Journal of
Ethics and Emerging Technologies

IEET

Book Review

"Living with Robots: What every anxious human needs to know". Ruth Aylett and Patricia A. Vargas, MIT Press, 2021.

Gabriela Arriagada Bruneau [1,2*]

[1] University of Leeds, Inter-disciplinary Applied Ethics Centre (IDEA), Leeds, England.
[2] Think tank 'Pensar en red', Director of Applied Ethics, Santiago, Chile.
* Correspondence: gabruneau@gmail.com

This book takes us to explore the 'life of robots' and presents us with a refreshing narrative that demystifies their recurrent anthropomorphic understanding. General ideas of what robots are and what they can do often lack knowledge about the limitations, functionality, and complexity needed to turn a robot into a fully operational machine. In this book, the authors portray a grounded and accessible description of current research developing robots. They are insightful, yet still allow the reader to understand basic processes and requirements of robotics. They deliver on their goal to show robots as human artifacts instead of placeholders for anxiety, i.e., a threat. However, they underdeliver regarding the ethical insights of robots as moral agents in the final chapter. Withal, this is an essential introductory reading for anyone interested in learning about the functioning of robotics and AI, and their integration to society.

To unpack the real contributions robots can bring into society, Aylett and Vargas engage the reader with a bit of storytelling. Chapter 1 starts with the tale of the Pygmalion, showing how robot-like machines are often based on narratives filled with awe and fascination. They present the conceptualisation of robots in history and the influence of pop-culture to form general ideas of robots as a threat to humans. The "*Frankenstein complex*" refers to robots that can turn against their creators. In part, this is related to the psychological distinction between in-group and out-group recognition. We favour ideas or values we share, thus robots are understood as 'the others'. This, according to the authors, makes the whole discussion about living with robots, fundamentally, a human affair.

From Chapter 2 onwards, thematic chapters highlight (challenging) characteristics required to build different types of robots. The second chapter is dedicated to *appearance*, unfolding concepts such as embodiment and agency. The authors explain the importance of the concept of social affordance for robots, i.e., robots reflecting their functionality in a way that is intuitive for humans. Overall, the relevance of human-robot interaction is presented as a worry about *form* and *function*. Robot appearance creates an expectation of the type of interaction we can have with it, and of its physical functionalities. Accordingly, the risk of entering the *uncanny valley*, which refers to the resemblance a robot has to a human being (physically and behaviourally) is key. An imperfect yet too close of a resemblance provokes a feeling of *strangeness* in users interacting with these machines. This chapter tells us about the importance of robots looking, acting, and being like humans, questioning if that is something we should really pursue.

Chapter 3 introduces *movement*. Here the authors show us how integrating movement into a machine can be done with *degrees of freedom*, controlling the different independent movements a machine can carry out using software. However, the more degrees of freedom, the higher the risk of *redundancy* (different combinations producing the same desired movement), which demands more complexity and computational power. To that, we must add the challenges of incorporating sensorial input. We often take the complexity of biological movement for granted, and it is an extremely difficult task to translate into

Citation: Arriagada-Bruneau, Gabriela. 2021. "Living with Robots: What every anxious human needs to know". Ruth Aylett and Patricia A. Vargas, MIT Press. *Journal of Ethics and Emerging Technologies* 31, (2).

Received: 04/08/2021
Accepted: 22/12/2021
Published: 22/12/2021

Publisher's Note: IEET stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

machines. Robotics often uses bio-inspired principles and *biomimetics* to take inspiration from animals like birds, lizards, or fish to engineer movements, as a way of integrating them into our *social space*. Yet, limitations to achieve the ideal biped robot persist.

Chapter 4 and Chapter 5 introduce us to two key concepts for robots to engage with their environment: *senses* and *geographical awareness*. Sensors in robots are an extremely important source of data for many things, including localization and navigation. Robots process everything in numbers, they do not see or feel a tree, they amount pixels to numbers and recreate a pictorial scene. A main issue in robotics is the presence of *noise* in most of the data's input to the machine, making sensor data highly *uncertain*. As noted by Aylett and Vargas, sensors give data, but robots need *information* to engage with their environment (language interaction, object recognition, etc.). And acquiring information remains a costly task.

