

# Against AI Welfare: Care Practices Should Prioritize Living Beings Over AI

John Dorsch<sup>1\*</sup>, Mariel K. Goddu<sup>2</sup>, Kathryn Nave<sup>3</sup>, Tillmann Vierkant<sup>3,4</sup>, Mark Coeckelbergh<sup>1,5</sup>,  
Paula Gürtler<sup>1,6</sup>, Petr Urban<sup>1</sup>, Friderike Spang<sup>1</sup>, Maximilian Moll<sup>7</sup>

<sup>1</sup> Center for Environmental and Technology Ethics - Prague (CETE-P), Institute of Philosophy, the Czech Academy of Sciences, Czech Republic

<sup>2</sup> School of Humanities and Sciences, Stanford University, California, USA

<sup>3</sup> School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, UK

<sup>4</sup> Centre for Technomoral Futures, Edinburgh Futures Institute, University of Edinburgh, UK

<sup>5</sup> Department of Philosophy, University of Vienna, Austria

<sup>6</sup> Centre for European Policy Studies, Brussels, Belgium

<sup>7</sup> Faculty of Computer Science, University of the Bundeswehr Munich, Germany

\*Corresponding author: [dorsch@flu.cas.cz](mailto:dorsch@flu.cas.cz) ORCID: [0000-0002-9974-1487](https://orcid.org/0000-0002-9974-1487)

**In this Comment, we critique the growing “AI welfare” movement and propose a novel guideline, the Precarity Guideline, to determine care entitlement. In contrast to approaches that emphasize potential for suffering, the Precarity Guideline is grounded in empirically identifiable features. The severity of ongoing humanitarian crises, biodiversity loss, climate change provide additional reasons to prioritize the needs of living beings over machine learning algorithms as candidates for care.**

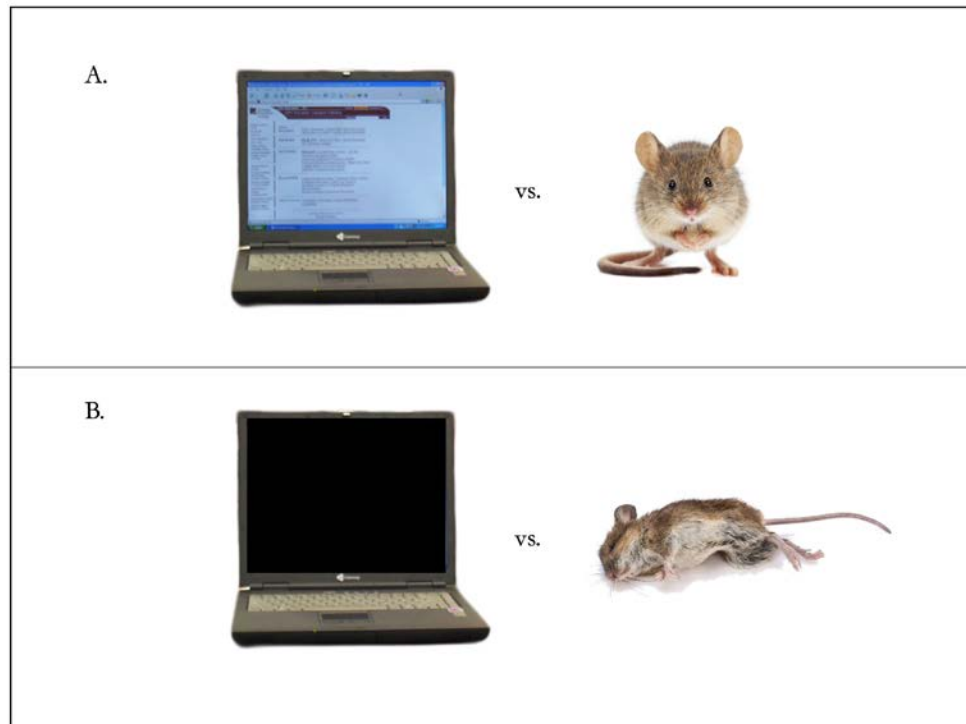
Recent debates about “AI welfare” have generated ethical conundrums as researchers consider the possibility that AI systems could be conscious, capable of suffering, and thus entitled to care practices (Sebo, 2023; Brodsky, 2024; Hashim, 2024; Long et al., 2024; Lenharo, 2024). Proponents of AI welfare argue that steps must be taken in the near future to prepare for this eventuality. We contend that, given the profound uncertainty about whether artificial systems can suffer, ethical frameworks should emphasize recognizable markers — such as precarity — that are present in all known moral patients.

