Children of the Fourth Revolution

Luciano Floridi

1 Introduction: the Two Souls of AI

It is a well-known fact that artificial intelligence (AI) research seeks both to *reproduce* the outcome of our intelligent behaviour by non-biological means, and to *produce* the non-biological equivalent of our intelligence. As a branch of engineering interested in *intelligent behaviour reproduction*, AI has been astoundingly successful. We increasingly rely on AI-related applications (smart technologies) to perform tasks that would be simply impossible by un-aided or un-augmented human intelligence. But as a *branch of cognitive science interested in intelligence production*, AI has been a dismal disappointment. Productive AI does not merely underperform with respect to human intelligence; it has not joined the competition yet. The fact that Watson—IBM's system capable of answering questions asked in natural language—recently won against its human opponents when playing *Jeopardy!* only shows that artefacts can be smart without being intelligent.

The two souls of AI, the engineering and the cognitive one, have often engaged in fratricidal feuds for intellectual predominance, academic power, and financial resources. That is partly because they both claim common ancestors and a single intellectual inheritance: a founding event, the Dartmouth Summer Research Conference on Artificial Intelligence in 1956, and a founding father, Turing, with his machine and its computational limits, and then his famous test. It hardly helps that a simulation might be used in order to check both whether the simulated *source* has been produced, and whether the targeted source's *behaviour* or *performance* has been reproduced or even surpassed. The misalignment of their goals and results has caused endless and mostly pointless diatribes. Defenders of AI point to the strong results of reproductive, engineering AI, whereas detractors of AI point to the weak results of productive, cognitive AI. Many of the current speculations on the so-called singularity issue have their roots in such confusion.

L. Floridi (✉)
Department of Philosophy, University of Hertfordshire, Hatfield, Hertfordshire, UK
e-mail: l.floridi@herts.ac.uk

In order to escape the dichotomy just outlined, one needs to realise that AI cannot be reduced to a "science of nature", or to a "science of culture", because it is a "science of the artificial", to put it with Herbert Simon (Simon 1996). As such, AI pursues neither a *descriptive* nor a *prescriptive* approach to the world: it investigates the constraining conditions that make possible to build and embed artefacts in the world and interact with it successfully. In other words, it *inscribes* the world, for such artefacts are new logico-mathematical pieces of code, that is, new texts, written in Galileo's mathematical book of nature.

Until recently, the widespread impression was that such process of adding to the mathematical book of nature (inscription) required the feasibility of productive, cognitive AI. After all, developing even a rudimentary form of non-biological intelligence may seem to be not only the best but perhaps the only way to implement technologies sufficiently adaptive and flexible to deal effectively with a complex, ever-changing and often unpredictable, when not unfriendly, environment. Such impression is not incorrect, but it is distracting because, while we were unsuccessfully pursuing the inscription of productive AI into the world, we were actually re-ontologising the world to fit reproductive, engineering AI. The world is becoming an infosphere increasingly well adapted to AI-bounded capacities. In robotics, an *envelope* is the three-dimensional space that defines the boundaries that a robot can reach. We have been enveloping the world for decades without fully realising it.

2 Enveloping the World

Enveloping used to be either a stand-alone phenomenon (you buy the robot with the required envelop, like a dishwasher or a washing machine) or implemented within the walls of industrial buildings, carefully tailored around their artificial inhabitants. Nowadays, enveloping the environment into an AI-friendly infosphere has started pervading any aspect of reality and is visible everywhere, on a daily basis. 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 reproductive AI and its limited capacities.

Enveloping is a trend that is robust, cumulative and progressively refining. It has nothing to do with some sci-fi singularity, for it is not based on some unrealistic (as far as our current and foreseeable understanding of AI and computing is concerned) speculations about some super AI taking over the world in the near future. But it is a process that raises the risk that our technologies might shape our physical and conceptual environments and constrain us to adjust to them because that is the best, or sometimes the only, way to make things work.

By becoming more critically aware of the re-ontologising power of reproductive AI and smart applications, we might be able to avoid the worst forms of distortion (rejection), or at least be consciously tolerant of them (acceptance), especially when it does not matter or when this is a temporary solution, while waiting for a better design. In the latter case, being able to imagine what the future will be like, and what adaptive demands technologies will place on their human users, may help to devise technological solutions that can lower their anthropological costs. In short, human

intelligent design (pun intended) should play a major role in shaping the future of our interactions with forthcoming technological artefacts and the environments we share with them. After all, it is a sign of intelligence to make stupidity work for you.

