

Chapter 3

Does AI Raise Any Distinctive Ethical Questions?

Abstract In developing codes of ethics for AI, it's important to consider how the ethical issues concerning AI relate to other ethical issues. Some of the questions overlap with those relevant to other rapidly developing technologies, especially where those technologies involve questions about how we humans understand ourselves and our place in the world. Issues that concern both AI and other areas of science and technology include the difficulty of considering ethical issues where predictions about the future may be difficult to make; where technology is embedded in complex ways in society, the economy and culture; where the analysis and use of large amounts of data are at issue; and where complex, large systems give rise to difficult questions about the distribution and attribution of responsibility. Characteristic ethical questions regarding AI concern its typical enhancement or replacement of human agency; crucially, questions of agency are at the heart of how we understand ethics. AI, like other technologies, is often hyped up; it's important to see through the hype to as realistic a picture as possible. Hype can distort our ethical thinking in dangerous ways, and some of the key hazards of hype regarding ethics are outlined.

3.1 Methodology: Focusing in on Ethical Questions

The ethical questions of AI are sometimes presented as if they are new and unique. But it's clear that some of the questions that AI raises are common to other concerns, such as other developing technologies and changes to economic practices. Understanding the links with other issues is important methodologically, for this can alert us to common ways of thinking that can be drawn on in moving the debate forward. But it's also important to consider if there are any distinctive ethical questions in AI, since if we are faced simply with a list of very broad ethical issues which are not focused upon the specifics of AI, we may end up with codes and principles which lack sufficient grip on concrete circumstances to achieve substantive impact. There will always be arguments about whether certain features are unique to AI. So, here, I consider if there are any features which are characteristic of AI, and which have significant ethical import.

3.1.1 *How Do We Identify Ethical Problems as New?*

Could certain applications of AI render more visible to the clear light of day, ethical issues which had been there all along? Perhaps this might be by magnifying the speed and efficiency with which certain tasks are done. Perhaps this might be by the reach and power of its wider effects e.g. on social organisation and on economics. Perhaps it might be because AI may articulate step by step what is needed to simulate human reasoning, and hence in building AI, we might become more acutely aware of the capacities and limits of our own reasoning, and perhaps more acutely aware of how we think about issues with moral implications. For all such reasons, and possibly more, AI might discover, as well as perhaps create, ethical issues for us to ponder.

Regulating Medical Robots: Is this an Issue for AI, or an Issue for Medicine?

The European Parliament's report on European Civil Law on Robotics (European Civil Law Rules in Robotics 2016) discusses the question of how precisely robots are defined. Legislation needs clear definitions. And robots come in a variety of forms. Robots may not involve AI as such.

One issue is whether we treat robots which don't have autonomy differently from those that do. The report points out that medical robots fall into the 'master/slave' category of robots, since they remain under the control of the surgeon or surgical team. 'Nevertheless, the European Union absolutely must consider surgical robots, particularly as regards robot safety and surgeon training in robot use.' (p. 9)

But note that practice of medicine, surgeon training, and medical devices are already subject to close scrutiny and regulation. In addition to law and professional regulation, there are also many active patient interest groups who concern themselves with such matters. Medical robots rightly deserve ethical and regulatory attention, but there is no particular reason—certainly not one to be found in this report—to suppose that existing systems for scrutiny and response in medicine are not adequate to the job. The absence of autonomy in these 'master/slave' robots is relevant to reaching such a conclusion.

3.2 **Many Ethical Issues in AI Are Shared with Other Rapidly Developing Technology**

Here are just a few.

Problems of prediction: The very fact that AI is a rapidly developing technology means that it's hard to predict what will occur even in the near future. As well, it is hard to anticipate ethical questions in advance and to produce codes and regulation as fast as the AI develops. But this is the case in other areas of science and technology as well.

Interface with social and cultural issues is an issue in all technology and AI is no exception. In considering ethical issues we need to consider how there may be subtle, yet pervasive, social, cultural, and economic impacts from the use of AI.

The manipulation of data: Some issues, especially in certain areas of AI, concern the use and manipulation of masses of data, and this issue arises in many other areas too. There has been work for some time now on ethics of data management, for example, in genomics and in other medical research.

