Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues

by

Malcolm Peltu
Oxford Internet Institute

Yorick Wilks
Oxford Internet Institute

Foreword

This paper summarizes discussions at the multidisciplinary forum[1] held at the University of Oxford on 26 October 2007 entitled *Artificial Companions in Society: Perspectives on the Present and Future*, as well as an open meeting the previous day addressed by Sherry Turkle[2]. The event was organized by Yorick Wilks, Senior Research Fellow for the Oxford Internet Institute (OII)[3], on behalf of the e-Horizons Institute[4] and in association with the EU Integrated Project COMPANIONS.

COMPANIONS is studying conversational software-based artificial agents that will get to know their owners over a substantial period. These could be developed to advise, comfort and carry out a wide range of functions to support diverse personal and social needs, such as to be 'artificial companions' for the elderly, helping their owners to learn, or assisting to sustain their owners' fitness and health. The invited forum participants, including computer and social scientists, also discussed a range of related developments that use advanced artificial intelligence and human–computer interaction approaches. They examined key issues in building artificial companions, emphasizing their social, personal, emotional and ethical implications.

This paper summarizes the main issues raised. Position papers[5] prepared for, and generally summarized at, the forum by participants formed a core resource for the event. Most direct quotes from participants in this paper come from the position papers. Appendix 1 contains examples of current artificial companions and related research projects mentioned at the forum.

Acknowledgements

---

[1] For more background on the forum, see: http://www.companions-project.org/events/
[2] Sherry Turkle is Abby Rockefeller Mauze Professor of the Social Studies of Science at MIT and Technology Director, MIT Initiative on Technology and Self. Her talk was entitled *Cyberintimacies/Cybersolitudes* (see http://www.oii.ox.ac.uk/events/details.cfm?id=150).
[3] See: http://www.oii.ox.ac.uk
[4] The e-Horizons Project is a unit of the James Martin School of the 21st Century at the University of Oxford. It focuses on critically assessing competing visions of the future of media, information and communication technologies and their societal implications (see http://www.e-horizons.ox.ac.uk).
[5] The position papers are available at: http://www.companions-project.org/events/

CONTENTS

Overview: the nature and significance of artificial companions

What is an artificial companion?[6]

Artificial companions (ACs) are typically intelligent cognitive 'agents', implemented in software or a physical embodiment such as a robot. They can stay with their 'owner' for long periods of time, learning to 'know' their owner's preferences, habits and wishes. An AC could enter a close relationship with its owner by chatting to, advising, informing, entertaining, comforting, assisting with tasks and otherwise supporting her or him. In doing this, the companion should make no technical demands on the user.

This view of artificial companions, also known as 'digital companions' or 'embodied conversational agents (ECAs)'[7], overlaps with a number of similar research conceptions crossing a variety of academic disciplines, such as 'affective computing' (Picard 1997)[8], 'emotion-oriented systems'[9], 'relational artefacts' (Turkle 2006) and 'humanoid robotics'[10]. The COMPANIONS project itself emphasizes language-centered software agents, rather than robotics.

These kinds of artificial companions could alter the way we think about the relationships of people to computers, the Internet and other information and communication technologies (ICTs). These changes could also affect relations, both individual and social, between human beings who own companions. Discussion at the e-Horizons forum moved beyond this conception to encompass a wide spectrum of real and possible entities that could be described as 'artificial companions'. These range from everyday electronic gadgets (e.g. mobile phones and GPS car navigation systems) to 'iEverything' intelligent toys, 'pets', vacuum cleaners, refrigerators and an ever growing range of digital innovations.

Figure 1 illustrates a few of the many artificial companions mentioned at the forum. The core areas of interest in these discussions were the more advanced AI-based software developments that move toward creating new forms of more intimate 'emotional' relationships between people and computers.

Figure 1. Examples of current artificial companions*

Type	Examples
'Intelligent' toy, 'pet' and domestic robots, from which valuable knowledge on users'	Sony Aibo robot dog
	Furby, an owl-like robot that appears to learn English

---

[6] See Appendix 1 and the Reference section for information on related research, products and publications providing further background to the issues raised in this paper.

[7] See, for example, Cassell et al (2000) and Pelachaud (2005; 2007).

[8] See, for instance, the work of MIT's Affective Computing Research Group (http://affect.media.mit.edu/index.php http://affect.media.mit.edu/people.php?id=hoda).

[9] See, for example, Sloman (2006), the HUMAINE research network and the Emotionally Intelligent Interface research led by Peter Robinson, Cambridge University based on the theory of mind of Professor Simon Baron-Cohen, Director of the Autism Research Centre at Cambridge (http://www.cl.cam.ac.uk/research/rainbow/emotions/).

[10] See, for example, MIT's Humanoid Robotics Group (http://www.ai.mit.edu/projects/humanoid-robotics-group/index.html).

emotional reactions to ACs can be drawn	Paro, a seal-like robot that seeks to create an emotional attachment in the owner.
	Primo Puel talking doll
	Tamagotchi, virtual creatures that can 'grow' from being a 'child' to 'healthy adult', but 'die' if they are neglected.
	Roomba vacuum cleaner robot
Software-based companions	BASIC (Believable Adaptable Socially Intelligent Character)
	Beating the Blues computer-based therapy
	Greta, a virtual 3D embodied conversational agent
	Laura fitness health adviser
Humanoid and other forms of advanced robotics	Ecobot energetically autonomous robots
	Kismet 'sociable machine'
	KASPAR, a child-like robot for studying human–robot interaction
	Pearl nursebot
	Radar learning robot

*See Appendix 1 for more details on these and other ACs and robots

Congenial support for the vulnerable: the COMPANIONS approach

The European Commission's COMPANIONS[11] project—which was the prime mover behind the e-Horizons forum—adopts a widely held perception of ACs as being most suitable as a 'friendly' and socially-valuable support for a wide range of targeted groups, like the elderly, fitness-conscious, young or disabled. It is particularly interested in creating software companions that could act as an enhanced interface to the future Internet by providing new and much more personal ways of dealing with the 'overload' of information opened up by the World Wide Web and other ICTs. In particular, it is exploring the management of personal information about the owner's life and memories.

Such agents would be integrated as the human interface to the development of a 'more intelligent Web', such as the proposed 'Semantic Web'[12]. These agents would be able to exploit the richer and more rigorous linking of data, which Shadbolt et al (2006) argue is a key goal of the envisaged progression from the World Wide Web of documents to the Semantic Web of data. This will enable Web content to be accessed directly by users or software 'agents' through an understanding of the

---

[11] See http://www.companions-project.org for details of the COMPANIONS project. An example of an AC being developed for it is at: http://www.asanangel.fr/morgan/
[12] For example, see Berners-Lee et al. (2001) for a discussion on the emergence of the concept of a Semantic Web.

meaning of the content of texts on the Web (see also Wilks 2006: 16–17). Artificial companions will be able to draw on such developments to support their owners[13].