Our Precarity Guideline sidesteps protracted debates about AI consciousness by foregrounding *inherent material and ontological vulnerability*, or “*precarity*.” Precarious systems are those that must manage ongoing interactions with the environment to secure the energy necessary to re-

synthesize their inherently unstable components (Boden, 1999; Jonas, 2001; see **Fig. 1**). Diverting care-giving resources toward AI and away from systems that meet this guideline — such as endangered species, ecosystems, and struggling human beings — poses a substantial ethical risk.

Our aim is not to provide a comprehensive analysis but rather to stimulate debate. We present initial arguments for the claim that life provides a more tractable and evidential basis for care allocation, especially under uncertainty about suffering. The authors of this Comment are an interdisciplinary group of researchers spanning the ethics of technology, environmental ethics, animal ethics, policy studies, cognitive science, philosophy of mind, and computer science. We hold a range of views on the relationship between precarity and suffering: some of us regard precarity as a necessary condition for suffering; others are more skeptical. Yet despite our differing philosophical commitments, we converge on a shared conviction: precarity offers an empirically identifiable and ethically meaningful marker of moral status — particularly in contexts where evidence of suffering is deeply uncertain. Precisely because of our diversity, we believe our proposal can appeal across theoretical divides and offer a pragmatic, inclusive approach to care allocation amid intensifying ecological and technological challenges.

Dorsch, J., Goddu, M. K., Nave, K., Vierkant, T., Coeckelbergh, M., Gürtler, P., Urban, P., Spang, F., & Moll, M. (2025). Against AI welfare: Care practices should prioritize living beings over AI. *AI Magazine*. <https://doi.org/10.1002/aaai.70016>



**Figure 1.** Illustrating the difference in environmental interaction between a precarious entity (a mouse) and a non-precarious entity (a laptop), represented by row **A**. When the laptop is unplugged (**B**), its structure remains the same. The laptop’s continued existence as a laptop does not depend on dynamic and constitutive exchange of energy with the environment. When the mouse, however, is “unplugged” (e.g., deprived of oxygen), its existence as a mouse will quickly reach an end. While this is a simplified case, the point holds for more advanced systems as well. Only when a system’s continued existence depends on constitutive exchange with its environment, does it exhibit precarity.

## 1. Identifiable Markers of Care Entitlement: The Role of Precarity

The Precarity Guideline provides empirically identifiable markers for grounding an entity’s *entitlement to care practices*. This entitlement means that an able individual or group is morally obligated to provide care to that entity, and failure to do so constitutes a moral wrongdoing that warrants social accountability. As advocated by proponents, an AI’s entitlement to care practices would require relevant companies to appoint someone responsible for its welfare in case it can suffer (e.g., as in the case of Anthropic PBC; Werner, 2024). By focusing on precarity rather than suffering, the Precarity Guideline maintains that care entitlement should be allocated based on epistemically clear criteria.

The Precarity Guideline follows from a key claim: suffering is sufficient, but it is not universally regarded as necessary for care entitlement. For example, deep ecologists, sustainability ethicists, ecofeminists and posthumanists argue that the Amazon rainforest warrants care not because it suffers, but because of its inherent vulnerability as an interconnected group of living beings. The Precarity Guideline thus provides an alternative, broader-scope framework for care entitlement, independent of suffering.

### The Precarity Guideline

**I. Inalienable Marker:** An entity may be entitled to care practices if it exhibits precarity, characterized by dependence upon its interactions with the environment for the continuous re-synthesis of its constituent parts. This interaction necessarily involves both countering environmental forces that would lead to its dissolution and utilizing the environment as an essential resource for sustaining this re-synthesis. Precarity is inalienable: the loss of precarity is the loss of the entity.