Now, using information presents another challenge. Navigation comes relatively easy to *us*, however implementing this in a robot is extremely difficult. In Chapter 5, the authors show why robots are still highly dependent to the environments they have been trained on. Processing new information can be an almost impossible task, which has taken researchers to integrate SLAM (simultaneous localization and mapping), where a robot can estimate current positions based on previous landmarks. Perception and anticipation continue to be one of the main struggles for robotics.

Chapter 6 focuses on *touch and handling*. This chapter specifies the difficulties of using robots for common daily tasks, especially if they require fine motor skills. It also explains the research behind technologies capable of mimicking animal muscle effects, like *electroactive polymers* and textile-based solutions such as *shape memory*. The authors introduce a sensible aspect of HRI (Human Robot Interaction), explaining the alternatives to apply a human-like skin into machines, drawing parallels with some previously discussed characteristics such as movement or appearance (*uncanny valley*). Towards the end of the chapter Aylett and Vargas criticize a sensational take concerned with the *singularity* risk, robots becoming 'superior' to humans —as they presented it, and *transhumanism*, the idea that technology can transform humans physically and intellectually: "Into some kind of super-cyborg species, 'the bionic man' Somehow it always does seem to be "man."" (pg. 103). This last statement insinuates an additional feminist critique, but we are left with that, and the next paragraph continues with the challenges of developing prosthetics and the application of medical devices. I believe that further engagement with these concepts was warranted. Singularity and Transhumanism are a rich and thriving area of AI, Robotics, and Ethics that inevitably interact with the meaning and attributions we give to technology, including robots, embodied AIs, and human enhancements. This was a missed opportunity to give the reader a more developed critical stance about our understanding of robots and their interactions with humans.

Chapter 7 presents us the debate about *intelligence* and the question of whether robots can be artificially intelligent. Across the chapter, the authors highlight different abilities or functions associated with the concept, including learning, understanding, memory, and judgement, and explore the ways in which something can be called intelligent. Here, we are once more introduced to key concepts for robotics, including the *architecture* of the robot and its *reactive layers*, that tell us how the robot *behaves*. A fundamental distinction between data and information is given, exemplifying why robots having access to the internet is not going to make them riskier by enabling them to pursue enlarged goals. Context and environment limit robots to perform poorly at re-planning and adapting goals, something even humans struggle with. The end of the chapter reminds us that AGI (Artificial General Intelligence) is still a far-fetched possibility.

Chapter 8 unpacks the concept of *learning*. *What* can be learned and *how*, are important questions providing insight into what is required for a robot to learn something. The authors take us through the three types of computerized learning: supervised, unsupervised, and reinforcement learning. Computerized learning means giving an input and getting the *right*

output. If the robot covers enough input-output pairings, it can then select the *right outputs*, i.e., *learn*. Then they tell us more about the evolution of AI with examples of ANNs (Artificial Neural Networks), integration of AI into robotics for socio-medical interests, including *developmental robotics* (combining developmental sciences, robotics, and cognitive science) and *neurorobotics* (artificial models of regions of the brain embodied into robots).

Chapter 9 introduces the idea of *collaboration*. Inspired by the *stigmergy* (communication of information via the environment) of insects, the computing world has integrated this mechanism to an area of research called swarm robotics, which controls groups of robots. The research attempts to find the cheapest ways to develop them as efficiently as possible, and to look for useful applications for these collaborative robots. The authors also point out an interesting distinction between *cooperative* and *coordinated robots*. The former refers to robots that act as a team, whereas the latter refer to robots that simply collaborate. Robots acting as a 'team' in most cases would not be considered "robots" based on the definitions given in Chapter 2, because they are teleoperated. This chapter shows the need for robots to collaborate and depend on humans to pursue new avenues of technological progress that neither robots nor humans would be able to achieve by themselves.

