Developing robots: The need for an ethical framework

Alex Leveringhaus

Abstract
This article discusses the need for an ethical framework for emerging robotic technologies. The temptation, arguably driven by sci-fi treatments of artificial intelligence, is to ask whether future robots should be considered quasi-humans. This article argues that such sci-fi scenarios have little relevance for current technological developments in robotics, nor for ethical approaches to the subject: for the foreseeable future robots will merely be useful tools. In response to emerging robotic technologies, this article proposes an ethical framework that makes a commitment to human rights, human dignity and responsibility a central priority for those developing robots. At a policy level, this entails (1) assessing whether the use of particular robots would result in human rights violations and (2) creating adequate institutions through which human individuals can be held responsible for what robots do.

Keywords
Artificial intelligence, Robotisation, Ethics

Introduction
While the concept of a robot is hardly new, in recent years robots have received a lot of press. And that press has not always been positive. The reasons for this are evident from the following examples. Since software robots, though lacking nuts and bolts, can be used to steal sensitive information on the Internet, their widespread availability has the potential to usher us into an age of increased cyberwarfare, where non-state actors compete with, and pose a threat to states. In the meantime, on the physical battlefield of the future, weaponised ‘killer robots’ will mow down everything in their path (Human Rights Watch 2012). But robots will not only be thieves and killers; they will also be ‘lovers’, as robotic

Corresponding author:
A. Leveringhaus, Department of Politics, University of Surrey, Guildford, UK.
Email: a.c.leveringhaus@surrey.ac.uk

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (http://www.creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
sex dolls become ever more affordable and widely available. Beyond the confines of the battlefield and the bedroom, robots are predicted to find their way into the world of the office, where white-collar workers, having experienced stagnant wages since the financial crash in 2008, are anxiously wondering whether they will soon be landing on the metaphorical scrapheap. And, to put the icing on the cake, some academics and business figures worry that advances in artificial intelligence (AI) and robotics will not just replace white-collar workers, but will pose an existential threat to humanity itself (Cellan-Jones 2014).

Admittedly, much of the above is hyperbole designed to grab headlines. But this is not to say that there should not be some concern about the impact of robotics on our lives. There are, then, important choices to be made by businesses and policymakers with regard to the development, production and deployment of robots. Roughly, there are two preconditions for meaningful choice. The first consists of clarity about the subject matter: what actually is a robot? The second consists of identifying a framework that can provide content for the choices open to developers. Here, this author suggests that any sound ethical framework for robotics must centre on the protection of human rights, the preservation of individual dignity and a commitment to human responsibility.

**Being realistic about robots**

Let us begin by asking what a robot is. Put crudely, a robot is an artificial device that is designed to carry out useful tasks. Now, there are, of course, many artefacts that are designed for precisely that purpose: the humble spade makes gardening so much easier. What makes robots distinctive, however, is that they can carry out useful tasks without human oversight, once they have been programmed. In order to do so, robots have four central features:

1. *A sensor suite.* This allows the robot to perceive its environment.
2. *A body.* This enables the robot to move through and interact with its environment. For instance, it could use its arm to pick something up and put it somewhere else.
3. *A power source.* Like other electronic devices, robots cannot function without energy. They may have batteries, solar panels or be connected to the mains.
4. *Governing software.* Robots need to be programmed in order to function and carry out their tasks.

Two brief observations on the governing software are necessary. First, all too often robots are considered quasi-humans that will eventually gain consciousness and start to think for themselves. Whether a machine is capable of thinking is an interesting philosophical question. However, for the foreseeable future, such technological developments are not on the horizon. The delivery robot that will drop off your parcel from a large online retailer will not think about what it is doing. It will do it. Nor should one be overly
confident that an AI ‘super-intelligence’, with which humans could be locked in an evolutionary struggle for survival, will be developed.\(^1\) Certainly, it is possible to imagine such a super-intelligence *theoretically*. But in the immediate future, robots are likely to remain pre-programmed artefacts that carry out useful tasks for their owners.

Second, the governing software installed in some robots may include learning mechanisms. It is important to emphasise that human programmers will determine what the robot learns. An autonomous robotic vehicle, for instance, could learn how to optimise its driving performance, but only, say, on even surfaces and straight stretches of roads. The real significance of machine learning is not that learning robots will one day have had enough and turn on their human masters, but rather that the execution of their programmed tasks will become increasingly unpredictable. We know *what* a robot has been programmed to do. The question arising from machine learning is *how* it will carry out its assigned task. To illustrate the point, a programmer may programme a fully autonomous (robotic) unmanned aerial vehicle (UAV) to fly from A to B. We know that the UAV will arrive at B. But compared to a more conventional automated system, we will not be able to make exact predictions about its flight path.

**Why ethics is relevant to robotics**

What, then, are the potential frameworks that can provide content for what are, ultimately, human choices in the development of robots? Law, economics and politics all impact on the development, production and deployment of new technologies. Because of this, some may wonder what ethics could add to the mix. For one thing, there are numerous disagreements between ethicists on even the most basic of issues. For another, the law already provides a regulatory framework that has the benefit of being enforceable via the legal system. Despite this, there are two reasons why ethics should play a role in the debate on robotics.

