e-mentor
No 5 (22) / 2007
www.e-mentor.edu.pl/eng

Artificial Companions and their Philosophical Challenges

Luciano Floridi

At the beginning of Much Ado About Nothing*, Beatrice asks "Who is his companion now?" (Act 1, Scene 1). These days, the answer could easily be "an artificial agent". The technology to develop* artificial companions *(henceforth AC) is largely available, and the question is "when" rather than "whether" they will become commodities*[1]*. Of course, the difficulties are still formidable, but they are not insurmountable. On the contrary, they seem rather well-understood, and the path from theoretical problems to technical solutions looks steep but climbable*[2]*. So, in the following pages, I wish to concentrate not on the technological challenges, which are important, but on some philosophical issues that a growing population of AC will make increasingly pressing.*

We know that AC are embodied (perhaps only as avatars, but possibly 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 (NLP); to be sociable, so that they can successfully interact with human users (their human companions, to be e-politically correct); 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

[1] For an introduction to artificial companions see Yorick Wilks (2005), "Artificial companions", *Interdisciplinary Science Reviews*, 30, 145-152 and Yorick Wilks (2006), http://www.companions-project.org/downloads/comp.6pp.ven.pdf "Artificial Companions as a new kind of interface to the future Internet", Oxford Internet Institute, Research Report 13, www.oii.ox.ac.uk/research/publications/RR13.pdf

[2] See for example the European projects on AC developed by the Adaptive Systems Research Group of the University of Hertfordshire on Human-Robot Interaction, http://adapsys.feis.herts.ac.uk/, and the COMPANIONS project a EU Integrated Project IST-34434 (Intelligent, Persistent, Personalised Multimodal Interfaces to the Internet), http://www.companions-project.org/

of some unforeseeable breakthrough in Good Old Fashioned AI. They are more the social equivalent of *Deep Blue*: they can deal successfully with their tasks, even if they have the intelligence of a refrigerator.

Although ACs are neither Asimov's robots nor *Hal*'s children, their nature posits several philosophical questions. Take some very elementary artificial agents, such as *Virtual Woman*[3], or the more recent and fancier *Primo Puel*[4], *Paro*[5] and *KASPAR*[6]. One ontological question is: when is *x* a companion? Could the previous examples be considered members of a first generation of simple companions? Is any of them better than a child's doll, or a senior's goldfish? Is the level and range of interactivity that matters (but then, the goldfish may not count) or the emotional investment that the object can invoke and justify (but then, the old Barbie might count). 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. All 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 2007, for example, an estimated $40.8 billion will be spent on biological pets in the U.S. alone[7]. The arrival of a whole population of ACs could change all this dramatically.

Suppose one may solve the previous questions to one's satisfaction. It is often said that artificial companions will help the disadvantaged. This is true, but a proviso is in order in the case of elderly users. Technology, demography and IT-skills follow converging lines of development. Future generations will be used to interact with digital artefacts in a way that we can only partly appreciate. To them, it will be natural and unproblematic to be in touch with artificial agents and to be related to the world through them. The more the threshold between online (or life-on-line) and "onlife" (or life-off-line) becomes blurred, the easier it will be to accept and be able to socialise with and through synthetic, hybrid, artificial companions. Future generations of senior citizens

[3] Available since the late 1980s, http://virtualwoman.net/
[4] More than one million sold since 2000 by Bandai, interestingly the same producer of Tamagotchi.
[5] http://paro.jp/english/index.html
[6] http://www.iromec-project.co.uk/
[7] Source: http://www.appma.org/press_industrytrends.asp

won't be immigrants but children of the digital era. Missing this point may be an easy but serious mistake, with significant, financial consequences. It is not that our grandchildren, in their retirement age, will be unable to use some kind of information technologies, but that they may no longer be able to, more in the way in which one may still be perfectly able to read, but no longer without glasses. 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"[8]. When they retire, they will not need to be explained what a computerised agent is, or how to use a mouse. But they will definitely enjoy the help of a personal assistant, a facilitator understood as an interface to the rest of the infosphere. In this sense, the evolution of artificial companions might be moving in the direction of specialised computer-agents for intelligence-non-intensive (aka stupid), informational tasks. Like avatars, they may more likely be means to tele-socialise with other human agents, rather than social agents in themselves.

The last point raises a further consideration. It seems that the population of ACs will be growing and evolve in the future and, as in the case of vehicles, one may expect robust trends in specialization. Today, we see and plan ACs as:

1. social workers, which may cope with human loneliness, social needs and the desire for emotional bonds and interactions, not unlike pets;
2. service providers, in contexts such as education and communication, health, safety, training, etc.;
3. memory keepers (see the *Memories for Life* project[9], for example), as stewards of the informational space constituted by human memories, whether individual or socially shared.

In each case, different questions arise.

Regarding (1), 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? The question casts an interesting

[8] Source: http://www.theesa.com/facts/top_10_facts.php
[9] Kieron O'Hara et al. (2006), "Memories for life: a review of the science and technology", J. R. Soc. *Interface* 3, 351–365.

light on human nature, and it seems to belong to the sort of questions asked with respect to recreational drugs. Essentially: what's wrong with it? Different answers seem to be based on different philosophical anthropologies or conceptions of what it means to be authentically human.