This special issue, guest-edited by John Sullins, intends to be a contribution to our deeper understanding of the conceptual issues raised by robotic technology, in a world that is going to be increasingly populated by artificial agents. It includes papers on both the military and the civilian uses of robots because, technologically, war and peace are never so far from each other: iRobot produces both the *Roomba 700* that hoovers your floor and the *iRobot 710 Warrior* that disposes of your enemies' explosives. In the rest of this editorial, I shall briefly comment on each side of the phenomenon.

3 Cyberwar

Cyberwar is the continuation of conflict by digital means. It is a new phenomenon, which has caught us by surprise. With insight, we should have known better, for several reasons.

Take the nature of our society first. When it was described as industrial, conflicts had mechanised features: engines, from battleships to tanks to aeroplanes, were weapons, and the coherent outcome was the emphasis on petrol and nuclear power. There was an eerie analogy between assembly lines and warfare trenches. Conventional warfare was kinetic warfare, we just did not know it, because the non-kinetic kind was not yet available. The Cold War and the emergence of asymmetric conflicts were part of a post-industrial transformation. Today, in a culture in which the word "engine" is more likely to be preceded by the verb "search" than by the noun "petrol", advanced information societies are as likely to fight with bytes as they are with bullets, with computers as well as guns, not least because digital systems tend to be in charge of analogue weapons. I am not referring to the use of intelligence, espionage, or cryptography, but to cyber attacks of the kind witnessed in Estonia (2007) or in Iran (Stuxnet worm, 2010), or to the extensive use of drones and other military robots in Iraq and Afghanistan. Robotic weapons may be seen as the final stage in the industrialisation of warfare, or, more interestingly, as the first step in the development of information conflicts, in which command and control as well as action and reaction become tele-concepts. Either way, from robots in physical environments to software agents in cyberspace, we should not be too optimistic about the non-violent nature of cyberwar. The more we rely on ICT (information and communication technologies), the more cyber attacks will become lethal. Soon, crippling an enemy's communication and information infrastructure will be like zapping its pacemaker rather than hacking its mobile.

Consider next the nature of our environment. We have been talking about cyberspace for decades. We could have easily imagined, even without reference to science fiction, that this would become the new frontier for human conflicts. We have been fighting each other on land, at sea, in the air, and in space. Predictably, the infosphere—as I prefer to call it, given the constant erosion of the divide between online and offline—was never going to be an exception. Information is the fifth

element (Floridi 1999), and the military now speaks of cyber warfare as "the fifth domain of warfare". The impression is that, in the future, such fifth domain will end up dominating the others.

Finally, think of the origin of the Internet (Floridi 1995). We know it was the military outcome of the arms race and of nuclear proliferation, but we were distracted by the development of the Web and its scientific origins, and forgot DARPA (Defense Advanced Research Projects Agency). History has merely caught up with us.

The previous sketch should help one understand why cyberwar, or more generally information warfare, is causing radical transformations in our ways of thinking about military, political and ethical issues. Are we going to see a new arms race, given the very high rate at which cyber weapons "decay"? After all, you can use a piece of malware only once, for a patch will then become available, and only within, and against, a specific technology that will soon be out of date. If cyber disarmament is ever going to be an option, how do you decommission cyber weapons? Drones can be hacked: will Pony Express make a patriotic come back in 2060 as the last line of defence against an enemy that could tamper with anything digital and online? Some questions make one smile, but others are increasingly problematic. Let me highlight two sets of them that should be of interest to philosophers.

The body of knowledge and discussion behind Just War Theory is detailed and extensive. It is the result of centuries of refinements since Roman times. The methodological question we face today is whether information warfare is merely one more area of application, or whether it represents a disruptive novelty as well, which will require new developments of the theory itself. For example, within the *jus ad bellum*, which kind of authorities possesses the legitimacy to wage cyberwar? And how should a cyber attack be considered in terms of last resort? And within the *jus in bello*, what level of proportionality should be attributed to a cyber attack? How do you surrender to a cyber enemy? Or how will robots deal with non-combatants or treat prisoners?

Equally developed, in this case since Greek times, is our understanding of military virtue ethics. How is the latter going to be applied to phenomena that are actually reshaping the conditions of possibility of virtue ethics itself? Bear in mind that any virtue ethics presupposes a philosophical anthropology (Aristotelian, Buddhist, Christian, Confucian, Fascist, Nietzschean, Spartan and so forth), and information warfare is only part of the information revolution, which is also affecting our self-understanding as informational organisms (more on this later). Take, for example, the classic virtue of courage: in what sense can someone be courageous when manoeuvring a military robot? Indeed, will courage still rank so high among the virtues when the capacity to evaluate and manage information and act upon it wisely and promptly will seem to be a much more important trait of one's character?

