

The Ethical Value of LAI in Social Robotics

James W. Besse
Instytut Filozofii i Socjologii,
Polskiej Akademii Nauk
(Institute of Philosophy and Sociology,
Polish Academy of
Sciences), Warsaw, Poland
besse.jw@gmail.com

Rosangela Barcaro Ph.D.
Independent scholar
Genova, Italy
rosangela.barcaro@gmail.com

In this paper, we take the position that the ethically important aspect of artificial intelligence (AI) is the entry of unintelligent machines into human affairs. Elaborating upon the views of Luciano Floridi (1999), we show why an effective example of AI is a machine which engages in a simple array of tasks and processes. Intelligent machines, we hold, should be approached from a perspective which recognizes the reality of their lack of human-like intelligence, while still acknowledging their success as companions. The paper begins by explicating Luciano Floridi's critique of Alan Turing in *Philosophy and Computing* (1999) and advocacy of light artificial intelligence (LAI), and begins to explain some of the full implications of his view by showing the ways in which a passion for non-human intelligence existed even in Turing and his colleagues. In the following section, we move through the assumptions made by Cynthia Breazeal of MIT, and demonstrate social robotics' compatibility with Floridi's ideas. We examine several examples to defend our point about the successes of LAI in social robotics. In the final section, we examine the ethical consequences of LAI in social robotics, such as openness to alterity and realization of the human interrelatedness with technology.

Social robotics. Artificial intelligence (AI). Human-robot interaction. Strong AI. Light artificial intelligence (LAI). Robot ethics. Posthumanism. Technological mediation.

1. INTRODUCTION (WHAT AI?)

In this first section, arguing alongside Luciano Floridi (1999), we will advance a view of artificial intelligence (AI) which stresses the practical disadvantages of designing AI to mimic human intelligence. Our view is that task-oriented AI has marked advantages over mimetic approaches. In *Philosophy and Computing* (1999), Luciano Floridi expresses a critical stance towards the strong view of artificial intelligence (AI). The strong view of artificial intelligence (or Good Old Fashioned AI, GOFAI, as he prefers to call it) asserts the vision of creating machines "whose behaviour would eventually be at least comparable, if not superior,

to the behaviour characterising intelligent human beings in similar circumstances" (Floridi, 1999). Floridi's critique of GOFAI can be summarized in the following points:

- (i) Advocates of GOFAI enact a reduction of human intelligence to computation which they claim to be an equation, without considering the individual and embodied nature of human intelligence.
- (ii) Given that the rules of Turing's Test are such that answers require only symbolic processing, a machine made by GOFAI, even if it succeeded by GOFAI's own terms and consequently passed the test, would

not be intelligent by human standards.

Turing's Test does not allow for oral communication, open-ended questions, critical questions, context-sensitive questions, dialogue between the interviewer and the interviewees, or questions where "linguistic pragmatic context" is used (Floridi, 1999). Floridi writes: "TT-conditions are so constraining that all intelligent behaviour of both C and W is strictly reduced to symbolic processing tasks." (Floridi, 1999)

By making these points, Floridi casts the whole project of GOFAI into doubt, showing the ways in which it is a reductionist project which does not take into account the actualities of human intelligence. An intelligent machine which would pass Turing's Test, Floridi points out, would only need to respond in an accurate way to questions which were suited to a reductionist (rational and computational) conception of human intelligence. Turing, by developing this theoretical framework which 'stupefies,' simplifies, or narrows, human intelligence, is charged by Floridi as having lowered the bar for what counts as an intelligent machine. It also demanded that human intelligence (albeit in a caricaturesque, mistaken depiction of it) be the goalpost for AI research.