First, some concerns about the impact of certain technologies are not reflected in the law. The use of semi-autonomous (or remote-controlled) robots, such as weaponised UAVs in armed conflict, serves as a good example in this respect.\(^2\) The deployment of UAVs during an armed conflict is perfectly legal under international humanitarian law. Yet many people feel uneasy about the ability of drone operators to kill enemy combatants over great distances without incurring any physical risks themselves. International humanitarian law does little to articulate, or respond to, these concerns. Ethics, by contrast, does. Similarly, it is not unreasonable to assume that people will also have ethical concerns about the development of robots for civilian purposes.

Second, ethical objections can sometimes have an impact on reform of the law. It is not entirely inconceivable that the legal regulation surrounding the development of certain robots could be tightened in response to moral concerns. Conversely, if there are strong ethical reasons in favour of the development of certain robots, the relevant legal regulation could be relaxed.
How not to do ethics

Now, it is one matter to acknowledge the relevance of ethics for robotics; it is quite another to articulate one’s ethical concerns. There are two extreme approaches best avoided. The first, which opposes the development of certain robots, consists of simply shouting ‘yuck!’ Often, however, ‘yuck!’ merely reflects prejudice. As robots become increasingly visible in the private and public domains, it is not clear whether those who shout ‘yuck!’ now would do so in 10 years’ time. Some may find the idea of a robot carrying out open heart surgery disturbing, but this does not mean that there is anything ethically wrong with developing robots for this very purpose. In 50 years, people may shake their heads in disbelief that their grandparents objected to robotic surgeons. The ‘yuck factor’ is thus a notoriously unreliable guide to ethics.

The second approach, which seeks to bolster the ethical case in favour of the development of certain robots, involves highly idealised assumptions. Some roboticists, for instance, argue that the occurrence of war crimes could be minimised by creating killer robots (Arkin 2010). The reasoning behind this argument is that, unlike soldiers, robots cannot lose control over their emotions and commit war crimes as a result. They may short circuit, but that is about it. As a result, the development of killer robots could lead to the adoption of more restrictive targeting standards by the military. This seems reasonable. The question remains, however, as to whether existing killer robots could be abused to perpetrate war crimes due to rogue programming or hacking. Moreover, once potential opponents in armed conflict have caught up technologically by creating effective anti-robot weapons, the military could, in response, lower its targeting standards. This may not necessarily debunk the ethical case in favour of killer robots. It shows, though, that one should not have unrealistic expectations of robotic weapons technology.

Similarly, when considering the impact of robotics on employment rates among white-collar workers, it is fanciful to assume that potential job losses in existing white-collar positions would be offset by job gains in the technology sector (Frey et al. 2016). Perhaps accountants, investment bankers and legal researchers could be retrained as AI programmers. But if experiences with automatisation, de-industrialisation and outsourcing in Western economies are anything to go by, it is unlikely that all white-collar workers who have been displaced by robots are going to find satisfactory alternative employment. Some may be reabsorbed into the economy but encounter worse pay and conditions compared to their previous jobs. Others may not be so lucky and might find themselves consigned to long-term unemployment. Again, this does not necessarily debunk the case for the introduction of robots into white-collar environments, but their adverse effects on employment must surely be part of any credible ethical assessment of contemporary robotics.

Rights, responsibility and dignity

But how can one make an ethical assessment of robots in the first place? There is no straightforward answer to this question, for two reasons. First, it is highly unlikely that
there is a one-size-fits-all ethical solution that is applicable to all areas of robotics. The ethical issues thrown up by military robotics are not necessarily the same as those arising from healthcare robots: robotic weapons will be designed to harm; healthcare robots will be designed not to harm but to help. Second, given that contemporary ethics is characterised by vigorous disagreement, it is unlikely that there is a single authoritative approach to resolving the critical issues arising from the development of robots. However, there is one approach in ethics that, this author thinks, is well suited to dealing with the challenges posed by robotics. It assumes that

- the ability of human individuals to lead independent lives must be respected,
- human individuals are holders of human rights that limit what can permissibly be done to them, and
- human beings are endowed with a dignity that forbids treating them as mere means (things) or de-humanising them in other ways (Nozick 1974; Rawls 1999; Kant 2012).

The repercussions of this approach are threefold. First, the protection of human rights needs to be central to the development of robots. Does the development and eventual deployment of a specific robot enhance or undermine human rights? If, on the one hand, the use of a particular robot enhances protection for human rights, this would count as a strong reason in favour of developing it. Imagine that the deployment of robotic weapons would indeed significantly reduce the occurrence of war crimes. This would count as a potential argument in favour of military robotics. If, on the other hand, the use of a specific robot undermines respect for human rights, this would count as a strong argument against its development. For instance, suppose that a software robot was specifically designed to steal personal information, thus violating an individual’s right to privacy. The development of such a robot would be immoral.