COMPANIONS is contributing to, and draws on, wider AC and related developments.[14] It is developing demonstrators consisting of 'persistent' (long-term) personalized agents that will engage in ongoing conversations with their users to build a narrative about that person's life and needs over a long period of time. These are targeting two main areas of support. One is a Senior Companion[15] agent that will communicate with its user by adapting to his or her voice, needs and interests in order to provide company for the lonely and to help access information and services, including how to react to emergencies. The other is a Health and Fitness Companion[16] that will support healthy eating habits and fitness activities by maintaining records of its user's health-related, eating and exercise information. These are typical of the AC applications envisaged by many forum participants.

The project is concentrating on three key AC technologies in building these demonstrators: memories for life and identity (e.g. Wilks 2006)[17]; natural language processing (NLP) and speech technology[18]; and agents and the Semantic Web (Wilks forthcoming). The demonstrators will deploy multimodal methods of communication, including NLP dialogues, speech, advanced visual technologies[19] (e.g. tracking eye movements[20]), sophisticated facial expressions and other non-verbal cues (e.g. André et al 2004; André 2007; Pelachaud 2007), touch screens and sensors.

Roots and branches of the artificial companions family

Two main streams of research were identified as the prime sources of current artificial companion work: artificial intelligence (AI) and human–computer interaction (HCI). Taylor and Swan (2007) summarize in their e-Horizon forum position paper how these streams have converged to influence AC design and development.

---

[13] Bryson et al (2002) propose that the semantic web (particularly Web services) could be thought of neither as passive content nor other agents to be negotiated with, but rather as extensions to the mind of an agent's user. Users determine their agents' motivational structure and communicate their needs to it, then the Web provides the agent with capacities to better meet those needs.

[14] COMPANIONS' project leader Yorick Wilks referred at the forum to a 'cloud' of projects undertaking work in similar area. These include: the European Commissions' projects HUMAINE, AMIDA, CALLAS and INDIGO; Birmingham University School of Computer Science's CoSy; SRI International's CALO; and the US CARTE initiative. See Appendix 1 for further details.

[15] See http://www.companions-project.org/demonstrators/senior.cfm for an example of a COMPANION's approach to finding out the requirements of senior citizens.

[16] See http://www.youtube.com/watch?v=KQSiigSEYhU for a video of an early prototype.

[17] See also, for example, http://www.memoriesforlife.org regarding the UK EPSRC research project Memories for Life (M4L), which is seeking to help define and solve the problems caused by people storing increasingly large quantities of information about themselves using digital versions of life's memories (in the form of photographs, documents, video, etc).

[18] For more on NLP and related areas see, for example, research at COMPANIONS partners at the University of Sheffield NLP Group (http://nlp.shef.ac.uk) and University of Oxford's Department of Linguistics and Phonetics (http://www.clg.ox.ac.uk). The speech technology research of Roger Moore at the University of Sheffield, part of the COMPANION's team, was also highlighted at the forum (see Zyga 2007 and http://www.dcs.shef.ac.uk/~roger/).

[19] See, for instance, Birmingham University's 3D vision research (http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#compmod07)

[20] See, for example, Rehm and André (2005).

They point out that AI and associated robotics approaches have undergone 'significant and dramatic changes since the 1950s' (e.g. see Turing 1950; Newell and Simon 1963; Sloman 1978; Weizenbaum 1984; Crevier 1993; Papert 1993; Whitby 1996; Brooks 2002; Boden 2006)[21]. They comment that the traditional 'top-down, brute force' AI approach has been, over time, largely replaced by an AI 'which envisages learning as something that evolves from the ground up'.

In this evolution, according to Taylor and Swan, the AI focus has moved away from building autonomous machines. Instead, they say: 'No longer is it exclusively assumed that the sort of intelligence to be attained in machines should simulate human intelligence.' They identify Suchman's (1987; 2006) critique of some assumptions implicit in AI as a pivotal development opening new possibilities for fresh thinking on framing ideas about machine and human intelligence. Suchman's work has contributed to a growing recognition in HCI of the 'situated' character of human action: the way that even planned behaviour is contingent on the contexts we find ourselves in (see also, for example, Avgerou 2004; Jacko and Sears 2003; Latour 1993). Even more significantly, they say, are Suchman's insights into the 'constitutive' nature of the human–machine intersection: how human–computer actions and interactions can reconfigure the ways in which humans and machines are understood. Taylor and Swan (2007) note that this is also related to the re-casting of distinctions made between humans (and animals) and 'things', for instance through developments such as Actor Network Theory (Law and Hassard 1999).

Turkle (2006) describes this trajectory in terms of a move in debates on AI from being centred around the question of whether machines could 'really' be intelligent, in terms of the capabilities of the objects themselves, to current debates about relational and sociable machines (e.g. Breazeal 2002), where the focus is 'not about the machines' capabilities but about our own vulnerabilities'. Traditionally, she adds, AI 'concentrated on building engineering systems that impressed by their rationality and cognitive competence—whether at playing chess or giving "expert" advice. Relational artefacts, by contrast, are designed to impress not so much through their "smarts" as through their sociability'.

A shadow over this history has been the frequently unfulfilled predictions of the more optimistic AI enthusiasts about timescales for the building of intelligent machines that resemble humans. This perceived underachievement lay behind the Lighthill (1973) report for the UK Science Research Council, which was damning about the achievements and prospects of AI. This had a strong negative influence on funding for related research in the 1970s. Reservations about what was seen to be continuing AI overhyping were expressed at the e-Horizons forum.

However, Wilks (2006: 6) rejects the notion of AI as a 'failed project', as he says 'it is simply everywhere'. He notes that AI 'is in the computers on 200-ton planes that land automatically in dark and fog and which we trust with our lives; it is in chess programs like IBM's Big Blue that have beaten the world's champion; and it is in the machine translation programs that offer to translate for you any page of an Italian or Japanese newspaper on the web. And AI certainly is present in the computer technologies of speech and language'.

---

[21] For further background on AI, see also: http://www.aaai.org/AITopics/html/welcome.html

Wilks (ibid) also points out that many AI pioneers had always seen their field's mission as modelling the normal, rather than the 'superhuman', in order to capture 'the shorthand of reasoning, the tricks that people actually use to cope with everyday life. Only then would we understand the machines we have built and trained and avoid them becoming too clever or too dangerous.' It is in these ways that AI has been primarily contributing substantially to the development of the kinds of artificial companions that were the central concern of the forum reported here.

