

Chapter XIV

Artificial Moral Agency in Technoethics

John P Sullins

Sonoma State University, USA

ABSTRACT

This chapter will argue that artificial agents created or synthesized by technologies such as artificial life (ALife), artificial intelligence (AI), and in robotics present unique challenges to the traditional notion of moral agency and that any successful technoethics must seriously consider that these artificial agents may indeed be artificial moral agents (AMA), worthy of moral concern. This purpose will be realized by briefly describing a taxonomy of the artificial agents that these technologies are capable of producing. I will then describe how these artificial entities conflict with our standard notions of moral agency. I argue that traditional notions of moral agency are too strict even in the case of recognizably human agents and then expand the notion of moral agency such that it can sensibly include artificial agents.

INTRODUCTION

The various technosciences of artificial agency such as, artificial life (ALife), artificial intelligence (AI), and robotics present a rather challenging problem to traditional ethical theories whose norms rely on an explicit or tacit notion of human personhood since these entities will share only some, but not all, of the qualities of the humans they will interact with.

The field of technoethics must disentangle this problem or be faced with the charge of inco-

herence. This is due to the fact that technology extends the biological limits of the human agent in such a way that it is often difficult to draw a clear line between the human agent and the technologies she uses. Artificial creations such as software bots, physical robots, and synthetic biological constructs are unlike anything we have encountered yet and in them something like individual agency is beginning to evolve. This quasi-individual agency is already placing these entities in conflict with the goals and desires of human agents, creating apparently moral

interrelations. What is the nature of these moral interrelations?

We have three possible answers to this question (see Sullins, 2005). The first possibility is that the morality of the situation is illusory, we simply ascribe moral rights and responsibilities to the machine due to an error in judgment. The second option is that the situation is pseudo-moral; a partially moral relation but missing something that would make the actors fully moral agents. A final possibility is that even though these situations may be novel, they are still real moral interrelations. I argue that technoethics must address this latter possibility.

BACKGROUND

It is not an obvious move to grant moral concern to the nonhuman objects around us. It is common to hold the view that the things we come into contact with have at best instrumental value and that only humans have moral rights and responsibilities. If some nonhuman thing elicits moral concern, it does so only because it is the property of some human through whom these rights extend. This all seems very straight forward and beyond question. But here is my worry—we have been mistaken in past about our definition of what it takes to be a human moral agent. Historically women, low caste men and children have been denied this status. We have come to regret these past indiscretions, it is possible that that our beliefs about moral agency are still misguided.

Some people may be willing to grant moral rights to animals, ecosystems, perhaps even plants. If machines were shown to be similar to these things might they not also be reasonable candidates for moral rights? If so, what happens if these entities acquire agency similar to that of a human, then must they also bear moral responsibilities similar to that of a human agent? The answer to the latter question is simple; of course anything that displays human level agency enough

to satisfy even harsh critics would be a candidate for moral rights and responsibilities because it would have undeniable personhood and all persons have moral worth. The possibilities for this happening any time soon though are fairly low. But, they have made some progress on attaining interesting levels of agency, so what we need to inquire into is whether or not these meager qualities are enough to grant moral agency and worth to artificial agents.

What is an Artificial Agent?

The term “agent” has a number of related, but potentially confusing, meanings. An agent is most simply; a thing that exerts power and is the opposite of a patient which only reacts to or receives the consequences of the actions of an agent. Thus it may be used to talk about a person or a thing that has some causal effect on its environment or other agents. Commonly, it is used to refer to persons who act on others behalf. These slightly different meanings converge in the definition of the term “artificial agent”, which I will use to refer to any technology created to act as an agent, either as a locus of its own power, or as a proxy acting on behalf of another agent. So an artificial agent might have its own goals that it attempts to advance or, more likely, it is created to advance the goals of some other agent.

Certain criteria are used by technologists to distinguish autonomous (artificial) agents from other objects:

*An **autonomous agent** is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future.* (Franklin and Graesser, 1996)

This definition could be used to describe a wide range of artificial entities, some of which could be very minimal and philosophically uninteresting. Franklin and Graesser (1996), go on to list a num-

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/artificial-moral-agency-technoethics/21581

Related Content

Application of Error Control Coding for Multimedia Watermarking Technologies

Mehul S. Raval (2013). *Digital Rights Management: Concepts, Methodologies, Tools, and Applications* (pp. 873-891).

www.irma-international.org/chapter/application-error-control-coding-multimedia/71009

Review of Sorgner's Philosophy of Posthuman Art

Andrei Nutas (2022). *International Journal of Technoethics* (pp. 1-7).

www.irma-international.org/article/review-of-sorgners-philosophy-of-posthuman-art/313197

ELATEwiki: Evolving an E-Learning Faculty Wiki

Shalin Hai-Jewand Roger McHaney (2013). *Digital Rights Management: Concepts, Methodologies, Tools, and Applications* (pp. 929-951).

www.irma-international.org/chapter/elatewiki-evolving-learning-faculty-wiki/71012

Social Constructivism vs. Pragmatism: A Search for a Suitable Social Work Paradigm for Research on Immigrants

Ziaul Islam Jewel (2024). *Methodologies and Ethics for Social Sciences Research* (pp. 123-137).

www.irma-international.org/chapter/social-constructivism-vs-pragmatism/337053

Technoethical Inquiry into Ethical Hacking at a Canadian University

Baha Abu-Shaqraand Rocci Luppini (2016). *International Journal of Technoethics* (pp. 62-76).

www.irma-international.org/article/technoethical-inquiry-into-ethical-hacking-at-a-canadian-university/144827