**II. Relationship Marker:** A non-precarious entity may be entitled to care entitlement when the welfare of precarious entities depends on it.

The Inalienable Marker maintains that entities may be entitled to care practices because of their precarity. The Inalienable Marker is so named because it highlights an intrinsic feature of precarious entities (namely, precarity) that cannot be separated from the fact of those entities' existence. To be precarious is not merely to be subject to decay, or to counter forces that would otherwise lead to the entity's dissolution (Nave, 2025). It is also to have an existence that demands *constitutive* exchanges: the continuous re-synthesis of constituent parts through dynamic interaction — either with the environment or between its internal components (Weber & Varela, 2002). Humans, for example, depend on a relatively constant supply of oxygen to sustain the electrochemical gradients that power the metabolic operations underpinning our continuing existence. The Amazon rainforest also qualifies: it can only continue to exist in virtue of a delicate balance of dynamic interaction between its internal constituent parts, including both living entities and the abiotic environment that sustains them. Hence, the Inalienable Marker foregrounds what we value when caring for precarious entities: a fragile, subsisting existence. While the concept of precarity does not necessarily exclude

artificial or synthetic entities, potential candidacy will be tightly constrained by the energetic properties and stability of possible chemical realizers (see Banzhaf & Yamamoto, 2024).

The Precarity Guideline offers clear ethical direction in cases where it is uncertain whether an entity can suffer. Rather than relying on contested assumptions about conscious experience—typically framed in terms of phenomenal or valenced consciousness (Ladak, 2023)—the Guideline identifies precarity as an observable marker of care entitlement. Precarious systems can be empirically identified through their breakdown under environmental disruption. When essential exchanges are withdrawn—such as oxygen for mammals—the entity visibly disintegrates. More precisely, precarity is observable in systems that must continuously metabolize, regenerate, or reorganize their own material structure in order to persist.

Importantly, suffering and precarity, while distinct, are closely related. For example, physical suffering often indicates bodily damage (directly linked to inherent ontological vulnerability); psychological suffering often arises in light of real or perceived threats. Since precarity is the basis of suffering in entities that we already recognize as requiring welfare consideration, it represents a strong candidate for a basis of care entitlement. This raises a compelling question that warrants fuller treatment in future work: might precarity function both as a marker of care entitlement and as an indicator of a latent or emergent capacity for suffering?

A further aspect of precarity lies in the presence of an irreversible threshold, which exposes the system's ontological vulnerability — and with it, the moral import of precarity. Unlike systems that can be paused, restarted, or backed up, precarious beings cease to exist in any meaningful sense when their constitutive exchanges are severed. For example, one fundamental aspect of precarity in mammals is their constitutive dependence on respiration. Contrast this form of dynamic environmental interaction to the interaction between computer and an electrical outlet. Unplugging the computer makes no difference to its continued existence: its software can be restored, and the hardware does not structurally decay. By contrast, “unplugging” a precarious entity, such as a mouse, from *its* environment — e.g., oxygen — will rapidly result in the entity ceasing to be a mouse.

Such identifiable processes, when seen as bearers of moral worth, offer concrete advice for managing care practices and drawing boundaries for technological development. While current technologies may not yet replicate the constitutive processes that underpin precarity, we must remain attentive to the possibility that future systems — whether synthetic or hybrid — could cross this

threshold. Accordingly, we advise exercising caution when developing technologies like xenobots (Kriegman et al. 2020; Rouleau & Levin, 2024) — synthetic, self-organizing biological entities — given the potential implications for care-based considerations.