Key social, psychological and ethical issues raised by artificial companions

While there was a general acknowledgement that artificial companions can bring substantial practical benefits in a variety of socially valuable applications, the e-Horizons forum emphasized a number of deeper issues where these and associated developments raise profound social, psychological and ethical issues.

Rethinking what it means to be human

A strong theme in forum discussions revolved around questions about what ACs could mean in rethinking notions of human and machine 'intelligence'[22], personhood and the implications for relationships between humans and other entities. In her pre-forum talk, Sherry Turkle, Director of the MIT Initiative on Technology and Self Program, emphasized important implications as some societies become increasingly composed of 'tethered selves': people in continuous connection to virtual and real worlds, never alone although often so in terms of direct physical contact with other people. In this environment, ACs have a distinctive and important role because they focus on engaged relationships that recognize and attempt to deal with the personal emotions and social settings of users and their digital artefacts.

Such 'close encounters' with entities that may to some extent become extensions of our selves could challenge traditional understanding of the meanings of 'self' and 'personhood'. For instance, O'Hara (2007) argues that much philosophical discussions on memory (e.g. Warnock 1987: 1–14) assumes that personhood is more or less a matter of spatio-temporal continuity of the body, and that the human mind is more or less identical with the human brain. What then of the ability of an AC to build a similar database of memory, which could make users dependent on their ability to recall memories shared with their artificial companion?

Another key focal point of forum debates was the implications of ACs for ideas about the value of authenticity, especially concerning human emotions. Turkle (2006) points out: 'When we are asked to care for an object, when that object thrives and offers us attention and concern, we feel a new level of connection to it.'

Wilks (2006: 3–4) highlights the implications of this for ACs in the way an elderly Japanese woman related to a talking doll called Primo Puel[23]: 'she kissed it while it talked to her, and she said how safe she felt with it there, even when chattering away to itself in the next room. It was so much better against loneliness, she said, than praying in front of her dead husband's shrine, which had been no help, she

---

[22] For more background on the ideas developed by Turkle, see her seminal book The Second Self (Turkle 2005).
[23] See http://news.bbc.co.uk/1/hi/programmes/this_world/golden_years/4373857.stm

admitted'. Although the doll is a relatively primitive AC, it can record how Akino moves about the house, and can phone the health authorities if sensors in the rooms show that her routine changes. Wilks stresses that this illustrates how artificial companions are already entering our society 'gently and by stealth'. He suggests: 'We must think now about their technical basis, their limits, what role we want them to have, and how to protect ourselves from them and the effects of their arrival, should it become necessary'.

Figure 2 summarizes some important questions highlighted during the forum. Differing perspectives on how to address these are discussed later in the paper.

Figure 2. Key social, psychological and ethical challenges

What does it mean to be human when many once seemingly distinctive aspects of being human can now be simulated by digital systems?

What are the limits in computer modelling of human behaviour and thinking?

What do ACs say about the changing relationships between people, and between humans and non-human entities? Are relationships with non-biological entities intrinsically different?

How significant is authenticity of feeling in relationships? In particular (see Boden 2007), to what extent:
- Could an AC be made to appear to do/feel a certain emotion?
- Would the human user believe the AC could do/feel X?
- Would we want the user to believe this?
- Would this have an effect on the user's relations with other people?

Does it matter that ACs can be 'cheap dates' by creating empathy through very simple AI techniques (e.g. 'intelligent' toys like the Tamagotchi and Paro)?

Can it be good for us if our experience with relational artefacts is based on a fundamentally deceitful interchange in which the artefact persuades us that they know of, and care about, our existence? Or might it be good for us in the 'feel good' sense, but bad for us in a moral sense? (Turkle 2006).

Can (or should) empathy or love with an AC be as—or even more—rewarding than with another human?

What does it mean to be an 'intelligent machine' or 'artificial life' (e.g. an artificial companion)? Do these entities have rights and feelings in any meaningful sense?

What are the implications if ACs and robots develop an 'alien' culture (Winfield 2007) inscrutable to humans?

Balancing potential benefits and harms

Turkle (2006) articulated a concern frequently expressed during the forum: that some people might find ACs better companions than humans. She illustrates this by referring to a comment she recorded during her research. When talking about her relationship with Aibo, Sony's household entertainment robot, one woman suggested this device 'is better than a real dog ... It won't do dangerous things, and it won't

betray you ... Also, it won't die suddenly and make you feel very sad.' This kind of response to a machine was seen as something potentially disturbing by some, but welcomed by others who thought such relationship could be of great benefit to some people

The two-edged nature of ACs, as with most ICTs (Dutton 2004), was illustrated frequently during the forum. For instance, one of the key advantages of an AC, such as the COMPANIONS Senior demonstrator, is that it can acquire intimate knowledge about the user's life. This is essential if it is to gain a rapport with its owner to support his or her needs over a long time. However, this also raises potential data security and privacy issues (e.g. if the companion shares that knowledge with other ACs). This issue could be exacerbated if, as Winfield (2007) suggests, ACs and sociable robots develop a shared culture that is 'quintessentially alien, in effect an exo-culture…inscrutable to humans, which means that when bots start gossiping with each other about you, you will have absolutely no idea what they're talking about because—unlike them—you have no theory of mind for your digital companions.'

Lowe (2007) highlights the benefits artificial companions can bring by helping their users to reframe those choices that can be difficult to make if they require expertise they do not have, or when an apparently easy and desirable short-term choice has harder long-term consequences (e.g. in relation to a diet or educational requirements). Here, AC support could be helpful as an 'enforcer' (e.g. to ensure the full course of antibiotics is taken) or conscience (e.g. on what not to eat when on a particular diet). He graphically explained this process as: 'When the sirens of plan-breaking temptation sing, it may be your Companion that ties you to the mast'.

The user could face severe harm if an AC in an 'enforcer' role is imposed by a state, organization or individual with malign intent. However, Lowe also argues that by persistently prioritizing the long-term over the immediate and short-term or by relying on information about the user and her situation that is over-rigid or outdated even an otherwise benign AC may reduce the spontaneity and quality of its owner's life.

Underlying clashes of values and perceptions

Some forum participants welcomed the possibility of developing artefacts that could be of great benefit to people who, for various reasons, are unable to have equivalently fulfilling relationships with people. For instance, Levy (2007a) states he is 'convinced that, within twenty years at the latest, there will be artificial emotion technologies that can not only simulate a full range of human emotions and their appropriate responses, but also exhibit non-human emotions that are peculiar to computers (and robots).' At the forum, Margaret Boden argued that it would be 'tragic' to mistake falling in love with a robot or AC with falling in love in the 'deepest most important sense of personal human love' (e.g. see Fisher 1990). However, others welcomed, or at least accepted, the validity of love with an AC as one type of love across a spectrum ranging from human love, through pets and plants to emotional attachments to inanimate object like mobile phones. Levy (2007 a, b), for instance, welcomes of the prospect of people 'not only be falling in love with software companions, but also having sex with them'.