Second, the idea that human individuals have rights and should be capable of living independent lives highlights the importance of responsibility and accountability. More precisely, individuals and institutions should be held responsible for their actions when these affect others. Those affected deserve an explanation of what has happened to them and why. Those who are responsible must offer such an explanation, and should be punished for any harm they have caused, if necessary. The use of learning robots, in particular, has the potential to undermine a commitment to responsibility. Robots, it goes without saying, are not moral agents and thus cannot be held responsible for what they do. That said, it may not always be fair, or indeed possible, to hold a robot’s programmer or owner responsible for what the machine does. Given the unpredictability inherent in learning mechanisms, neither the programmer nor the owner may reasonably be able to foresee the ways in which their robot, once deployed, might behave. Who is responsible in such situations? In response to this problem, ethics demands that adequate structures for the assignment of responsibility are created. A commitment to preserving human responsibility for robotic behaviour must, therefore, be central to the development of all robots.
Third, the type of ethics that the author is proposing here views treating persons as mere things as a violation of their dignity. To avoid this, the development of robots needs to be assessed in relation to whether it enhances or undermines the dignity of human persons. On the one hand, the development of certain robots could prevent the treatment of individuals as mere things. Some tasks, as roboticists like to remind us, are so dull, dirty and dangerous that it is preferable to let them be carried out by a robot. Consider a train that connects airport terminals. Suppose that it only goes back and forth between two terminals and stops every two minutes. It is hard to see how using a human driver to drive the train for eight hours a day does not treat that unlucky person as a thing. The task is so mind-numbingly dull and repetitive that it would drive any person mad. Or imagine clean-up operations after nuclear accidents. Can we really say that we treat human persons with the respect and dignity they deserve if we knowingly expose them to lethally high dosages of radiation? In both instances, there is a strong case in favour of using robots rather than humans.

On the other hand, there can clearly be instances where the use of robots undermines the dignity of individuals. The use of robots in social care, for instance, has the potential to dehumanise the elderly, whose immediate medical needs may have been met but whose emotional needs are being neglected. Interestingly, care robots not only have the potential to dehumanise recipients of care but also the human providers of care. The development of care robots suggests that the provision of care could be done by a machine. This fails to acknowledge the emotional investment that many carers have in their work, as well as the importance of an appropriate professional ethos that sustains a humane care environment. For some, after all, care work is a vocation, rather than merely a job. Care, then, is not a simple, automated process. And those (humans) who provide it are not automatons that can easily be replaced.

Conclusion

Even if one remains realistic about the kinds of robots whose development is technologically feasible in the foreseeable future, a more detailed case-by-case ethical assessment of specific robotic technologies is necessary: there is no one-size-fits-all ethical solution that is applicable to all types of robots. More concretely, at a policy level, such an ethical assessment should, following the approach outlined in this paper, be two-fold. First, it should be assessed whether the development, production and use of a specific type of robot potentially leads to human rights violations. Second, policymakers need to develop adequate structures and institutions through which users and developers of robotic technologies can be held responsible for what their machines do. These two steps should form the basis of any ethically informed policy approach to emerging robotic technologies.

Notes

1 The notion of super-intelligence was developed by the philosopher Nick Bostrom (2014) in *Superintelligence*. A more critical perspective is provided by Margaret A. Boden (2016) in *AI: Its Nature and Future*. 
To be clear, not all UAVs are necessarily weaponised: they can be used for civilian purposes. For a longer version of this argument, see Leveringhaus (2016).

References

Arkin, R. C. (2010). The case for ethical autonomy in unmanned systems. *Journal of Military Ethics*, 9(4), 332–41.

Boden, M. A. (2016). *AI: Its nature and future*. Oxford: Oxford University Press.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press.

Cellan-Jones, R. (2014). Stephen Hawking warns artificial intelligence could end mankind. *BBC News*, 2 December. http://www.bbc.co.uk/news/technology-30290540. Accessed 5 December 2017.

Frey, C. B., Osborne, M. A., Holmes, C., Rahbari, E., Garlick, R., Friedlander, G., McDonald, G., Curmi, E., Chua, J., Chalif, P., & Wilkie, M. (2016). *Technology at work v2.0: The future is not what it used to be*. Citi Group and Oxford Martin School, 27 January. https://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work_2.pdf. Accessed 15 January 2018.

Human Rights Watch. (2012). *Losing humanity: The case against killer robots*. 19 November. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots. Accessed 15 January 2018.

Kant, I. (2012). *Groundwork on the metaphysics of morals*. Cambridge: Cambridge University Press.

Leveringhaus, A. (2016). *Ethics and autonomous weapons*. Basingstoke: Palgrave Macmillan.

Nozick, R. (1974). *Anarchy, state and Utopia*. New York: Basic Books.

Penrose, R. (1989). *The emperor’s new mind*. Oxford: Oxford University Press.

Rawls, J. (1999). *A theory of justice*. Oxford: Oxford University Press.

Author biography

Alex Leveringhaus is a lecturer in the Department of Politics, University of Surrey, where he teaches political philosophy. He is also the author of the book Ethics and Autonomous Weapons (*Palgrave Macmillan, 2016*).