Artificial Intelligence’s New Frontier: Artificial Companions and the Fourth Revolution

Luciano Floridi¹,²

¹Research Chair in Philosophy of Information and GPI, University of Hertfordshire; ²St Cross College and IEG, University of Oxford.

Address for correspondence: Department of Philosophy, School of Humanities, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK; l.floridi@herts.ac.uk

Abstract

In this paper I argue that recent technological transformations in the life-cycle of information have brought about a fourth revolution, in the long process of reassessing humanity’s fundamental nature and role in the universe. We are not immobile, at the centre of the universe (Copernicus); we are not unnaturally distinct and different from the rest of the animal world (Darwin); and we are far from being entirely transparent to ourselves (Freud). We are now slowly accepting the idea that we might be informational organisms among many agents (Turing), inforgs not so dramatically different from clever, engineered artefacts, but sharing with them a global environment that is ultimately made of information, the infosphere. This new conceptual revolution is humbling, but also exciting. For in view of this important evolution in our self-understanding, and given the sort of IT-mediated interactions that humans will increasingly enjoy with their environment and a variety of other agents, whether natural or synthetic, we have the unique opportunity of developing a new ecological approach to the whole of reality.
At the beginning of *Much Ado About Nothing*, Beatrice asks “Who is his companion now?”. Today, the answer could easily be an artificial agent.

Artificial companions (Lee et al. [2007]) (henceforth ACs) come in various forms. Examples include the Wi-Fi enabled rabbit *Nabaztag*, the therapeutic robot baby-harp seal *Paro* (Wada and Shibata [2007]), the child-sized humanoid robot *KASPAR* (Cole [2007]) or the interactive doll *Primo Puel*. This first generation of simple ACs is interactively sociable, informationally skilled and capable of some basic natural-language processing (AISB [2005]). Later generations are expected to become more autonomous, and hence behave in self-initiated, self-regulated, goal-oriented ways, and to be able to learn from their users, in the machine-learning sense of the expression (Wilks [9 November 2007]). The technology is largely available already, and the question is when rather than whether ACs will become commodities (Benyon and Mival [2007]).

Bandai, interestingly the same producer of the Tamagotchi, has sold more than one million copies of *Primo Puel* since 2000. ACs are a technological success because they are not the outcome of some unforeseeable breakthrough in Good Old-Fashioned AI, but the social equivalent of *Deep Blue*: they can deal successfully with their interactive tasks, even if they have the intelligence of a toaster. And they are philosophically significant precisely because they are neither Asimov’s robots nor *Hal’s* children. Out of the realm of thought experiments and unrestrained speculations, they posit very concrete, ethical challenges (Floridi [2007b]), which usher in what may be defined as a fourth revolution in humanity’s self-understanding (Floridi and Sanders [2004]). Let me explain.

How we build, conceptualise and interact with ACs will influence our future ability to address humanity’s needs and wishes, with a serious impact on standards of living and related economic issues. In 2007, for example, an estimated $40.8 billion was spent on biological

It is often argued that ACs will become increasingly popular the more they are able to assist elderly users satisfactorily and cost-efficiently (Mival et al. [2004]). This is true and encouraging, especially for countries where there is an aging population, like Japan and parts of Europe. However, we should remember that future generations of senior citizens will not be “e-migrants” but children of the digital era. Here the gaming industry provides useful projections. Today, “sixty-seven percent of American heads of households play computer and video games” and “the average game player is 33 years old and has been playing games for 12 years”.\footnote{Source: Entertainment Software Association, http://www.theesa.com/facts/top_10_facts.php} When they retire, it is not so much that they will be unable to use IT products, as that they may need help to do so, in the same way that one may still be perfectly able to read, but no longer without glasses. Thus, they may welcome the support of a personal assistant in the form of an AC, which can act as an interface to the rest of the world. ACs should be planned more with the digitally impaired in mind rather than the computer illiterates.

The last point suggests that, in the long term, ACs may be evolving in the direction of specialised computer-agents, dedicated to specific informational tasks, following trends already experienced in other technological industries. Three are already envisionable.