One discussion related to the existence of possible underlying 'essentialisms' influencing these conflicting perceptions about the degree to which people welcomed

or were apprehensive about AC developments. As indicated in Figure 3, one such extreme could be characterized as being that of 'humanists' who privilege human emotional authenticity and the uniqueness of human intelligence. At the opposite pole could be the view of the more extreme proponents of artificial life who would like to create a new, more perfect set of 'beings' (e.g. see Helmreich 1997). A spectrum of views between these poles was reflected at the forum.

Figure 3. Conflicting perspectives: Humanist v Artificial Lifer

'Humanist'	*'Artificial Lifer'*
Human relationships are the most important	Relationships with ACs can be as—or more—rewarding than human ones
Human and other biological life forms are unique and have a unique value	Artificial life artefacts can have equivalent value and rights to other life forms
Authenticity of human understanding and emotions is essential to engaged relationships	'Fit for a purpose' functionality is a good enough aim for AC and intelligent robot developments
ACs should primarily seek to mediate and support human–human communication	Interactions between a human and AC can be a valuable end in itself.
ACs should support human needs, including being a servant or even 'slave'	ACs represent a new breed of autonomous artificial life of value in its own right.
Computer modelling of human emotions and behaviour is too difficult to achieve in a realistic manner	Within a reasonable period it will be feasible to simulate a full range of human emotions and their appropriate responses.
Users should make influential contributions to the design of AC capabilities	The focus should be on technological innovation to create new choices for users.
Machine intelligence is fundamentally different to human intelligence	Machine intelligence can simulate and even emulate human intelligence
There can be no meaningful, reciprocated loving relationship with artificial life.	Falling in love with a AC or robot is as natural as other forms of love.

Challenges to creating effective appropriate artificial companions

Wilks (2007) says there is still some confusion and lack of consensus among specialists in this field over what is desirable in the self-presentation of a possible long-term artificial companion, how we should view or deal with them and what social and personal roles we want them to adopt with us. This is reinforced by Sloman's (2007) comment that the detailed requirements for ACs are 'not at all obvious, and will be found to have implications that make the design task very difficult in ways that have not been noticed'. However, he adds that overcoming these difficulties is 'perhaps not impossible if we analyse the problems properly'.

Much time at the forum was devoted to discussion of various approaches to designing and building artificial companions which would fulfill the main hopes in this field. This included an acknowledgement of the dangers of basing ACs on computing models that eliminate essential features of the real world being simulated to make the model manageable—without taking account of the richness, subtleties and complexities of human motivation and behaviour, social contexts and the complex interactions that take place with the physical environments in which we (and perhaps AC robots) move and live. Figure 4 summarizes some key questions raised about designing and build artificial companions.

Figure 4. The main artificial companion design and build questions

- What are the most appropriate Human–AC relationships?

- To what extent should ACs complement or replace human–human communication?

- How human-like should the AC be?

- What are the limits of computational modelling of personal emotions, intelligence and behaviour within realistic social and physical environments?

- How best can potential users of ACs influence designs, particularly when they may be unaware of what can be achieved and how their own requirements may change once they experience life with an artificial companion?

- How can the user requirements of vulnerable groups be elicited most effectively, such as for the elderly (see Newell 2007).

- What are the technical and cost constraints determining the feasibility of building ACs with appropriate performance levels (e.g. to achieve believable interactions, such as the realism of graphics in well-financed movie animation)?

- How can ACs be made as flexible as possible to enable them to meet, in a cost-effective manner, the distinctive needs of each user in their particular situations—and their evolving requirements over a long period of time?

- How transparent should an AC be in making clear the non-human nature of emotions and its machine-like 'thinking' processes?

- Should ACs be governed by ethical 'rules' relating to their behaviour towards users?

- Should ACs be regarded as having 'rights' (e.g. against 'abuse')?

Summary of remainder of this paper

The next section explores potential solutions and some barriers to resolving the design questions outlined here. This is followed by a discussion on the main policy issues relating to the growing use of ACs. These need to be better understood if a framework is to be created that supports wider use of artificial companions, while minimizing potential harms. The paper concludes with a summary of the longer-term

implications of living in societies populated by a growing range of artificial companions in continuous close engagements with people.

Approaches to designing and building artificial companions

What features are likely to make a 'good companion'?

As indicated in Figure 1 above, there are a wide range of entities that could be regarded as artificial companions. However, a number of broad AC types and core features emerged from discussions at the e-Horizons forum. These were broadly captured in Wilks' (2007) description of the qualities of a typical Victorian lady's companion. Many of these, he noted, would now be considered as part of the general caring and social services. Floridi (2007) identifies three classifications of such companions: as social workers; service providers in specific application contexts such as education, health, safety and communication; and memory keepers, as in the Memories for Life project (see Appendix 1). Boden (2007) expresses concern about the privacy aspects of being a 'confidant' gleaning intimate personal information, and so prefers an emphasis on ACs as 'conversationalists'.

Figure 5 summarizes some features the designers of ACs could consider, based on the Victorian lady's companion and other characteristics suggested at the forum. Many of these characteristics are actually extremely difficult even for humans to achieve (e.g. being supportive, engaging in long-term relationships, correctly interpreting emotions, being witty).

Figure 5. Potentially desirable features of an artificial companion

Able to recognize its 'owner' as an individual distinguishable from other people, animals and things in the environment.

Able to understand its owner's emotions and intentions, and respond in an appropriate manner.

Self-presentation by the AC that is coherent and believable.

Dependable, predictable behaviour in the service of the owner

Well-informed, particularly about its owner's needs, memories and social relationships

Supportive as a mediator in human–human communication, to complement existing communication methods (e.g. as a more friendly intelligent interface to the Web).

Long-term relationship if possible

Able to act as a caring expert advisor.

Independent, not requiring much effort from the user to use, sustain and nurture.

A discreet confidant in knowing what should or should not be communicated, and who or what (e.g. another AC) can be told particular information

Knowing its place relative to the human it supports and serves

Trustworthy

Responses firmly under control to best suit its owner

Good conversationalist and sympathetic listener

Polite where necessary, but firm if that is in the owner's interest

Modest and realistic about its own capabilities

Diverting, able to be witty and entertaining where appropriate

Cheerful even when faced with difficult problems

Specific looks not important provided they do not antagonize or irritate

Operationally reliable, resilient and requiring little special maintenance

Source: e-Horizons forum discussions (e.g. see Pulman 2007, Romano 2007, Wilks 2007)

One of the most distinctive features of ACs as digital artefacts is their integral need to engage with the emotional states of their users. This is emphasized by Cowie (2007), who points to the 'pervasive emotions' that colour most of life as being of particular importance to ACs—for which briefer intense emotional episodes may be less significance. Human–computer interfaces, he commented, were traditionally oriented towards people who could be expected, at least temporarily, to adopt a more or less unemotional stance to the task in hand. 'The populations who need artificial companions are likely to find it harder than average to adopt and sustain unemotional stances,' he adds.

