Who Needs Needy Machine Consciousness?

Max Jones
Department of Philosophy
University of Bristol
Bristol, UK
max.jones@bristol.ac.uk

***Abstract*— Much of AI Ethics has focused on the dangers of superintelligent machine consciousness. However, recent work on the homeostatic/affective basis of consciousness and on developments in soft robotics suggest that we may be able to build simple conscious machines in the relatively near future. As such it is pressing to ask whether we should build machines that may be conscious according to this theory. Here, I argue that we should avoid doing so because we don't need and certainly don't want needy conscious machines.**

Keywords—AI Ethics, Soft Robotics, Homeostatic consciousness.

I. Feelings, Genuine Intelligence, and Consciousness

Cantwell Smith argues that we may be further from creating genuinely intelligent AI than many assume [1]. While our current and emerging sophisticated AI systems may be capable of supreme reckoning abilities, they lack the capacity for judgement. In order to be capable of true judgement, AI systems must be capable of representing the world in a way that matters to them. Their decisions must have consequences for their own survival. In short, they must have needs.

Recent work on the neuroscience of consciousness suggests that the possession of needs and the ability to act in relation to them may be central to the function of consciousness [2, 3]. On this homeostatic/affective view, consciousness is inherently tied to affect, with a subject's affective states intimately linked to their overarching goal of maintaining homeostatic viability. This suggests that creating what Cantwell Smith sees as genuine intelligence may require the creation of (at least rudimentary) machine consciousness.

Man & Damasio have argued that soft robotics may provide a route to both genuine intelligence and machine consciousness [4]. By designing robots to be vulnerable in their exchanges with the environment and with a drive for self-preservation, it may be possible to render their interactions with the environment as meaningful in a way that rigid robots' interactions are not. One can similarly expand the repertoire of robots "feelings" by introducing other vulnerabilities, such as hunger or the need to care for offspring [5].

If these theorists are correct and genuine intelligence and machine consciousness just require us to build robots with needs, then the technologies for achieving these goals may not be as distant a prospect as some assume. There is already work on vulnerable soft robots that can act in relation to their own self-healing properties [6]. Robots with artificial guts that power themselves with energy from the environment have been developed [7, 8]. Self-replicating robots are already a possibility [9]. Moreover, one could presumably also engineer robots to care for the survival of their progeny (e.g. by implementing a biased ethical model based on simulation theory akin to [10]). On could go as far as to say that, if we wanted, we could build a robot with needs akin to biological organisms in the next 10 years. Perhaps we should stop thinking of these near future robots as machines and instead conceptualise them as artificial organisms [11].

II. Dennett's "No More Colleagues" Challenge

Supposing that the affective/homeostatic approach to machine consciousness is correct and engineering conscious robots isn't such a distant prospect, the pressing questions become whether and why we should want to build conscious machines of this kind. Dennett has argued that we should avoid creating conscious machines even if it is possible [12]. We have plenty of conscious colleagues already. "If you want a conscious agent, we've got plenty of them around and they're quite wonderful, whereas the ones that we would make would be not so wonderful." For Dennett, the costs of producing conscious machines may be great and in the end we just end up with more colleagues, which we were never short of in the first place. The value of AI systems and robots may lie precisely in their not being conscious, since they are able to carry out impressive feats of intelligence relatively cheaply without needing to experience anything and without having any needs that must be satisfied.

We don't have to worry about our AI systems and robots refusing to work because they don't *feel* like it or because they *feel* like they are being treated unfairly. We struggle to organise the economy in a way that allows us to look after the wellbeing and mental health of countless humans. Perhaps this should give us pause in aiming to introduce a whole swathe of new entities whose needs should be also met. "We want smart tools, intelligent tools, not artificial colleagues".

III. The Dangers of Needy Conscious Robots

Dennett's critique arguably applies to all forms of machine consciousness. However, it is particularly pressing in the context of the affective/homeostatic approach to machine consciousness. If machine consciousness requires robots with real needs that impact on their survival, then the continued existence of machine consciousness requires that those needs be met. This will clearly come with some significant dangers and costs. It's not hard to imagine the direct dangers to humanity that could result from creating vulnerable and hungry robots with a drive for self-preservation, a capacity to reproduce and a drive to also protect their vulnerable progeny. Since we are organisms with our own energetic needs, we would inevitably be in competition with such robots for energetic resources, so such robots would be right to infer that we are a threat to their own and their progeny's survival. Hungry robots can decide to eat you or to kill you to protect the resources that we are competing to secure.

One potential way to avoid these kinds of nightmare scenario and to mitigate the potential costs of machine consciousness is to build robots that can only meet their energy requirements by consuming materials that we are unable to consume. Even better would be to do this with materials that we are keen to get rid of.

For example, one could build robots that are hungry for and capable of digesting the plastics that we need to remove from the ocean [8]. In short, we can try to ensure that the robots can meet their needs in a way that aligns with our own needs. However, the problem with this line of thought is that it's not clear why conscious robots would be needed to fulfil tasks that happen to align with the presence of unwanted resources. Most that see applications for conscious robots tend to envisage them being better at interacting with humans, but we tend to avoid locations that are full of toxic of polluting resources. Moreover, this may be precisely the kind of application where machine consciousness could potentially be dangerous, as it might enable the kind of behavioural flexibility that allows the robot to set its own goals that deviate from what we desire. If there are ways of solving similar problems (perhaps using robotics) that nonetheless don't involve the costs and risks of machine consciousness, then these should be preferred.