Floridi follows his argument against GOFAI with an advocacy of what he calls LAI (Light Artificial Intelligence). This is an approach to AI which is task-oriented, and does not seek to reduce human intelligence by positing that a machine can do it just as well as a human being. Due to its lack of success in modeling human intelligence using computation, GOFAI can be supplanted by an approach which creates machines in a task-oriented way. This approach recognizes that human intelligence does not need to be the standard for machine intelligence, and in fact that human intelligence is often less suited for certain tasks than is machine intelligence. Floridi writes that "LAI is performance-oriented or *constructionist*, not *mimetic*" (Floridi, 1999). This means that the same tasks can be performed in several different ways and after having established a specific task, the "computer-way" of doing it needs to be discovered (Floridi, 1999). The purpose is not creating a program "that can mechanically prove a logic theorem exactly as a human being would, because we do not need either the latter's mistakes

and distractions, or his or her insights and intuitions" (Floridi 1999). Nor do we want our GPS to be as smart as a human being, because we do not want the inconsistency and second-guessing that are characteristic of human intelligence. We also do not want our GPS to respond to some of our requests by asking "Do you really want to drive there?" A GPS can fit perfectly into the social environment of a long drive, and do so precisely because it does not have human-like intelligence.

Floridi expands upon his point that LAI is successful by developing his theory of "enveloping" or "re-ontologizing" (Floridi 2011). This argument posits that tasks have been changed such to accommodate their being done by machines. Referring to driverless vehicles, he writes: "If drones or driverless vehicles can move around with decreasing troubles, this is not because productive AI has finally arrived, but because the 'around' they need to negotiate has become increasingly suitable to[...] AI and its limited capacities" (Floridi 2011). So driving in the highly developed urban and residential infrastructure has become a task which accommodates machines. It is not that cars have become smart enough to replicate the intuition and judgement of a human driver. One could simply do an experiment and bring a self-driving car to an extremely rural or unpredictable environment (for example rural Maine or the pedestrian saturated streets of New Delhi) and see the glaring failures and disasters this might cause.

Another good example is the dishwasher. This is not a machine which is in any way intelligent by human standards, yet a dishwasher tends to be far better at washing dishes than human beings. Why is this? Because the task of washing dishes, for the designers of dishwashers, is redesigned to be done according to the capacities of a machine. Effective and interesting machines tend to be suited to the environment in which they are deployed, they are not built with the goalpost set at human intelligence but find value and interest in non-human sorts of intelligence. We can ascertain that the better an engineer sets purposes and tasks for a machine, the better the interaction with human beings, and the better the success in a specific domain in which the machine can be introduced.

However, despite what we consider to be his invaluable contribution to theorizing AI, we think Floridi is missing the point that could make his own

argument stronger, that is namely that Turing had an astoundingly inclusive view of social relationships, and saw no problem with caring for a machine which was not intelligent in a human sense. Turing's relationship with the machines he was creating was an affective one, and he was not alone in this fact. Frustration, surprise, and care, all played an important role in what kind of relationship existed between engineer and machine. "The early logbooks of the University of Manchester's Ferranti Mark I suggest that an enormous amount of adult care was needed to keep this juvenile machine functional. User entries in late 1951, for example, document many unexpected (funny) events, a lot of trial-and-error learning, and ongoing emotional engagement" (Wilson 2010). Wilson documents the extent to which, in this early period of artificial intelligence research, people found themselves in relationships with machines which would surprise them, make them laugh, demand care from them, and affirm the extent to which intelligence is not always a condition for love and respect. What this suggests is that artificial intelligence was developed on the basis of human willingness to engage emotionally with machines, not merely a reductive understanding of intelligence. Far from being a simply cognitive affair, AI in the early period was founded on enthusiasm which was not bound by the engineers' focuses on intelligence. Of course, we accept and endorse Floridi's argument about Turing's views of intelligence. However, it is important to remember that affective engagements existed between humans and machines, regardless of what is traditionally thought to be the point of AI, that is intelligence. We hold that if Floridi's argument against GOFAI is read in this context, it becomes even more interesting and valuable. An openness to the diversity of intelligence within sociality can be considered a major consequence of the refutation of GOFAI.