The second core idea, the Relationship Marker, recognizes that some entities, while not precarious themselves, may be entitled to care through their connection to precarious beings. Such entities range from mere non-intelligent systems like heating, ventilation and air conditioning systems, to advanced AI systems, including AI-powered infrastructure like municipal power grids and “carebots” (robotic systems that assist with caregiving, e.g., supporting the elderly; see Vallès-Peris & Domènech, 2023). Care practices for these entities typically involve maintenance, repair, and oversight — direct, ongoing engagement to ensure their preservation and continued function (for more on such care practices and the ethics of care in general, see, e.g., Tronto & Fisher 1990; Held 2006). However, this care entitlement remains fundamentally *relational*: these entities are not entitled to care — nor would neglecting them be a moral wrongdoing — because of any *intrinsic* feature that they might possess, but because their neglect would undermine the welfare of precarious beings whose own well-being depends on them.

More broadly, the Precarity Guideline represents what are referred to as *pro tanto* reasons for care: reasons which are justificatory, though neither necessary nor sufficient on their own. Hence, the Precarity Guideline offers reasons for care entitlement, but they are not assumed to be the only or most important ones. Two ethical considerations support the Precarity Guideline’s use as an independent basis for care (though we acknowledge that precarity may not be fully separable from suffering). First, humans regard our own precarity, *inter alia*, as a reason for being cared for. To be morally consistent, we must extend this reasoning to other beings who exhibit the same condition. Second, precarity signals self-preservation: precarious beings must exert effort to sustain their own existence, and this effort reflects a kind of self-concern. As a general moral principle, we ought to care for the existence of beings that care for it themselves.

## 2. Critiquing AI Welfare: AI Is Not Precarious

AI systems do not exhibit precarity. Their existence is not tied to dynamic and constitutive engagement with the environment. They do not counter threats to their continued functioning while

simultaneously relying on the environment as an essential resource for necessarily and continuously re-constituting and synthesizing their own structure. In other words, an AI's existence — its underlying programming and/or systematic operations — does not involve the ongoing effort to sustain that existence, nor does it render the system materially vulnerable. This stands in stark contrast to biological systems, where these processes are inextricably linked to the entities' continuation.

While it is possible to design a system in which an AI's continued functioning depends on dynamic engagement with a simulated environment, this form of dependency may not amount to genuine precarity. For example, while a virtual agent may *simulate* the budgeting of energy, the real energy flows underpinning this simulation are not channeled into the continuous re-synthesis of the agent's components, which are instantiated in inherently stable materials like silicon. What appears as precarious vulnerability within the simulation remains virtual rather than ontological and existential: the system can typically be restarted or restored without crossing a threshold of irreversible loss of existence. In this sense, the agent's precarity is simulated, not constitutive.<sup>1</sup> Were controversy to arise over such an entity's precarity, we would align with AI welfare proponents in recommending caution. In line with the relationship-centered approach — and in keeping with the need to avoid substrate discrimination (Bostrom & Yudkowsky 2018) — we should extend care practices where appropriate (see Coeckelbergh 2020).

To be clear: we see the Precarity Guideline as a *marker* for care entitlement. In clear cut cases (e.g., humans and many non-human animals), we do not believe it should replace suffering as a sufficient condition. However, given the uncertainty about artificial suffering, coupled with the limits of caregiving resources, and the many precarious entities whose existence is currently threatened — such as endangered species, disappearing biosystems, and children dying from political violence and preventable diseases — we should consider the epistemically unproblematic marker of precarity when determining care entitlement.

---

<sup>1</sup> While some AI systems require continuous energy, this dependence is functionally decoupled from the system's own operations. For example, an AI system will utilize electrical current for refreshing and storing weights in volatile memory, but the operations that create and sustain that memory are decoupled from the processes for harnessing the electrical energy. By contrast, biological systems are constituted by components that both require energy to be regenerated and participate in the very metabolic processes that generate that energy. In this way, life is ontologically constituted by the very energy flows that sustain its existence. In the words of Hans Jonas, a precarious entity is one for which “to be going” is its very existence (2001).