Cowie (ibid) sounds a note of caution to designers of ACs: 'Among the commonest misunderstandings is the idea that engaging with emotion is necessarily about devices that mimic human emotional life.' Nevertheless, he emphasizes that a basic level of attention to emotion cannot be left out of work in this field.

How human should an artificial companion attempt to be?

The question of how much like a human the AC should present itself is central to debates about the design of an AC's self-presentation and behaviour. There was a general feeling at the forum that looking human was not an intrinsic requirement (e.g. see Wilks 2006; Pulman 2007), provided the form of presentation is appropriate to the applications and is believable (Romano 2007).

Different views were expressed about the relevance of the concept of an 'uncanny valley', which is said to affect the believability of synthetic human-like presentations. Romano (2007) explains that this valley was initially identified by Mori (1970) to represent the way emotional responses to an artificially generated character is thought to move through a curve that rises slowly up to a point, then descends rapidly into the uncanny valley as the presentation increasingly resembles a human—but not perfectly. After the lowest point is reached, the curve rises almost vertically to induce a positive reaction in humans to the presence of actual humans, which doesn't occur in relation to artificial entities that may closely resemble humans.

Cowie (2007) questions the uncanny valley thesis, as he regards it as a fallacy to suggest there is a fundamental obstacle to the acceptance of devices that are human-like. Rather, he sees 'signs of a long valley lying in wait for any approach to human-like qualities, or isolated potholes that a sensible team will steer round'. Nevertheless, he acknowledges that 'a near-miss can be a disaster' in terms of getting close to a human-like appearance. Predictability of behaviour, with expectations tailored to what can be delivered, is seen by many to be more important than human-like features in many AC presentations. Figure 6 summarizes some of the specific advice on designing an AC highlighted at the forum.

Figure 6. Examples of how to design effective AC user interactions

Ensure there is coherence in the appearance, characterization and behaviour of an AC. For instance, having photorealistic people does not make them necessarily more believable if their behaviour, speech and language are not of the same level as the visual qualities. Any discordant element can make the character unbelievable.

Consider the aesthetic qualities of the AC's movements, its behaviour and its ability to interpret and respond verbally and non-verbally to the human user (e.g. humour and politeness can make a character more human-like).

Build-in an autonomous modelling capability to enable the AC to generate its own behaviour according to its perception of the interaction with the user and the environment, generating appropriate expressions, emotions and credible behaviour.

Understand how to recognize and respond to attentive and motivational cues in real-time, in a manner that achieves natural dialogue behaviour and adequate response times. This requires recognition and response processes to be very fast, and synchronized, in order not to be perceived as awkward.

Distinguish decreasing engagement from those grounding problems where the speaker and listener have conflicting views regarding which object the conversation is about. For example, the fact that a user looks at an object that is not meant by the speaker does not necessarily mean the user is no longer engaged in a conversation.

Find the right level of sensitivity in responding to the user's attentive and motivational state, for instance by not overreacting to the user's state (e.g. Eichner et al. 2007).

The AC need not always display its emotion immediately in response to a user cue, as in some circumstance it may be more appropriate to mask, disguise or delay the response.

Source: Derived from André (2007), Romano (2007) and discussions at the e-Horizons forum

Understanding and meeting actual user requirements

A prime motivation for the development of HCI as a discipline in the 1960s and 70s was a growing awareness that most computer systems had been designed by technical experts, with little conception of what users actually wanted and needed. Despite many HCI advances and a greater awareness of the importance of meeting user requirements, many forum participants expressed concern that AI innovations like artificial companions are still driven too much by the perceptions and wishes of technically-oriented designers.

Prioritizing user-led design requirements

The frequently very close personal engagements envisaged for human–AC relations, involving sensitive feelings and intimate information, led many at the forum to emphasize the need for a creative and flexible understanding of human relations and psychology to be the foundation of artificial companion design and development. Newell (2007) strongly argues that it is important to understand the characteristics of potential users of artificial companions, as they can be very different from the characteristics generally favoured by designers of new technologies and their traditional user bases.

This is particularly significant for those users in the vulnerable groups who are likely have a special need for such companions. For example, Newell (ibid) explains: 'Older people have multiple minor disabilities, sensory, motor, and cognitive, and may also have major disabilities. The vast majority will have substantially different experiences of, and emotional attitudes to, new technologies than younger cohorts. Many of them will not understand, or be familiar with the jargon, metaphors, or methods of operation of new technologies, and have major lack of confidence in their ability to use them.'

Successful ACs that meet the needs of such groups require the development and adoption of appropriate methods of eliciting knowledge of their needs. Some potential users with no experience of associated technologies may not even be aware of the feasibility of technical capabilities that could provide the support they would like. Any requirements elicitation methods used for such potential AC users should take account of their special physical, sensory and cognitive characteristics, together with their experiences of, and attitudes towards, current technologies. Novel requirements-gathering techniques are likely to be needed for specific groups. For instance, Newell (ibid) says theatre techniques that dramatize future possibilities[24] have proved to be effective with older people.

It will also be important to determine what safeguards are required, for example to protect intimate AC-held information that the system's owners might not want to be divulged to close relatives. In addition, attempts should be made to find out why some people may decide they do not want to use certain types of technology in particular situations. Such 'digital choices' are evident, for example, in decisions by elderly people not to use the Internet even when it is freely available for other members of the family in their home (Dutton 2005). Turkle[25] described how a disabled colleague had been abused by nurses in a care home. Although he is aware of the development of 'nursebots' to care for people in such homes, he told her that he would still prefer to be dealt with by a human nurse, even one of his abusers, as he could still connect with the 'human narrative' of that person.

---

[24] See also Rice et al (2007) and http://www2.napier.ac.uk/companions/Movie%203.html regarding such use of theatre in which Newell has been involved, as well as, for instance, the technology and art approaches developed for Carnegie-Mellon University's Oz Project (http://www.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/oz.html).
[25] In her pre-forum talk.

Allowing adjustments to meet bespoke user needs

In constructing artificial companions, the potential to be adapted to different user needs should be considered as a fundamental capability. This would help to support people to tailor their companion and the way they use it to fit varying configurations of their own relationships with people, pets, technologies and other entities. It would also allow the AC to evolve with changing user requirements. Finding out about, and better understanding, such diversity in user requirements will be assisted by a user-led approach to AC design.