2. SOCIAL ROBOTICS

Social robotics emerged in the late 1990s and early 2000s with researchers who began to apply the ideas of artificial intelligence to the atomization of social, as opposed to merely cognitive, intelligence. Cynthia Breazeal and her colleagues working at the MIT Media Laboratory pioneered social robotics in the United States, and began creating artificially intelligent machines which could engage in social and emotional tasks. The field of social robotics has been marked, to an even greater extent than

artificial intelligence, by affective connections between humans and machines. The roots of this phenomenon, we have shown, go back to the beginning of artificial intelligence, but in social robotics it has become much more pronounced. This is because the social roboticists have worked with the explicit intention of building machines which provoke human beings into interacting with them; or as Turkle puts it, these machines push our "Darwinian buttons" (Turkle, 2011). In Floridi's opinion, by the standards of GOFAI, they are certainly not much smarter than the machines of the early period of artificial intelligence. The difference between the GOFAI program and social robotics appears to be in the purposes: in social robotics, the machines interact with human beings in order to perform everyday activities, in cooperation and personal assistance. These machines are specifically designed to be interactive social companions and assistants. To that end, we will show in the next section, non-human intelligence is an advantage.

In her early work describing her intentions for building the social robot Kismet, Cynthia Breazeal was clearly advocating for something similar to GOFAI (Breazeal, 2002). However, this does not mean she was bound to a strict idea of intelligence in her engagements with machines. The attachment Breazeal had to Kismet is well known. As Sherry Turkle, the sociologist of technology who was at the Media Lab concurrently with Breazeal, writes, "The last time Breazeal will have access to Kismet [, she] describes a sharp sense of loss. Building a new Kismet will not be the same. This is the Kismet she has 'raised' from a 'child.' She says she would not be able to part with Kismet if she weren't sure it would remain with people who would treat it well" (Turkle, 2011). Social robotics, we see in the case of its American pioneer, was an emotionally motivated affair, propelled by the desire of its engineers to engage emotionally with machines. This was also the case with the people who engaged socially with Kismet in the studies at the Media Lab. People, as noted by Breazeal and Turkle, were more than willing to open themselves and converse in deep and vulnerable ways with the robot. This was also the case with Joseph Weizenbaum's ELIZA chatterbot, with which students expressed a desire to articulate their feelings in intimacy. The early period of social robotics in the United States was marked by people's openness to emotional relationships with

machines, and a dramatic expansion of what could be said to have feelings and deserve emotional reciprocation. The age of machines is certainly here, but it has little to do with Terminator, Ultron, or the replicants of Philip K. Dick's vision; nor should it. The age of machines is about machines which are far from human standards of intelligence, yet are accepted open heartedly into our homes and lives.

3. LAI, SOCIAL ROBOTICS AND ETHICS

Florida's argument, applied to social robotics, gets us into ethical territory. We advance the position that, from an ethical standpoint, it is actually better for social robots not to have human intelligence. In this section, we will also make reference to the work of Peter-Paul Verbeek, which we think is a philosophy of technology particularly friendly to LAI. Our position rests on the following: 1) GOFAI preserves an outdated divide between humans and objects which can show some characteristics of human thinking, 2) This therefore excludes from the category of legitimate partners non-human beings, objects and those humans which can't think in the conventional sense, 3) LAI encourages people to engage socially and emotionally with things sans the criteria of normative notions of human intelligence, and breaks the line between the human and the non-human.

This is not merely that, as some (particularly scholars of the feminist critique of sex robotics, like Richardson, (2015)) suggest, reciprocity becomes no longer important in social relationships, but the criteria of legitimacy for partners in social relationships become less reliant upon strict ideas of intelligence and humanness. If we can embrace the openness of mind that social roboticists have given us, and follow the lesson of engaging emotionally and respectfully with the range of beings in our world (human or non-human, living, or non-living) regardless of their intelligence, important normative ethical practices are given to contemporary culture. Our position has clear applications to (and lends clear support to) the arguments of posthumanism. As Braidotti (2013) suggests, humanism was built on a strict universal idea of the white European man. The strictness of humanism's normative ideas of humanness and what and who can be legitimately respected and engaged with, as anyone basically familiar with the history of the 20th Century knows, has led to

unparalleled destruction. Of the European Jewish and Roma populations, of refugees, of 'livestock' such as pigs and chickens, and of the planet, examples have been made of the dangers of limiting who and what can be considered a legitimate partner in a relationship based on care and respect. Being able to enter into emotional relationships with non-humans sans criteria of intelligence is an ability which suggests something good is starting to take place in our culture. We are still responsible for these aforementioned crimes. To say anything else is inexcusable. At least, however, artificial intelligence suggests that, for some of us, the social world is becoming somewhat less exclusive.

