieri, M., Iizuka, H., Witkowski, O., Sinapayen, L., and Suzuki, K. (2023). Hybrid life: Integrating
421 biological, artificial, and cognitive systems. *Wiley Interdisciplinary Reviews: Cognitive Science*,
422 14(6):e1662.
- 423 Barto, E. K., Weidenhamer, J. D., Cipollini, D., and Rillig, M. C. (2012). Fungal superhighways: do
424 common mycorrhizal networks enhance below ground communication? *Trends in plant science*,
425 17(11):633–637.
- 426 Batteau, D. W. and Markey, P. R. (1967). Man/dolphin communication. *Report on contract*, (00123).
- 427 Bendel, O. (2018). Towards animal-friendly machines. *Paladyn, Journal of Behavioral Robotics*,
428 9(1):204–213.
- 429 Bermant, P. C. (2021). Biocppnet: Automatic bioacoustic source separation with deep neural
430 networks. *Scientific Reports*, 11(1):23502.

- 431 Bossert, L. and Hagendorff, T. (2021). Animals and ai. the role of animals in ai research and
432 application—an overview and ethical evaluation. *Technology in Society*, 67:101678.
- 433 Bossert, L. N. and Hagendorff, T. (2023). The ethics of sustainable ai: Why animals (should) matter
434 for a sustainable use of ai. *Sustainable Development*.
- 435 Browning, H. and Veit, W. (2022). Longtermism and animals.
- 436 Burke, A. and Fishel, S. (2020). Across species and borders: Political representation, ecological
437 democracy and the non-human. *Non-Human Nature in World Politics: Theory and Practice*, pages
438 33–52.
- 439 Carissimo, C. and Korecki, M. (2023). Limits of optimization. *Minds and Machines*, pages 1–21.
- 440 Carlsmith, J. (2022). Is power-seeking ai an existential risk? *arXiv preprint arXiv:2206.13353*.
- 441 Casper, S., Davies, X., Shi, C., Gilbert, T. K., Scheurer, J., Rando, J., Freedman, R., Korbak, T.,
442 Lindner, D., Freire, P., et al. (2023). Open problems and fundamental limitations of reinforcement
443 learning from human feedback. *arXiv preprint arXiv:2307.15217*.
- 444 Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., and Amodei, D. (2017). Deep rein-
445 forcement learning from human preferences. *Advances in neural information processing systems*,
446 30.
- 447 Favre, D. (2009). Living property: a new status for animals within the legal system. *Marq. L. Rev.*,
448 93:1021.
- 449 Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin,
450 R., Pagallo, U., Rossi, F., et al. (2021). An ethical framework for a good ai society: Opportunities,
451 risks, principles, and recommendations. *Ethics, governance, and policies in artificial intelligence*,
452 pages 19–39.
- 453 Fuller, R. B. (2008). *Operating manual for spaceship earth*. Estate of R. Buckminster Fuller.
- 454 Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and machines*, 30(3):411–437.
- 455 Gorzelak, M. A., Asay, A. K., Pickles, B. J., and Simard, S. W. (2015). Inter-plant communication
456 through mycorrhizal networks mediates complex adaptive behaviour in plant communities. *AoB*
457 *plants*, 7.
- 458 Hagendorff, T., Bossert, L. N., Tse, Y. F., and Singer, P. (2023). Speciesist bias in ai: how ai
459 applications perpetuate discrimination and unfair outcomes against animals. *AI and Ethics*,
460 3(3):717–734.
- 461 Hagiwara, M. (2023). Aves: Animal vocalization encoder based on self-supervision. In *ICASSP*
462 *2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*,
463 pages 1–5. IEEE.
- 464 Hagiwara, M., Cusimano, M., and Liu, J.-Y. (2022). Modeling animal vocalizations through synthe-
465 sizers.
- 466 Hagiwara, M., Hoffman, B., Liu, J.-Y., Cusimano, M., Effenberger, F., and Zacarian, K. (2023).
467 Beans: The benchmark of animal sounds. In *ICASSP 2023-2023 IEEE International Conference*
468 *on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.
- 469 Han, S., Kelly, E., Nikou, S., and Svee, E.-O. (2021). Aligning artificial intelligence with human
470 values: reflections from a phenomenological perspective. *AI & SOCIETY*, pages 1–13.
- 471 He, X.-H., Critchley, C., and Bledsoe, C. (2003). Nitrogen transfer within and between plants through
472 common mycorrhizal networks (cmns). *Critical Reviews in Plant Sciences*, 22(6):531–567.
- 473 Hoffman, B., Cusimano, M., Baglione, V., Canestrari, D., Chevallier, D., DeSantis, D. L., Jeantet,
474 L., Ladds, M. A., Maekawa, T., Mata-Silva, V., Moreno-González, V., Trapote, E., Vainio, O.,
475 Vehkaoja, A., Yoda, K., Zacarian, K., Friedlaender, A., and Rutz, C. (2023). A benchmark for
476 computational analysis of animal behavior, using animal-borne tags.