A core system for an AC could therefore be designed to meet the common needs of a substantial sector of the user base, but allow adjustability by user selection and controls (e.g. by including a software library of user-chosen options). For instance, the kind of digital learning companion proposed by Davies and Eynon (2007)—which seeks to support the needs of adult learners who have difficulties in other environments—could be adapted to fit different learning styles (e.g. Pask 1976).[26]

Policy to help shape appropriate artificial companion uses

The national and international policy significance of artificial companions and related AI developments is indicated by the support given by the Japanese government and industry to the development and use of such innovations to support its ageing population, thereby reducing the potential dependence on immigrant labour[27]. The opportunities and concerns outlined in this paper also raise other significant policy issues.

The targeting of artificial companions at supporting vulnerable groups in society could clearly help to close social discriminations and digital divides. However, this could also lead to such gaps widening if policy measures are not taken to ensure groups without the financial resources or skills and knowhow to make effective use of these systems are not given appropriate assistance.

Legal aspects, such as liability (e.g. see Wilks and Ballim 1990) were frequently mentioned. Wilks (2007: 8) emphasizes the practical significance of questions about where responsibility and blame may lie when an AC acts as a person's agent and something goes wrong. He explains: 'At the moment, Anglo-American law has no real notion of any responsible entity except a human, if we exclude Acts of God in insurance policies. The only possible exception here is dogs, which occupy a special place in English law, at least, and seem to have certain rights and attributions of character separate from their owners. If one keeps a tiger, one is totally responsible for whatever damage it does, because it is *ferae naturae*, a wild beast. Dogs, however, seem to occupy a middle ground as responsible agents, and an owner may not be responsible unless the dog is known to be of "bad character".' Wilks (ibid: 4) wonders whether AC developments could influence changes in the law so that things that are not human could be liable for damages. Until now, if a machine goes wrong it is always the maker or the programmer, or their company, which is at fault. Bryson (2000) argues that liability should always be firmly associated with a human,

---

[26] For example, users engaging in conversation with the Virtual Woman 3D animated image (see Appendix 1) can choose from a range of characteristics for its presentation.
[27] Referred to by Turkle in her pre-forum talk (see also, for example, Jervis 2006 and Kitano 2007).

although whether it is the owner/operator or designer who is to be liable would remain an issue to resolve.

Turkle expressed a fear[28] that many at the leading edge of digital developments have begun to become too complacent about the potential for ICTs to be used to heighten surveillance and control over citizens. She said she has noticed a trend for many digital technology enthusiasts to argue along the lines: 'All information is good. We are observed all the time. If we give up information now freely, say on social networks like FaceBook and MySpace, we will be less troubled later. So there is no need to worry about surveillance. And in any case if you have nothing to hide there is no need to be concerned'. However, a great deal of such information has been used to repress people and political movements, so it seems over-confident to imagine that no regime would ever misuse data within your, or your data's, lifetime.

This view could reduce pressure to establish and implement clear policy guidelines to protect citizens' privacy and other rights which could be infringed by increased digital surveillance and privacy intrusions. It is possible that the ability of ACs to gather intimate personal information for what are seen to be socially valuable purpose could add a new dimension to this kind of abuse. The way even primitive 'intelligent' artefacts can be, in Turkle's phrase, 'cheap dates' in convincing people of their emotional sincerity also creates new opportunities for those with harmful intent to commit deception and fraud.

Figure 7 summarizes some of the main policy issues raised at the forum.

Figure 7. Key policy issues on the creation and use of artificial companions

Policy area	*Key issues*
Digital divides	Support to ensure appropriate ACs are developed and made accessible to all individuals in vulnerable groups, such as the elderly, disabled, ill and learners with special needs
	Capacity and skill building to help vulnerable groups make informed decisions about which ACs they could choose to use (or not use)
	Divides between older generations and those who are growing up in an always-on digital world
	The degree to which plans to provide ACs to vulnerable groups may reduce pressure to look for other, perhaps better, solutions for their care
Legal	Liability for incorrect advice or actions by an AC
	Implications of giving 'informed consent' for an AC to act on behalf of a person (e.g. after death)
	Status of 'professional privilege' when an AC gains confidential information in a particular 'expert' role (e.g. as medical advisor)

---

[28] In her pre-forum talk.

Privacy and surveillance	Control over what biographical information is recorded by an AC; its protection from unwarranted access and change; its use after a person's death; and its eventual possible deletion.
	Ownership of the biographical narrative acquired by an AC.
	Rights of others (e.g. spouse, children, doctor, police, government, business partner) to access biographical information gathered by an AC.
	Use of ACs as a business strategy to acquire personal information from customers (e.g. if a supermarket were to give simple ACs for free to chat about shopping with lonely and bored customers).
	Sharing of information by an AC with other ACs.
Safety	Deliberate deception to gain a person's confidence, similar to spam email 'phishing' for bank details, capable of being used for economic and emotional exploitation of the user.
	Identity theft (even after death) using information provided confidentially by an AC's owner.
Socio-economic	Using ACs to meet social, economic and political requirements, such as caring for the elderly or reducing the need for immigrant labour.

The future: Can we live in harmony with artificial companions?

The boundary-spanning value of artificial companions

The wide range of issues outlined above indicate that the artificial companion concept provides a very useful tool or, speculum mentis, spanning many disciplines, for examining issues at the boundary where the human meets the artificial and computational. For instance, it challenges computer scientists to consider diverse issues related to notions like that of an AC being an extension of the user's 'self'. They therefore need to consider deeply in their design processes the relevant non-technical factors determining the outcome of their innovations. Likewise, it forces social scientists to think of design issues for the computer sciences, such as what a companion should look like—and what kinds of communication will be essential or sufficient, both between companions and in the kinds of social relations between humans and ACs that will tend to be fostered or inhibited.

One of the main future design challenges is to better understand and deal with the real world environment in which the user of a virtual artificial companion lives, or in which an embodied AC robot moves. Sloman (2007) claims that giving machines an understanding of the physical and geometrical shapes, processes and causal interactions that occur in an ordinary house 'is currently far beyond the state of the

art' as 'it has proved extremely difficult to give machines the kind of intuitive understanding required for creative problem-solving in novel physical situations'[29].

Future implications beyond AC developments, as such, were highlighted by Turkle's concern about the implication for notions of authenticity as more people become 'tethered selves', always intertwined within virtual and real worlds (see also Reeves and Nass 1996). This places forum discussions at the centre of broader work on the social implication of the Internet and other ICTs[30]. At the same time, the 'cyberintimacies' displayed in some people's reaction to relatively simple ACs indicate that these developments could be opening distinctive new challenges concerning the relationships between people and machines.