First, ACs will address social needs and the human desire for emotional bonds and playful interactions, not unlike pets (Lee et al. [2007]), thus competing with the omnipresent TV for attention. Here a key question is whether allowing humans to befriend ACs might be morally questionable. Should their non-biological nature make us discriminate against them? Not necessarily, if one agrees with Descartes [1996], Huxley [1893] or Wiener [1961] who argued that animals are living machines anyway. On the other hand, the question casts an interesting light on our understanding of what kind of persons we would like to be. Perhaps
there is nothing wrong with pet-like ACs. After all, they already constitute a widespread and innocuous phenomenon. In January 2008 there were more than 220 million neopets online, owned by more than 150 million people. Nobody has yet raised any moral objection.

Second, ACs will provide ordinary information-based services, in contexts such as communication, entertainment, education, training, health and safety. Like avatars, ACs are likely to become means to interact with other people as well as social agents in themselves. In this context, one of the challenges is that their availability may increase social discriminations and the digital divide (Norris [2001]). In particular, with respect to individuals with relevant needs or disabilities, the hope is that they will be able to enjoy the support of an AC, just as the Motability Scheme in the UK, for example, provides disabled individuals with the opportunity to own or hire powered wheelchairs and scooters at affordable prices.

Finally, ACs will act as “memory stewards” (O’Hara et al. [2006]), creating and managing a repository of information about their owners. This is also good news. For leaving behind a lasting trace has always been a popular strategy to withstand the oblivion inevitably following one’s death. Nowadays, we can all be slightly less forgettable, insofar as we succeed in our mnemonic DIY. This trend will grow exponentially, once ACs become commodities. Storage capacity is increasing at an astonishing pace. “Between 2006 and 2010 [...] the digital universe will increase more than six fold from 161 exabytes to 988 exabytes”. It is only a matter of decades before a whole life will be recordable by an AC. But then, it will not be long before some smart application – based on a life-time recording of someone’s voice, visual and auditory experiences, expressed opinions and tastes, linguistic habits, millions of digital documents and so forth – will be able to simulate that person, to the point where one may interact with her AC even after her death, without noticing, or even

---

3 Source: Neopets, http://www.neopets.com/petcentral.phtml
4 Source: Motability, http://www.motability.co.uk/
deliberately disregarding, any significant difference. A personalised AC could make one “e-
mmortal”. After all, an advanced, customised ELIZA can already fool many people in Second 
Life. Our new memory stewards will exacerbate old problems and pose new and difficult 
ones. What to erase, rather than what to record (as is already the case with one’s emails), the 
safety and editing of what is recorded, the availability, accessibility and transmission of the 
information recorded, its longevity, future consumption and “re-playing”, the management of 
ACs that have outlived their human partners, the redressing of the fine balance between the 
art of forgetting and the process of forgiving (consider post-dictatorial or post-apartheid 
cultures), and the impact that all this will have on the construction of personal and social 
identities, and on the narratives that make up people’s own past and roots: these are only 
some of the issues that will require careful handling, not only technologically, but also 
educationally and philosophically.

The previous trends suggest that ACs are part of a wide and influential informational 
turn, a fourth revolution in the long process of reassessing humanity’s fundamental nature 
and role in the universe. We are not immobile, at the centre of the universe (Copernicus); we 
are not unnaturally distinct and different from the rest of the animal world (Darwin); and we 
are far from being entirely transparent to ourselves (Freud). We are now slowly accepting the 
idea that we might be informational organisms among many agents (Turing), inforgs not so 
dramatically different from clever, engineered artefacts, but sharing with them a global 
environment that is ultimately made of information, the infosphere. The information 
revolution is not about extending ourselves, but about re-interpreting who we are. When ACs 
become commodities, people will accept this new conceptual revolution with much less 
reluctance. It is humbling, but also exciting. For in view of this important evolution in our 
self-understanding, and given the sort of IT-mediated interactions that humans will 
increasingly enjoy with other agents, whether natural or synthetic, we have the unique
opportunity of developing a new ecological approach to the whole of reality. This approach is not just biocentric and does not privilege only the natural or the untouched, but treats as authentic and genuine all forms of existence and behaviour, even those based on synthetic or engineered artefacts. In the end, how we build, shape and regulate ecologically the new infosphere is the crucial challenge brought about by ACs and the fourth revolution (Floridi [2007a]). Beatrice would not have understood “an artificial companion” as an answer to her question. Yet future generations will find it unproblematic. It is going to be our task to ensure that the transition from her question to their answer will be as ethically smooth as possible.
References