Chapter 10 is about *emotions*. One of the idealized conceptions of robots is that they could empathize or love, as well as feel rage or hate. The expectations of including emotions into robots are mixed. Some believe robots should remain *'rational'*, a rather Cartesian outlook into reason being *slave* of the *passions*, and that reasoning (being intelligent) has no business messing with its antithesis (emotions). Others defend the idea that emotions are what *makes us humans* and should remain ours. The authors demystify both positions and argue that none of them are in sync with how emotion and reason interact in humans and how it can be implemented in robots. They clarify what it means to *give robots emotions* considering current applicability and explain the limitations of modelling *affect*.

Chapter 11 goes into specific analyses of *social interactions* of robots as pets, butlers, and companions. A recurrent problem with social robots is their *novelty effect*, a surprising factor derived from interacting with something *unknown*, which makes us overlook faults and overestimate robots 'capacities. The real challenge comes after this effect vanishes, and robots must sustain long term interactions without becoming repetitive, boring, or ineffective for their purpose. Throughout the chapter we see examples of pet robots like Aibo or care robots like Paro the seal. Paro, designed for dementia sufferers, was able to successfully attend to the needs of *those patients*, by keeping the scope of interactions tailored to their main users (with features a nonimpaired person might dislike). Robovie 2 is another example. This is a patrolling robot used in a shopping centre in Japan that was abused and mistreated by children, showing another dimension of how our perception of a robot can change our social interactions for better *and* worse.

Chapter 12 is about the ability to *speak* and process *languages*. The first distinction we are introduced to is chatbots (dialogue systems) versus robots, inciting the question: how can it be possible to make robots *speak*? And if they speak, do they *mean* what they say? Language communicates meaning, however chatbots just manipulate forms and functions without further understanding of it. This might be, nonetheless, a useful feature for *some* robots, such as a robot guide in a museum or shopping centre. Other options include ASR (automatic speech recognition systems), but the difficulties to train natural language systems remains a difficult task. The authors exemplify cornerstone views on analysing language processing, including the *Turing test*, Searle's *Chinese Room*, and Wittgenstein *language games*.

Finally, Chapter 13 talks about ethics and society and promotes the idea that robots are merely *human artifacts*. The authors present the concept of *harm*, in relation to how we judge the moral capabilities of robots, and glance over some ethical guidelines and regulations about trustworthy artificial intelligence. Later, they contrast the *realistic expectations of autonomous robots* and the requirement of human judgement for decision

making derived from their use. They exemplify which areas might be more likely for robots to replace humans, often including tasks that are *impossible* or hardly achievable by humans on their own. However, the expectations given by its title "Ethics and Society: Could a robot have morals?" are not met. The ethical analysis in this chapter lacked depth and range.

By not integrating further aspects of the debate of robot rights they neglect their main argument against attributing "degrees of humanness to such machines that they really do not have" (pg. 219). The same can be said about the treatment given to the question: "could robots have morals?". They emphatically deny this possibility stating that "People have morals; and machines do not" (pg. 219-20). They claim that the idea that robots should have rights or morals shifts responsibility from the makers to the machine. They present the problem as a misattribution of morality, and do not really engage with the core of the question which is *can* they be moral? At no point we are told *why* machines cannot or could not have morals and the implications of this. How they discuss the possibility of artificial moral agents and machine ethics is lacking. They fail to benefit from a richer discussion to strengthen their argument and educate the reader as they do in previous chapters. Their critique is too simplistic and heavily (perhaps even unfairly) loaded against other perspectives that *do not* focus on or accept troublesome concepts like the singularity, but instead work directly on *moral outcomes*, like the case of autonomous vehicles as 'moral machines'.

Overall, this book is an excellent introduction to the understanding of the world of robots and their *actual* role in society. It clarifies that for a machine to be called a robot it *needs* the ability to react and interact with changes in their environment as a minimum requirement, after all they are '*embodied agents'*. It engages the reader with an easy yet robust explanation to understand the implementation and development of robotics. Fundamental notions are clearly defined and exemplified, serving as a well-researched introduction to engage with this field. As I said, however, some concepts such as singularity, transhumanism, and the ethics of moral robots deserved more attention. Despite these minor flaws, I recommend this to anyone interested in getting to know more about the details behind robot development, and to form truthful expectation of what they can do for us now, and what we might expect from them in the near future.