For instance, from her research Turkle (2006) reports the reaction of a woman sitting with the robot Paro, a seal-like device advertised as the first 'therapeutic robot' because of its ostensibly positive effects on the ill, the elderly and the emotionally troubled. Depressed because her son had abandoned her, the woman turned to Paro and, as she stroked it, she said: 'Yes, you're sad, aren't you. It's tough out there. Yes, it's hard.' Turkle observes: 'And then she pets the robot once again, attempting to provide it with comfort. And in so doing, she tries to comfort herself.' Levy (2007a) refers to the strength of affection and caring shown towards Tamagotchi artificial pets, which he says can be as strong as the affection shown for live pets.

At the start of the forum, Bill Dutton, Director of the OII and Co-Director of the e-Horizon's Institute, showed a video of Apple Computer's 'Knowledge Navigator'[31]. Conceived around 1987, this used actors to show what a software-based online desk-top personal assistant could eventually become: a simulated human prompting, advising and responding to a manager in the way humans would naturally communicate in an office. To some, this could have reinforced the view that AI has once again underachieved in failing to turn vision into reality. For others, it remains a goal researchers are getting ever closer to achieving.

However, it would be wrong to dismiss such prospects and their implications as belonging to the realm of science fantasy. The actual experiences of 'cyberintimacies' explored at the forum, as well as the advances that continue to be made to make ACs more believable—if not yet at Knowledge Navigator perfection—indicate that the related issues raised in this paper will need to be addressed in many practical ways in the coming years.

Wilks drew on the earliest science fiction vision of an artificial companion, in Mary Shelley's Frankenstein, to warn that we must 'take seriously the possibility that everything may turn out differently from what we expect and Companions, however effective, may be less loved and less loveable than we might wish'. This kind of dystopian fear, as well as the many Utopian promises of caring artificial companions,

---

[29] The time scale on enabling robots to operate with physical, as well as mental, human-like capabilities is indicated by the international RoboCup challenge, which aims to develop a team of fully autonomous humanoid robots that can win against the human world champion team in soccer—by the year 2050 (see http://www.robocup.org).
[30] See for example the range of topics being investigated at the OII (http://www.oii.ox.ac.uk) and Oxford university e-Horizons Institute (http://www.e-horizons.ox.ac.uk).
[31] See: http://www.billzarchy.com/clips/clips_apple_nav.htm

lay behind a core thread running through the forum. This was a search for an answer to the question: 'What are the appropriate relationships between people and artificial companions?'

Can a person engage in an I–Thou relationship with an artificial companion?

Answers to the question about what makes an appropriate artificial companion can challenge traditional ideas of authenticity and selfhood, for example by claims that it is leading to emergent new meanings of intelligence and emotions in ACs. These were crystalized at the forum around a discussion of whether human–computer relation can ever have the quality of the 'I–Thou' ethical relationship, as described by the philosopher Martin Buber (2004).

What is it to be human?

In her talk, Turkle reported that she had found many digital technology and AI enthusiasts have begun talking about people having an I-Thou relationships with cybersystems. She feels this profoundly misunderstands Buber's conception of such a relationship. For this, she says, there needs to be 'mutual understanding of deep life expectancies' and the biological lifecycle. But she believes this cannot happen 'if you think someone's there but there isn't'. From this perspective, it is insufficient for an artefact without genuine emotions to seem to offer sympathy (e.g. an AC acting as a psychiatric counsellor talking to a person about sibling rivalry when it never had a mother).

Dutton suggested we may be asking too much of ACs in suggesting they can become 'human-like' in their understanding of the emotional states of others, when people find those judgements so difficult to make accurately. On the other hand, there has been much evidence traditionally that human relationships can be effective even without any special or correct intentions on the side of one participant. For instance, a psychiatrist may not always be listening intently to a patient, who may still draw benefit from that therapeutic session[32]; and a confession to a Catholic priest could bring comfort even if the priest was thinking of something else but provided a seemingly appropriate response.

Nevertheless, as discussed above, developments like artificial companions are challenging traditional notions of the self, with computer systems in some cases becoming virtual extensions of a tethered self. Lowe (2007) pointed to the complexities of the relationship between the 'self' of a user and the 'other' of an artificial companion. He said that ACs 'will need to be distinct and independent enough for us to treat their advice, information, and representation of our rules as real constraints on our action, but sufficiently closely aligned and sensitive to our goals that we do not feel them as an imposition'.

O'Hara (2007) sees the human–AC relationship closer to the Johnson–Boswell, biographer–subject, friendship. In this, the AC acquires its owner's biographical knowledge from conversations with the person and other relevant sources, building up a perception of the subject's life, but not acquiring the life itself. He suggests an

---

[32] See also Weizenbaum (1984) for a discussion on the relatively primitive Eliza software program that engaged in what some users thought were believable psychiatric counselling.

insight into the emerging relationship with ACs can be gleaned from a comment by Žižek (1997: 137) on how people relate to the avatars they have created to represent themselves in a virtual world: 'on the one hand we maintain an attitude of external distance, of playing with false images. … On the other hand, the screen persona I create for myself can be 'more myself' than my 'real-life' persona … in so far as it reveals aspects of myself I would never dare to admit in real life'.

Wilks (2007) speculates that future artificial companions could act as an owner's agent: e.g. on the Internet or, further in the future, perhaps holding power of attorney in case of an owner's incapacity or, with the owner's advance permission, of being a source of conversational comfort for relatives after the owner's death. O'Hara (2007) foresees the possibility of an artificial companion functioning on the owner's behalf after the bodily death of that person (e.g. for overseeing and administering trust funds and the execution of wills).

What kinds of ethics should govern human-AC relations

Science fiction writer Isaac Asimov's (1993) three laws of robotics was an early attempt to create an ethical code to govern the relationships between people and intelligent machines[33]:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

One activity characteristic of artificial companions where such ethical guidance is likely to be of value is in controlling the sharing of sensitive information. This goes beyond preventing unwarranted access to questions like to whom (or what other AC) certain information can be disclosed, and in what circumstances.

The ethical stance of humans towards artificial companions was also highlighted at the forum. For instance, Bryson (2007) recalls: 'I was astonished during my own experience of working on (non-functional) humanoid robots in the mid 1990s by how many well-educated colleagues asserted immediately (without prompting) that unplugging such a robot would be unethical.' Bryson and Kime (1998) have argued that such deep concern with (or fear of) the well-being of AI artifacts results from a misattribution of human identity and therefore our empathy. Bryson's (2007) suggestion that the best way of viewing ACs or robots is as slaves because they are 'wholly owned and designed by us, we determine their goals and desires' opened a discussion on broader ethics of 'abusing' artificial artefacts, and whether a term like 'abuse' has a real meaning in this context.

---

[33] On robot ethics and rights, see also Roboethics.org (http://www.roboethics.org) and research at Birmingham University (e.g. http://www.cs.bham.ac.uk/research/projects/cogaff/crp/epilogue.html and Sloman 1978).