Regarding (2), may the availability of ACs as service providers increase social discriminations and the digital divide? For example, should individuals with relevant disabilities have the right to be supported by ACs? Today, the Motability Scheme in the UK, for example, provides citizens with physical disabilities, or health conditions affecting their mobility, with the opportunity to own or hire powered wheelchairs and scooters at affordable prices[10]. Should something similar happen for ACs? Consider that ACs might easily become embedded in future technological artefacts engineered for mobility, as prosthetic memory agents for those with memory dysfunctions. As for the new generations of students, the more memories are exogenous rather than endogenous, the more the educational system will have to provide individuals with the sort of skills required to access and give sense to information. Languages (not only natural, but also mathematical and artificial) and culture (not mere facts, but an open-ended appreciation and understanding of human developments and achievements) will be crucial.

Regarding (3), creating ACs as artificially-living diaries will pose interesting challenges. Let us not forget that, short of the real thing, an ever-lasting memory is the second best choice to reach immortality. The accumulation of memory has been, for a long time, a crucial but friction-full business. Never before has the creation, reproduction, management and destruction of documents been just a click away and so cheap, in terms of computational and recording resources. This trend will only increase once ACs, as memory stewards, will become available. What to record, the safety and editing of what is recorded, the availability and accessibility of the information, its longevity and future consumption and "re-playing", the impact that all this will have on the construction of individuals', groups' and social identities and on the narratives that make up people's own past and roots, these are all issues that will require very careful handling, not only technically, but also ethically. For example, who will be the new, professional memory workers? In the past, the definition of a famous person could be

[10] Source: http://www.motability.co.uk/Templates/Internal.asp?nodeid=89861

provided in terms of someone whose memories were recorded and managed by professional memory workers, being these artists (poets, sculptors, painters, musicians, architects and so forth), chroniclers, historians or journalists. Nowadays, we are all famous and a little bit less mortal insofar as we succeed in being our own memory keepers. However, in the same way as the commodification of cameras has not made the profession of photographer disappear, but it has reshaped its nature and our understanding of it, will AC also cause a comparable transformation? Are we going to witness the emergence of new professional creators and managers of digital memories? And on a related but different note, what sort of memories will or should survive their human supports? And what are we going to do with the artificial companions that will have outlived their human partners? Reset them? Edit, cut and paste, reformat? Are we going to see memory hackers? When a couple will divorce, who will have the right to keep the AC that recorded the wedding and the first years of the kids? Will people be happy with duplicates or will they become attached to the specific artefact that holds the memories as well, the artefact itself (perhaps with its scratches and blips) having become humanly salient? Will someone's digital companion be more important than his old cufflinks or her old earrings? And how long will it take before some smart application, based on a life-time recording of someone's voice, interactions, visual and auditory experiences, tastes, expressed opinions, linguistic habits, million of documents (tax forms, emails, google searches, etc.) and so forth, will be able to imitate that person, to a point where you will write or even talk to someone actually dead without noticing any significant difference? An advanced, customised ELIZA could already fool many people in *Second Life*. Or will there be people working in Artificial Companions centre offering services who can impersonate dead people? Will some future service company offer you the possibility of downloading enough information about you as to make a you-liza available even when some time you are not there, or even when you are no longer there? And how will future generations cope with the art of forgetting, so often crucial for the process of forgiving?

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. We are not immobile, at the centre of the universe (Copernican revolution), we are not

unnaturally separate and diverse from the rest of the animal kingdom (Darwinian revolution), and we are very far from being Cartesianly transparent to ourselves (Freudian revolution). We do not know if we are the only intelligent form of life. But we are now slowly accepting the idea that we might be informational entities and agents among many others, and not so dramatically different from smart, engineered artefacts. 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 that it is going to be our task to make sure that the transition from her question to their answer will be as ethically smooth as possible.

References

- the European projects on AC developed by the Adaptive Systems Research Group of the University of Hertfordshire on Human-Robot Interaction, http://adapsys.feis.herts.ac.uk/
- the COMPANIONS project a EU Integrated Project IST-34434 (Intelligent, Persistent, Personalised Multimodal Interfaces to the Internet), http://www.companions-project.org/
- Kieron O'Hara et al. (2006), "Memories for life: a review of the science and technology", J. R. Soc. *Interface* 3, 351–365
- Yorick Wilks (2005), "Artificial companions", Interdisciplinary Science Reviews, 30, 145-152

- Yorick Wilks (2006), http://www.companions-project.org/downloads/comp.6pp.ven.pdf "Artificial Companions as a new kind of interface to the future Internet"
- Oxford Internet Institute, Research Report 13, www.oii.ox.ac.uk/research/publications/RR13.pdf
- http://www.appma.org/press_industrytrends.asp
- http://www.iromec-project.co.uk/
- http://www.motability.co.uk/Templates/Internal.asp?nodeid=89861
- http://paro.jp/english/index.html
- http://www.theesa.com/facts/top_10_facts.php
- http://virtualwoman.net/

Info about the author

Luciano Floridi holds the Research Chair in Philosophy of Information, Department of Philosophy, School of Humanities, University of Hertfordshire; GIP, University of Hertfordshire. He is the founder and coordinator of the IEG, the interdepartmental research group on the Philosophy of Information at Oxford University.