Complex systems and responsibility: AI often involves highly complex systems, with those working on different elements unknown to each other, and the effects of the operation of AI may be some distance from producers, for instance when machine learning produces algorithms which may affect millions of others far distant to the source. Questions about how responsibility works in such large networks of researchers is already receiving attention (Boddington 2011).

Nonetheless, there are some characteristic ethical challenges in AI.

3.3 Ethical Questions Arise from AI's Typical Use to Enhance, Supplement, or Replace the Work of Humans

It is difficult to give a precise characterisation of how the extension of human agency by AI differs from that offered by simpler automation. What is characteristic of AI is not just that it extends or enhances human agency; nor that it extends or enhances human reasoning, for something as simple as an abacus does this. AI characteristically enhances or replaces human decision-making and human judgement. It may enhance or replace human action and/or human perception, and, may attempt to simulate human emotions.

The extension, enhancement and replacement of human agency and reasoning in AI serve as the loci of many of the ethical issues that arise in its use, sometimes presenting us with vivid versions of old questions. Note that these will vary depending on the precise application and context, and on how far-reaching are the effects of the AI. In some instances, AI does what humans could do, if they had enough time; in other instances, it seems to enhance human agency and reasoning to provide calculations and perform actions that we simply could not do unaided. Some recurring issues arise.

AI and transparency: Some applications of AI give us answers that are in principle hard to check, and where it's hard to know how the answer was even reached, as when answers are produced by machine learning. There's currently debate about such questions, about whether, for example, we'll ever be able to train machines to explain their otherwise opaque decisions to us.

AI and the control problem: The more powerful AI is, as a general rule, the less we are going to be able to control it. There's no easy way to characterise this, and it's always going to be necessary to look in detail at specific cases. In some

instances, autonomy might give more control—an autonomous missile can be programmed with a more precise goal than a bullet which once released, is out of our control. But what tends to be characteristic of AI is autonomy of judgement, of decision, of action; and as we have seen, these are key concepts driving accounts of morality at a very deep level.

This means also that work in AI itself is directly tackling issues which are also of central concern to ethics. Artificial Intelligence is perhaps unique among engineering subjects in that it has raised very basic questions about the nature of computing, perception, reasoning, learning, language, action, interaction, consciousness, humankind, life etc. etc.—and at the same time it has contributed substantially to answering these questions (in fact, it is sometimes seen as a form of empirical research) (Müller 2012).

AI then tends to give rise to questions that we see of particular significance to ethics, and to our values more broadly. As we've seen, at the heart of any account of ethics is an account of moral agents, their powers and their limitations. And AI is precisely about the extension of our powers as agents. One issue of particular interest is how in thinking about this, we have a tendency to idealise both the agency of humans, and the agency of machines, and this will be discussed further in Chap. 7.

Robot Camel Jockeys: 'Pimp My Ethics'

Robots were often employed to replace child camel jockeys in various Gulf states where camel racing is a popular sport. The use of child camel jockeys was highly problematic: there is considerable evidence that very young children were taken from countries such as Bangladesh, Pakistan and Sudan, deprived of food to keep their weight down, treated very harshly, and subject to the frightening and dangerous task of riding camels, all for the sake of a sport. This practice was also against laws on child labour, yet continued despite these laws. There are widespread reports of sexual abuse as well (Gluckman 1992; Lillie 2013; Pejman 2005).

Firms created small and simple robots to replace the child jockeys. This has been hailed as a success; they have even been described as 'the robots that fight for human rights' (Brook 2015) and that 'there are some issues that really can be solved with innovation and technology' (Rasnai 2013). Although these robots are so simple that they scarcely count as AI, it seems then that this is an instance of robotics put to great moral benefit. There have been positive write-ups in various tech magazines, and Vogue has even produced a glamorous photoshoot of camel racing with robot jockeys (Shaheen 2017).

But this case also serves as an example of how attention to even this simple technology can divert attention from other issues.

Hailed as an ethical 'win-win', reading further down some of the tech magazine articles on this topic shows that the situation is more complex, and

(continued)

further research demonstrates this too. There are difficulties in repatriation of the boys (Pejman 2005), and one organisation estimates that 3000 of these boys have simply vanished (Ansar Burney Trust 2013). It's not even entirely clear if the practice of using young boys in appalling conditions has been completely eradicated (Peachey 2010). So, has the rush to praise the use of tech to solve a moral issue been premature?