Now we move onto examples. Take a look at two of the social robots which are currently popular in Europe and North America. They include Paro (the robotic harp seal used to comfort elders in nursing homes) and Kuri (a home robot). These robots, by human or even non-human animal standards, are quite stupid. However, they are fully capable of being emotional and social partners, and their lack of human intelligence is advantageous to that end. Paro is an interesting case of extremely successful LAI, and its social and emotional life with humans is actually improved by its lack of human-like intelligence.

Nursing homes usually do not accept animals: many elderly people have compromised immune systems which would make it dangerous to be in proximity to living animals, and hygienic concerns limit interactions with live animals. Nonetheless the elderly people might benefit from taking care of pets. Paro might provide a feasible solution, and has found its way into nursing homes across the world. It is hygienic, cute and cuddly, but can it serve as an effective companion? Turkle (2011) suggests as much, and shows that in improving the emotional wellbeing of the elderly, Paro has comparable results to human contact. For the purpose of completing its functions it does not need to be as intelligent as a biological harp seal, and its 'stupidity' seems to be an asset (as it is unlikely that a biological harp seal would tolerate long stretches of being held by the elderly). The people who engage with Paro do not tend to care that it does not have intelligence to rival a human being, or even a baby seal. Rather, they enthusiastically love, respect and engage with the robot in its

inanimacy and artificiality, without demanding more (Pettman, 2012).



Figure 1: Paro with an elderly person (Crowley n.d.).

Kuri is an additionally interesting example, as its task is to be a home robot which entertains, records family life, and can help with practical tasks (such as reading the weather forecast and grocery lists, or playing music). For the tasks Kuri is made, it is an extremely successful machine. It is even capable of expressing emotions in very simple ways (such as whistles, eye and head movements, and a light in its torso), which draw people into attachments with it. In family settings, Kuri can express satisfaction and dissatisfaction, and most importantly perform the limited range of tasks for which it is designed. Whereas Paro mostly interacts with elderly people in the context of nursing homes, what Kuri demonstrates is a case of non-clinical success of LAI in social robotics.

The reason we find GOFAI problematic is because it preserves an outdated line between human beings and technology, and relies upon a strictly human-centered idea of intelligence. By setting the goal post of artificial intelligence at the point where it becomes able to compete with human intelligence in human tasks, GOFAI not only engages in a pointless endeavour (because as Floridi writes, it's simply better to have a machine do machine tasks, or re-engineer tasks to be more easily automated), but hinders the social and cultural recognition of the role of technologies in human society which Simondon (2012) so ardently says is needed, and the bearing of the fruit of such a recognition. Perhaps the interconnectedness of human life to technologies could help to mitigate the damages of certain technologies and bring the benefits of certain others to fruition. Medical technologies such as care robots and data-driven outpatient care are two examples of where people could put their focus and efforts if they had a clear understanding of how humans and technologies

are interrelated as opposed to in competition. GOFAI, regardless its demonstrated inability to produce significant results to date, is a mistaken endeavour because it seeks to effectively sever the relationship between human beings and technology by making technology a stand-alone entity: Rather than "mediation," as Verbeek (2005) outlines, or "enveloping," as Floridi (2014) outlines, one gets Terminator or Ultron: Technology on one side, able to function with or without human beings, and humans on the other who are both the model for machine intelligence and held separate from it. GOFAI enacts a theoretic covering-over of the most interesting aspects of artificial intelligence: Namely, the mediation and re-ontologizing of our experience by an ever-growing range of partners which, although (and because) they are not sentient in a human sense, draw people away from a narrow human-centered view of the world. Our experience as human beings is saturated by machines and, qua process, is shaped by our interactions with them. Verbeek (2011) writes, "There is an interplay between humans and technologies within which neither technological development nor humans has autonomy. Humankind is a product of technology, just as technology is a product of humankind." Humans and technologies are not to be seen as separate from each other but rather, as Verbeek's theory of technological mediation suggests, they formed in the context of interaction. Technology mediates "the relation between humans and their world, amongst human beings, and between humans and technology itself" in such a way to shape each of them in a fundamental way (Verbeek, 2005). The human world cannot be separated from the world of machines. Like Floridi's aforementioned work, this theory holds a special place in our consideration of the ethics of AI because it is honest to the realities of human experience. The success of LAI makes us question what it means to be human not by adopting a reductionist approach and the assertion that a brain works like a computer, but rather by showing the co-constitutive relationships between human and non-human experience and processes. In our contemporary world, both human reality and experience include machines and are shaped by interactions with them. To deny this, as the pursuit of GOFAI seems to do, preserves an outdated line between humans and machines (and non-humans in general). As Floridi (2017) writes, artificial intelligence is about us. It is important to get our concepts right. We must understand why artificial