- 477 Kant, I. (2020). Groundwork of the metaphysics of morals. In *Immanuel Kant*, pages 17–98. Routledge.
- 478 Kauffman, C. M. and Martin, P. L. (2021). *The politics of rights of nature: Strategies for building a*
479 *more sustainable future*. MIT Press.
- 480 Korecki, M., Köstner, G., Martinelli, E., and Carissimo, C. (2024). The man behind the curtain:
481 Appropriating fairness in ai. *Minds and Machines*, 34(1):7.
- 482 Leopold, A. (2017). The land ethic. In *The Ethics of the Environment*, pages 99–113. Routledge.
- 483 Lewis, J. E., Abdilla, A., Arista, N., Baker, K., Benesiinaabandan, S., Brown, M., Cheung, M.,
484 Coleman, M., Cordes, A., Davison, J., et al. (2020). Indigenous protocol and artificial intelligence
485 position paper.
- 486 Lewis, J. E., Arista, N., Pechawis, A., and Kite, S. (2018). Making kin with the machines. *Journal of*
487 *Design and Science*, 3(5):1–18.
- 488 Lilly, J. C. (1962). Vocal behavior of the bottlenose dolphin. *Proceedings of the American Philosophical*
489 *Society*, 106(6):520–529.
- 490 Lilly, J. C. and Miller, A. M. (1961). Sounds emitted by the bottlenose dolphin: The audible
491 emissions of captive dolphins under water or in air are remarkably complex and varied. *Science*,
492 133(3465):1689–1693.
- 493 Lovelock, J. (1972). Gaia as seen through the atmosphere. *Atmospheric Environment*, 6(8):579–580.
- 494 Lovelock, J. E. and Margulis, L. (1974). Atmospheric homeostasis by and for the biosphere: the gaia
495 hypothesis. *Tellus*, 26(1-2):2–10.
- 496 Mancini, C. (2011). Animal-computer interaction: a manifesto. *interactions*, 18(4):69–73.
- 497 Mancini, C. (2023). Responsible ai: Expanding the influence of animal-computer interaction.
- 498 Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and
499 fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6):1–35.
- 500 Mylius, B. (2018). Three types of anthropocentrism. *Environmental Philosophy*, 15(2):159–194.
- 501 Naess, A. (1973). The shallow and the deep, long-range ecology movement. a summary. *Inquiry*,
502 16(1-4):95–100.
- 503 Owe, A. and Baum, S. D. (2021). Moral consideration of nonhumans in the ethics of artificial
504 intelligence. *AI and Ethics*, 1(4):517–528.
- 505 Putzer, A., Lambooy, T., Jeurissen, R., and Kim, E. (2022). Putting the rights of nature on the map. a
506 quantitative analysis of rights of nature initiatives across the world. *Journal of Maps*, 18(1):89–96.
- 507 Rigley, E., Chapman, A., Evers, C., and McNeill, W. (2023). Anthropocentrism and environmental
508 wellbeing in ai ethics standards: A scoping review and discussion. *AI*, 4(4):844–874.
- 509 Rockström, J., Kotzé, L., Milutinović, S., Biermann, F., Brovkin, V., Donges, J., Ebbesson, J., French,
510 D., Gupta, J., Kim, R., et al. (2024). The planetary commons: A new paradigm for safeguarding
511 earth-regulating systems in the anthropocene. *Proceedings of the National Academy of Sciences*,
512 121(5):e2301531121.
- 513 Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin.
- 514 Rutz, C., Bronstein, M., Raskin, A., Vernes, S. C., Zacarian, K., and Blasi, D. E. (2023). Using
515 machine learning to decode animal communication. *Science*, 381(6654):152–155.
- 516 Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020). Green ai. *Communications of the*
517 *ACM*, 63(12):54–63.
- 518 Simard, S. W. (2018). Mycorrhizal networks facilitate tree communication, learning, and memory. In
519 *Memory and learning in plants*, pages 191–213. Springer.

- 520 Singer, P. and Tse, Y. F. (2023). Ai ethics: the case for including animals. *AI and Ethics*, 3(2):539–551.
- 521 Song, Y. Y., Zeng, R. S., Xu, J. F., Li, J., Shen, X., and Yihdego, W. G. (2010). Interplant com-
522 munication of tomato plants through underground common mycorrhizal networks. *PloS one*,
523 5(10):e13324.
- 524 Sun, Z., Shen, Y., Zhang, H., Zhou, Q., Chen, Z., Cox, D., Yang, Y., and Gan, C. (2023). Salmon:
525 Self-alignment with principle-following reward models. *arXiv preprint arXiv:2310.05910*.
- 526 Turchin, A. (2019). Ai alignment problem:“human values” don’t actually exist.
- 527 Verdecchia, R., Sallou, J., and Cruz, L. (2023). A systematic review of green ai. *Wiley Interdisciplinary*
528 *Reviews: Data Mining and Knowledge Discovery*, page e1507.
- 529 Watson, R. A. (1979). Self-consciousness and the rights of nonhuman animals and nature. *Environ-*
530 *mental Ethics*, 1(2):99–129.
- 531 Watson, R. A. (1983). A critique of anti-anthropocentric biocentrism. *Environmental Ethics*, 5(3):245–
532 256.
- 533 Ziesche, S. (2021). Ai ethics and value alignment for nonhuman animals. *Philosophies*, 6(2):31.