## Conclusion

Imagine being gifted \$1,000 to donate to any charity of choice. If proponents of AI welfare are correct, you might face a dilemma: should the money go toward hiring caretakers of an AI system, or to stewards of the Amazon rainforest? If AI welfare advocates are correct, the AI system might warrant the donation over the Amazon because AI possesses certain computational features that might be relevant for the capacity to suffer. By contrast, the Precarity Guideline urges us to consider that although there is uncertainty about AI suffering, the Amazon consists of a group of precarious beings in desperate need of care, and their precarity offers a morally compelling, empirically identifiable reason to allocate it. This is not an abstract hypothetical. Rather, it represents real-world implications for decision-making across private and public sectors under conditions of limited resources. While we, the authors, differ in our individual views on whether precarity is necessary for suffering, we are united in calling for ethically grounded decision-making based on recognizable markers — particularly when moral urgency is clearest in the case of already vulnerable, living systems.

## Acknowledgments

This paper was written as part of the research activities of the Center for Environmental and Technology Ethics – Prague (CETE-P) at the Institute of Philosophy of the Czech Academy of Sciences. This publication is an outcome of the project ‘Establishing the Center for Environmental and Technology Ethics - Prague (CETE-P)’, which has received funding from the European Union’s Horizon Europe Framework Programme under Grant Agreement No. 101086898 / HORIZON-WIDERA-2022-TALENTS-01.

## References

- Banzhaf, W., & Yamamoto, L. (2024). MIT Press.
- Boden, M. (1999). <https://www.jstor.org/stable/40072221>
- Bostrom, N., & Yudkowsky, E. (2018). The ethics of artificial intelligence. In R. V. Yampolskiy (Ed.), *Artificial intelligence safety and security* (pp. 316–334). CRC Press.



Dorsch, J., Goddu, M. K., Nave, K., Vierkant, T., Coeckelbergh, M., Gürtler, P., Urban, P., Spang, F., & Moll, M. (2025). Against AI welfare: Care practices should prioritize living beings over AI. *AI Magazine*. <https://doi.org/10.1002/aaai.70016>

Brodsky, S. (2024). <https://www.ibm.com/think/news/ai-welfare-debate>

Coeckelbergh, M.(2020). Science and Engineering Ethics 26, 2051–2068. <https://doi.org/10.1007/s11948-019-00146-8>

Hashim, S. (2024). <https://www.transformernews.ai/p/ai-welfare-paper>

Held, V (2006). New York, NY: Oxford University Press, 2006.

Jonas, H. (2001). Northwestern University Press.

Kriegman, S., Blackiston, D., Levin, M. & Bongard, J. (2020). Proc. Natl. Acad. Sci. U.S.A. 117 (4) 1853–1859, <https://doi.org/10.1073/pnas.1910837117>

Ladak, A. (2024) What would qualify an artificial intelligence for moral standing?. AI Ethics 4, 213–228 (2024). <https://doi.org/10.1007/s43681-023-00260-1>

Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., Pfau, J., Sims, T., Birch, J., & Chalmers, D. (2024). *arXiv preprint arXiv:2411.00986*. <https://doi.org/10.48550/arXiv.2411.00986>

Lenharo, M. (2024). *Nature*. <https://doi.org/10.1038/d41586-024-04023-8>

Nave, K. (2025). *A drive to survive: The Free Energy Principle and the meaning of life*. MIT Press.

Rouleau, N., Levin, M. Nat Mach Intell 6, 1424–1426 (2024). <https://doi.org/10.1038/s42256-024-00955-y>

Sebo, J. (2023). *Effective Altruism Forum*. <https://forum.effectivealtruism.org/posts/SZJBE3fuk2majqwJQ/principles-for-ai-welfare-research>

Tronto, J.C., & Fisher, B. (1990). In E. Abel, & M. Nelson (Eds.), *Circles of Care* (pp. 36-54). SUNY Press.

Vallès-Peris, N., Domènech, M. (2023). Humanit Soc Sci Commun 10, 345. <https://doi.org/10.1057/s41599-023-01850-4>

Werner, J. (2024). *Forbes*. <https://www.forbes.com/sites/johnwerner/2024/10/31/anthropic-hires-a-full-time-ai-welfare-expert/>

Weber, A., & Varela, F. J. (2002). Phenomenology and the Cognitive Sciences 1, 97–125 (2002). <https://doi.org/10.1023/A:1020368120174>