Similar questions seem to invite new theorising, not mere application or adaptation of old ideas. Perhaps, instead of updating our old theories with more and more service packs, we might want to consider upgrading them by developing a new macroethics. Information warfare calls for an information ethics, and so does the civilian uses of robots, as we shall see in the next section.

4 Artificial Companions

At the beginning of *Much Ado About Nothing*, Beatrice asks "Who is his companion now?" These days, the answer could easily be "an artificial agent".

Artificial companions (henceforth, ACs) come in all sizes and shapes. Examples include the Wi-Fi enabled rabbit *Nabaztag*, the therapeutic robot baby harp seal *Paro*, the child-sized humanoid robot *KASPAR*, or the interactive doll *Primo Puel* (more than 1 million copies sold since 2000, it is produced by Bandai, interestingly the same producer of the Tamagotchi). They are part of an ever-widening species of smart robots used in health care, industry, business, education, entertainment, research and so forth.

The technology to develop ACs is largely available, and the question is "when" rather than "whether" they will become commodities. Of course, the difficulties are still formidable, but they are not insurmountable and seem rather well understood. ACs are embodied (sometimes only as avatars, often as robotic artefacts as well) and embedded artificial agents. They are expected to be capable of some degree of speech recognition and natural language processing; to be sociable, so that they can successfully interact with human users; to be informationally skilled, so that they can handle their users' ordinary informational needs; to be capable of some degree of autonomy, in the sense of self-initiated, self-regulated, goal-oriented actions; and to be able to learn, in the machine-learning sense of the expression. ACs are not the end-result of some unforeseeable breakthrough in productive, cognitive AI but the social equivalent of *Deep Blue* and *Watson*. They can deal successfully with their tasks, even if they have the intelligence of a toaster.

Perhaps because ACs are neither Asimov's robots nor *Hal*'s children, the philosophical questions they posit are very concrete. When is an informational artefact a companion? Is an AC better than a child's doll, or a senior's goldfish? If it is the level and range of interactivity that counts, then an AC performs better than a goldfish. If it is the emotional investment that the object can invoke and justify that matters, then the old Barbie might qualify as a companion as well as an AC. Is there something morally wrong, or mildly disturbing, or perhaps just sad in allowing humans to establish social relations with pet-like ACs? And why this may not be the case with biological pets? Is their non-biological nature that makes philosophers whinge? Not necessarily, since, to a Cartesian, animals are machines, so having engineered pets should really make no difference. These are not idle questions. Depending on their answers, one may be able to address human needs and wishes more effectively, with a deep impact on economic issues. In 2011, for example, an estimated $50.84 billion was spent on biological pets in the United States alone (source: American Pet Products Association). The arrival of a whole population of ACs could change all this dramatically.

5 Conclusion: the Fourth Revolution

The informational turn may be described as the fourth step in the process of dislocation and reassessment of humanity's fundamental nature and role in the universe (Freud 1917). Previous revolutions have made us realise that we are not

immobile, at the centre of the universe (Copernican revolution), that we are not unnaturally separate and diverse from the rest of the animal kingdom (Darwinian revolution), and that we are very far from being Cartesianly transparent to ourselves (Freudian revolution). We are now slowly accepting the idea that we might be informational organisms among many others, *inforgs* that are going to live and interact with other smart, engineered artefacts often not so different from biological agents (Turing revolution). When ACs will be commodities, people will accept this conceptual revolution with much less reluctance. It seems that, in view of this important change in our self-understanding and of the sort of IT-mediated interactions that we will increasingly enjoy with other agents, whether biological or artificial, the best way of tackling the previous questions may be from an environmental approach, one which does not privilege the natural or untouched, but treats as authentic and genuine all forms of existence and behaviour, even those based on artificial, synthetic or engineered artefacts. Beatrice would not have understood "an artificial companion" as an answer to her question. Yet future generations will find it unproblematic. It seems to be our task to make sure that the transition from her question to their answer will be as ethically smooth as possible.

References

Floridi, L. (1995). Internet: which future for organized knowledge, Frankenstein or Pygmalion? *International Journal of Human—Computer Studies, 43*, 261–274.

Floridi, L. (1999). *Philosophy and computing: An introduction*. London: Routledge.

Freud, S. (1917). A difficulty in the path of psycho-analysis. *The Standard Edition of the Complete Psychological Works of Sigmund Freud* XVII (1917–1919), pp. 135–144.

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge: MIT.