intelligence was developed with such an enthusiasm, and how it offers ways of humans rethinking our relationships to each other and to non-human beings. Our answer is that AI was developed so enthusiastically because of human emotional attachments to non-humans, and that it offers us a vision of human reality as given shape by our relations to non-humans. Social robotics, when it follows the program of LAI, can push people towards an embracing of alterity, and a comprehension of our relationship to machines. LAI in social robotics, more than simply a pragmatic approach, is an ethical one as well.

4. CONCLUSION

In this paper, we hope to have shown that looking at Floridi's ideas about LAI have significant ethical implications when applied to social robotics. Floridi takes a critical view of those early ARI researchers, such as Alan Turing, who were interested in creating machines with human intelligence. This criticism, we have found, is accurate and extremely important. However, Floridi (and this may be because it is simply not the point of his argument) does not adequately emphasize that AI researchers have, throughout the field's history, displayed an almost childlike enthusiasm for engaging with objects, and a deep willingness to care and feel emotions for them, regardless of their intelligence. It seems that LAI's success was foreshadowed as early as the 1950s. This makes Floridi's argument for LAI even stronger, especially when it is applied to social robotics. It also brings LAI into the domain of ethics, especially with regards to social robotics. We hope to have shown how social robotics has seen great success with LAI, and that how there are ethical issues at stake. Among these issues is our ability to form effective relationships with those who do not share our form, or level, of intelligence. LAI in social robotics also directly leads to a consideration of the relationship human beings have to technology. In understanding this, Floridi's work on the Infosphere is essential.

5. REFERENCES

Braidotti, R. (2013) *The Posthuman*. Polity, Cambridge, UK.

Breazeal, C. L. (2002) *Designing Sociable Robots*. The MIT Press, Cambridge, MA.

Crowley, S., n.d. Digital photograph, The New York Times, viewed 15 June 2018, <https://www.nytimes.com/2010/07/05/science/05robot.html?_r=2&pagewanted=1>

Floridi, L. (1999) *Philosophy and Computing. An Introduction*. Routledge, London and New York.

Floridi, L. (2011) Children of the Fourth Revolution. *Philosophy & Technology*, 24(3), pp. 227-232.

Floridi, L. (2014) *The Fourth Revolution. How the Infosphere is Reshaping the Human Reality*. Oxford University Press, Oxford.

Floridi, L. (2017) Roman Law Offers a Better Guide to Robot Rights Than Sci-Fi. Financial Times [online]. Available at: <https://www.ft.com/content/99d60326-f85d-11e6-bd4e-68d53499ed71> (Accessed 15 June 2018).

Pettman, D. (2009) Love in the Time of Tamagotchi. *Theory, Culture & Society*, 26(2-3), pp. 189-208.

Richardson, K. (2015) The Asymmetrical 'Relationship'. Parallels Between Prostitution and the Development of Sex Robots. *ACM SIGCAS Computer and Society. Special Issue on Ethicomp*, 45(3), pp. 290-293.

Simondon, G. (2012) *Du mode d'existence des objets techniques*. Flammarion, Paris.

Turkle, S. (2011) *Alone Together. Why We Expect More from Technology and Less from Each Other*. Basic Books, New York.

Verbeek, P.-P. (2005) *What Things Do. Philosophical Reflections on Technology, Agency and Design*. The Pennsylvania State University Press, University Park, PA.

Wilson, E. A. (2010) *Affect and Artificial Intelligence*. University of Washington Press, Seattle and London.

6. ENDNOTES

[1] We would like to thank Zachary Isrow for his help with editing this paper, and Peter-Paul Verbeek for his kind and instructive comments during and after our presentation.