There are moral pluses and minuses of the allure of technology. Note that one reporter describes how the robots were adopted because they were 'cool': 'the message is clear: pimp my camel' (Schmundt 2005). If we take this analysis at face value, it means that the very coolness and allure of technology served a good moral purpose, of helping to stop the use of boy jockeys where laws requiring this had not. That seems good, doesn't it?

But note then too, that if the camel racers and owners merely used the robots instead of the boys because they thought they were 'cool', this does not suggest that they were motivated by a moral awareness of the devastating impact on the lives of the child jockeys.

3.4 We Also Need to Consider the Methods of Production of AI

The development of technologies in AI gives increasing possibility for a few individuals, corporations or research groups to make significant advances with scant oversight, notwithstanding the current efforts of many to consider the ethical and policy issues involved. Work can be done by those working entirely outside the framework of any professional accreditation.

And there is an increasingly steep gradient of wealth and resources for those accumulating capital in IT in general including AI (Piketty et al. 2014). Again, this is not unique to AI, but is characteristic of much IT in general, and the wealth, and technological reach, of some of the big players in AI is so large that realistically, this is a major feature of the world economic and, indeed, political landscape. All these issues present challenges in considering the development of codes of ethics for AI, especially when coupled with the breadth and power of AI to affect the lives of people globally.

3.5 Hype in AI and Implications for Methodology in Ethics

It is easy to be dazzled by the myriad claims made about AI and the scent of hyperbole that often surrounds it. Hype can concern both the technology itself and the ethical issues it may raise. The hyping of technology and its potential dangers

and benefits tends to occur particularly when notions of speed, rapid technological change, and money, are present; for example, the presence of hype in genomics research is well established (Caulfield and Condit 2012). The framing of a technology as having a fundamental impact upon humanity and on our self-image, as has also happened with genomic technology, is also at play in many discussions of AI (Joy 2000); indeed, such claims may be true, making it all the more important that they are examined closely.

So, is the current attention to the ethical issues in AI a hyped ‘moral panic’?

Work in the analysis of rhetoric and language demonstrates how the framing of an issue, and precisely how its ethical problems are described, impacts upon our responses (Fischer and Forrester 1993; Majone 1989; Throgmorton 1993). Problems will be made sense of and acted upon through descriptions which frame the discussion, highlighting certain aspects, and making certain contrasts (Rein and Schon 1991; Hajer 1993).

Consider the biotechnology industry which has also been the target of multiple moral attacks (Newton 2001); these are not necessarily grounded in realistic assessments, with hard-to-explain regional differences in publicly-voiced concerns (Ford et al. 2015). Hype may be used, sometimes knowingly, to fulfil a particular purpose. For instance, accounts of synthetic biology lurch between hyping up how new it is, typically in the context of attracting investors, and insisting that it is a well-established field only doing what has been done for millennia, thereby downplaying the need for specific regulation (Parens et al. 2009).

Likewise, we are warned about the moral dangers of AI, and also the moral dangers of clamping down on progress in this technology. Hype about future possibilities can distract us from present realities (Crawford and Calo 2016). For instance, hype about a possible malevolent superintelligence being developed some time in the future could distract from the need to consider ways in which AI is already affecting our everyday lives now. Any work in the ethics of AI then needs to think carefully through the hype and the reality; while taking into account that the very uncertainty of the field and of how it might develop makes it hard to work out what is hype and what is realistic.

3.5.1 Hype Can Both Distort Our Ethical Reasoning, and Reveal Things of Potential Interest

Questions in ethics require close attention to all aspects of a situation that are morally relevant, and must weigh these against each other, a difficult task requiring close attention. We need to be aware of how our attention is being commanded, and consider if we are being drawn to something of prime relevance, or if we are being blindsided. There is no easy answer to this, and it’s one of the reasons why ethical debate and dialogue with others is vital (Bazerman and Tenbrunsel 2011).

For example, let us examine a claim in a widely-read article that states issues involving AI present ‘the key question for humanity today’ and that artificial intelligence holds unparalleled promise because ‘everything that human civilisation has to offer is a product of human intelligence’ (Hawking et al. 2014). The last remark is a loose and inflated view of the power of human intelligence. Note the ambiguity of ‘product of human intelligence’. This could mean merely that human intelligence had something to do with all the good things of civilised life. But it seduces us to consider that ‘everything’ of value in civilisation is the designed result of the application of intelligence, and it may lead to hubris, if we consider that nothing is due to serendipity or other factors.