Given these risks and costs, the potential benefits of creating machine consciousness would have to be large to make its development a worthwhile endeavour. Unfortunately, the benefits of machine consciousness are far from clear. Some have argued that machine consciousness may be necessary for robots to feel empathy, and that this may be a prerequisite for their being able to engage in moral deliberation [4]. This may be true, but it just points to the further question of why we would want robots to be capable of such deliberation. The capacity for moral deliberation is not the same as a guarantee of morally acceptable action. After all, we have plenty of examples throughout history of conscious humans, capable of empathy and moral deliberation, doing abhorrent things to one another.

IV. Protecting Against Needy Robots

The deeper lesson that we should learn from considering the homeostatic/affective approach to machine consciousness is that we shouldn't assume that consciousness is aligned with superintelligence. Most work concerning consciousness in AI ethics has focused on worries about the singularity and super-intelligent conscious AI [13], but the homeostatic/affective approach to consciousness in living things takes it to be relatively widespread and possessed by relatively simple and unintelligent creatures. We therefore need to be careful, as building needy and not so intelligent conscious robots may be something already or nearly within our grasp, as a result of developments at the intersection of robotics and artificial life.

What can we do to protect against these costs and dangers? One option is to give robots needs without giving them the capacity to register those needs. One can build a robot that is, in a sense, hungry (in that it is energetically autonomous and must consume environmental resources to maintain its own viability) without building a robot that represents its own homeostatic state. In doing so, we may have to forgo the potential benefits, for example, in terms of behavioral complexity, of a robot that, in some sense, genuinely *feels* hungry. Yet, this may be a price worth paying to avoid the dangers of a robot with needs that it is aware of.

A second option is to accept that the advent of needy robots is inevitable and to ensure that we have the requisite moral framework and safeguards in place. Importantly, the relevant moral framework is unlikely to emerge from considerations of superintelligence and the singularity. Rather, we should turn to established issues in non-human animal ethics and the bioethics of artificial life [14, 15].

Most importantly, any discussion of whether we should create potentially conscious machines with their own homeostatically driven needs, should be conducted in the context of considering the needs of already existing humans and nonhuman animals. By creating new entities with needs of their own, we inevitably create more needs that must be satisfied with a still finite pool of resources. Conscious machines may be just around the corner. All we need to do is give them needs and the capacity to register and act upon those needs. What isn't clear is whether and why *we* need a proliferation of needy companions, particularly when we already struggle to cater for the needs of those that already exist.

References

[1] B. Cantwell Smith, The Promise of Artificial Intelligence: reckoning and judgement. Cambridge, MA: MIT Press, 2019.

[2] M. Solms, The Hidden Spring: A journey to the source of consciousness. London: Profile Books, 2021.

[3] A. Damasio and H. Damasio, "Homeostatic feelings and the biology of consciousness," Brain, vol. 145, pp. 2231-2235, May 2022.

[4] K. Man and A. Damasio, "Homeostasis and soft robotics in the design of feeling machines," Nature Machine Intelligence, vol. 1, no. 10, pp. 446-452, 2019.

[5] D. Parisi and G. Perosino, "Robots that *have* feelings," Adaptive Behavior, vol. 18, no. 6, pp. 453-469, 2010.

[6] E. Roels et al., "Processing of self-healing polymers for soft-robotics," Advanced Materials, vol. 31, no. 6, 2022.

[7] C. Melhuish, I. Ieropoulos, J. Greenman, and I. Horsfield, "Energetically autonomous robots: Food for thought," Autonomous Robots, vol. 21, pp. 187-198, 2006.

[8] H. Philamore, J. Rossiter, A. Stinchcombe, and I. Ieropoulos, "Row-bot: an energetically autonomous water boatman," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3888-3893, 2015.

[9] K. Lee and G. S. Chirikjian, "Robotic self-replication," IEEE robotics & automation magazine, vol. 14, no. 4, pp. 34-43, 2007.

[10] D. Vanderelst & A Winfield, "An architecture for ethical robots inspired by the simulation theory of cognition," Cognitive Systems Research, vol. 48, pp. 56-66, 2018.

[11] J. Rossiter, "Soft robotics: the route to true robotic organisms," Artificial Life and Robotics, vol. 26, no. 3, pp. 269-274, 2021.

[12] J. Brockman, D. Chalmers and D. Dennett, "Is Superintelligence Impossible? On Possible Minds: Philosophy and AI with Daniel C. Dennett and David Chalmers", Edge Conversations, 4.10.19.

[13] V. Müller, "Ethics of artificial intelligence and robotics," Stanford Encyclopedia of Philosophy, 2020.

[14] T. Douglas, R. Powell, and J. Savalescu, "Is the creation of artificial life morally significant?" Studies in History and Philosophy of Science Part C, vol. 44, no. 4, pp. 688-696, 2013.

[15] A. Christiansen, "Synthetic biology and the moral significance of artificial life," Bioethics, vol. 30, no. 5, pp. 372-379, 2016.