Moreover, human beings are not simply characterised by intelligence but by sociability. It’s common in the context of AI to read casual remarks about how intelligence is what has brought humans so far, such as this comment from Guruduth Banavar: ‘From the evolutionary point of view, humans have reached their current level of power and control over the world because of intelligence. . . AI is augmented intelligence’ (Conn 2017a). Such remarks often slide into claims that it will be, therefore, artificial *intelligence* which leads us forward to the future (Kurzweil 2001). However, although some claim that it is human intelligence which has formed the impetus behind our evolution, other evolutionary scientists beg to differ: our sociability, pair-bonding, and religion, are also prime candidates to explain our extraordinary success (Barrett et al. 2002; Zauzmer 2017). And with regard to developments in AI, some sensible humility about the impacts of human ‘intelligence’ on civilisation, and a measured appraisal of our abilities to control AI, is precisely what’s at issue. Yet the phrase seduces us into a world view that breathes the promise of human control over what we do. Such tunnelling of thinking towards intelligence alone can be characteristic of much talk about AI, and can have a real impact on how ethical issues are discussed.

3.5.2 Hype About AI Can Channel Our Thinking About Solutions

A focus on the dangers of AI, coupled with optimism about its potential, can lead to an overreliance on AI as a solution to our ethical concerns—to an ‘AI on AI’ approach. Problems are seen as solely technological, hence as requiring technological solutions. But creative thinking can often produce a wider range of solutions to problems. The common view that ‘technical cause = technical solution’ is not necessarily valid, just as it is not necessarily valid in other areas such as medicine or psychology. Bear such points in mind when the question of how we assess the harms and benefits of something as complex and as deeply embedded in our life as AI.

Critiquing the Core Values of a Profession

The practice of modern medicine has saved and extended many lives, and improved the health of many (along with other developments such as plumbing, electricity, widening education, and the mechanisation of agriculture), so it's easy to take its value for granted. A set of professional regulations is unlikely to critique the core activities of the profession as such. Such critique tends to come from those outside the profession, such as groups representing public concern, critical social scientists, or mavericks within the profession. Yet such critiques can be very helpful. Often, giving a name to concepts helps move debate forward.

Healthism is the notion that health is the supreme value that should be pursued above all others; this can be questioned (Fitzpatrick 2001; Crawford 1980).

The concept of medicalisation has been influential in debates about the benefits of medicine. '*Medicalisation*' is the process by which phenomena are analysed in medical terms, with solutions sought in medical terms, where this represents a narrowing of possible options (Illich 1976). For example, to treat 'low mood' as only a medical issue, and with a pharmaceutical response only, may in some cases overlook the root causes of an individual's problem and may mask helpful solutions.

Geneticisation refers to the same phenomenon, applied to genetic explanations of conditions (Lippman 1991), where this blinds us to other aspects of the causation and treatment of disease and health.

Likewise, *iatrogenesis* names the phenomenon whereby the practice of medicine actually causes problems, which are then treated with more medicine (Illich 1976).

Virtually everyone is against eugenics. But many practices in modern medicine may be described as *liberal eugenics*, where the state does not mandate the practice, but actions by individuals under the banner of individual reproductive choice occur which change what kind of children are born. Some hail this; others note the strong arm of the state may be replaced by other pressures implicit in much modern medicine. Either way, the label 'liberal eugenics' is useful for catalysing debate (Agar 2004).

Are there equivalents of medicalisation and iatrogenesis happening in relation to AI? *Artificialintelligenciation* is a bit of a mouthful which I freely admit may not make it into the Oxford English Dictionary. But perhaps we ought nonetheless to look out for it. And perhaps codes of ethics ought explicitly to consider this too.

And where AI concerns the 'enhancement' of humans in various ways, the questions raised about how our thinking may be tunnelled concerning the benefits of medicine may be of direct relevance.

3.5.3 *Impacts of Hype on Moral Thinking*

Here are some of the ways that hype can impact how we think:

Distortions in thinking: On occasion, discussion of AI can simplify, exaggerate and idealise what it is to be human, and what it is to be an intelligent machine. Hype may not simply and harmlessly draw attention to the ethical issues, but may distort the very questions that we need to address. Hype can shape the (dis)appearance of humanity in how we perceive technology, the relationship between humanity and machines, and the attenuation of our visions of humanity. We will discuss the idealisation of agency in thinking about AI and ethics in Chap. 7.

Hype, ethical dangers, and public image: One important potential impact of hype is that fear of being branded one of the ‘bad guys’ may lead to individual or collective attempts to promote oneself as on the side of the angels. This might be for intrinsic or explicitly strategic reasons. There are serious dangers for institutions which fall for this trap.

At its worst, appearing to be ethical might trump concerns about being ethical. Hying the moral dangers of AI might produce what has come to be known as ‘virtue signalling’ (Bartholmew 2015), where an individual or organisation proudly proclaims their ethical credentials in ways which acts as a substitute for actual action (Hogben 2009). Content then might perhaps be simply empty displays of ethical probity (‘we are passionate about the future of the human race’, and so on).

Hype may have a tangible impact upon the content of codes. The rush to go address prominent ethical issues may lead to foregrounding one moral value, possibly at the expense of others; the ethics is set to 11, to paraphrase the film *Spinal Tap*; this has been dubbed ‘ethical enthusiasm’ (Hammersley 2009). For example, concern for data privacy might create conditions that make valuable research onerous or even impossible.

Emphasising one moral value without attention to balance has particular hazards. Aristotle long ago argued that at the extremes, virtues turn to vice (Aristotle 1999). Research in moral psychology and personality traits corroborates this; personality traits spread out along a normal distribution, which indicates that there are disadvantages of inhabiting the extreme ends of the distribution. For instance, the trait of empathy at its extreme can lead to overlooking hazards to self and other in concern for those who are the object of that empathy, and aggression towards any perceived threat to the object of empathy; yet aggression in some circumstances is very valuable. Values need to be balanced with each other, against the realities of a particular context. Hying up the importance of one value, to the exclusion of others may occur when panicking about the impacts of AI.

Emphasising One Ethical Value: Excluding the Less Well Resourced

Policies designed to produce good ethical practice are sometimes easier to comply with, if you are well resourced. For example, the data sharing policies of many scientific research institutions include clauses that data sharing should be encouraged, by sharing data only with those who enter into reciprocal agreements—who are dubbed ‘good data sharers’. But, those from less well-resourced institutions may have less to share, or may lack resources to guarantee certain standards of data curation. The least privileged researchers may end up with the worst deal, unless some special provision is given (Boddington 2012).

But then, focusing on one highlighted value may indeed be what an organisation, on reflection, decides to do.

Here’s an example from AI. The Icelandic Institute for Intelligent Machines (IIIM) Ethics Policy for Peaceful R&D eschews military funding, and has rules about collaboration based on this:

2.3 IIIM will not collaborate with any institution, company, group, or organization whose existence or operation is explicitly, whether in part or in whole, sponsored by military funding as described in 2.2 or controlled by military authorities. For civilian institutions with a history of undertaking military-funded projects a 5–15 rule will be applied: if for the past 5 years 15% or more of their projects were sponsored by such funds, they will not be considered as IIIM collaborators (IIIM 2015).

Researchers who are not in the position of finding alternative sources of funds will therefore be excluded from potentially beneficial collaboration.

This is not a commentary on the rights or wrongs of the IIIM’s stance; the IIIM is perfectly free to set their own policies. These issues might however be a central concern for government agencies, or for international professional bodies, who probably would not wish to disadvantage any members, least of all those already struggling with resource access.

3.6 Conclusion

I have argued that thinking about AI will require us to think clearly and deeply about some fundamental questions in ethics. AI and ethics are intimately concerned with fundamental questions of agency. Various questions in ethics concern how we relate to friend and stranger, to the world at large, and how we treat ourselves. AI promises to make changes to all of these dimensions of our lives. And we need to consider how far AI has common concerns with other areas, and how far it raises distinctive questions, while being careful not to fall for the distorting effects of hype upon our moral thinking.

Our task then is this: we need to develop codes of ethics in a situation of uncertainty, derived not just from the rapid development of technology, and not

just from the diverse views about this technology, but from the rapid technology which is changing potentially how we relate to others, to the world around us, and our sense of self—in other words, impacting upon the very foundational basis of any ethic.

But here, we are not looking in general at the ethics of AI. We are concerned with developing codes of ethics for AI. I turn now to consider the essential elements of professional codes of ethics, before discussing how AI challenges these elements.