Cowie (2007) identified three main issues in the ethics of artificial companions' engagement with emotion:

1. Deception. What is the ethical status of building a device whose behaviour signals emotions that it does not actually feel, in any straightforward sense? Wilks (2006: 6) believes 'companions are not at all about fooling us as to their true natures', as in the case of an updated version of the 'Turing test' scenario where (an originally teletype-based) computer passed the test if it could make the user think a person was hidden at the other end of the line.

2. The 'lotus eater' problem. What if making a companion too emotionally engaging risks eroding the person's motivation for engaging with human beings, who are not always emotionally engaging. This could increase isolation instead of reducing it.

3. Technical limits of emotion detection. For the foreseeable future, it is likely that artificial companions will have a relatively modest ability to recognize emotion-related attributes. This could lead to misattributions that could seriously affect what happens to the person (e.g. if a false attribution of a negative mood in the AC's owner is fed by the companion into the medical or care system, such as by the AC calling—or failing to call—an emergency service in the wrong circumstance).

Meeting future real and virtual needs

As always in discussions on how to develop advanced innovations for an unknowable long-term future, the most frequent design advice on artificial companions was to make them as adjustable as possible to changing needs. Fulfilling Wilks' goal of providing a more friendly interface to the Internet to help manage its oceans of overflowing information would in itself be a significant contribution to giving people more control over ICTs. The more naturally human the dialogue between ACs and their users become, the more smoothly will ACs be able to adapt to the changing requirements of their owners over a long period, without the user having to continually learn new skills and knowledge—and forget old ones—in order to interact effectively with the artificial companion.

New forms of artificial life are also likely to emerge from bioengineering developments, including the introduction of biological innovations into traditional AI robotics[34]. A critical divide for the future could be the generational one, particularly in terms of issues like authenticity, as younger generations may find interacting with artificial beings more 'natural' and appealing than earlier generations. For instance, Turkle (2007) describes how the reaction of some young people to a real tortoise at a Darwin exhibition in New York suggests that, in certain circumstances, the emerging digital-savvy generation will think that 'aliveness doesn't seem worth the trouble'. She reports a 12-year-old girl adamantly stating: 'For what the turtles do, you didn't have to have the live ones.' Her father looked at her, uncomprehending: 'But the point is that they are real, that's the whole point.'

---

[34] For instance, see Ecobot (Appendix 1).

References

André, E. (2007) Towards More Sensitive Artificial Companions: Combined Interpretation of Affective and Attentive Cues. In OII (2007).

André, E., Rehm, M., Minker, W. and Buhler, D. (2004) Endowing Spoken Language Dialogue Systems with Emotional Intelligence. In André, E., Dybkjaer, L., Minker, W. and Heisterkamp, P. (eds), Affective Dialogue Systems, (Berlin: Springer Verlag), 178–187.

Asimov, I. (1993 [1950]) I, Robot (London: HarperCollins).

Avgerou, C., Ciborra, C. and Land, F. (2004) (eds) The Social Study of Information and Communication Technology: Innovation, Actors and Contexts (Oxford: Oxford University Press).

BBC News 24 (2007) My Special Partner. Available at: http://news.bbc.co.uk/1/hi/programmes/this_world/golden_years/4373857.stm

Berners-Lee, T. Hendler, J and Lasilla, O. (2001) The Semantic Web Scientific American 284(5), 34–43

Bickmore, T. W. (2003) Relational Agents: Effecting Change through Human-Computer Relationships, PhD Thesis (Cambridge MA: MIT). Available at: http://www.ccs.neu.edu/home/bickmore/bickmore-thesis.pdf

Boden, M. A. (2006) Mind as Machine: A History of Cognitive Science (Oxford: Clarendon Press).

Boden, M. A. (2007) Conversationalists, Maybe—But Confidants? In OII (2007),

Breazeal, C. (2002) Designing Sociable Robots (Cambridge MA: MIT Press).

Brooks, R. A. (2002) Flesh and Machines: How Robots Will Change Us (New York: Pantheon).

Bryson, J. J. (2000) A Proposal for the Humanoid Agent-builders League (HAL). In Barnden, J. (ed) AISB'00 Symposium on Artificial Intelligence, Ethics and (Quasi)Human Rights, 1–6. Available at: http://www.cs.bath.ac.uk/%7Ejjb/ftp/HAL00.pdf

Bryson, J. J. (2007) Robots Should Be Slaves. In OII (2007).

Bryson, J. J. and Kime, P. (1998) Just Another Artifact: Ethics and the Empirical Experience of AI. In Fifteenth International Congress on Cybernetics, 385–390.

Bryson, J. J., Martin, D., McIlraith, S. I. and Stein, L. A. (2002) Toward Behavioral Intelligence in the Semantic Web, IEEE Computer 35(11), 48–54.

Buber, M. (2004 [1923]) I and Thou (London: Continuum)

Cassell, J., Sullivan, J., Prevost, P. and Churchill, E. (2000) Embodied Conversational Characters (Cambridge, MA: MIT Press).

Collodi, C. (1996 [1922]) The Adventures of Pinocchio (Oxford: Oxford University Press)

Cowie, R. (2007) Companionship is an Emotional Business. In OII (2007).

Crevier, D. (1993) AI: The Tumultuous Search for Artificial Intelligence (New York: BasicBooks).

Davies, C. and Eynon, R. (2007) Some Implications of Creating a Digital Companion for Adult Learners. In OII (2007).

Dutton, W. H. (2004) Social Transformation in the Information Society (Paris: UNESCO Publications for the WSIS).

Dutton, W. H. (2005) The Internet and Social Transformation: Reconfiguring Access. In W. H. Dutton, B. Kahin, R. O=Callaghan and A. W. Wyckoff (eds), Transforming Enterprise (Cambridge, MA: MIT Press).

Eichner, T., Prendinger, H., André, E. & Ishizuka, M. (2007) Attentive Presentation Agents. In The 7th International Conference on Intelligent Virtual Agents (IVA), 283–294.

Ellis, P. M. and Bryson, J. J. (2005) The Significance of Textures for Affective Interfaces. In Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P. and Rist, T. (eds) The Fifth International Working Conference on Intelligent Virtual Agents, (Berlin: Springer), 394–404. Available at: http://www.cs.bath.ac.uk/%7Ejjb/ftp/iva05pme.pdf

Fisher, E. M. W. (1990) Personal Love (London: Duckworth).

Floridi. L. (2007) Philosophical Issues in Artificial Companionship. In OII (2007).

Helmreich, S. (1997) The Spiritual in Artificial Life: Recombining Science and Religion in a Computational Culture Medium, Science as Culture 6(3), 363–395.

Jacko, J. A. and Sears, A. (2003) The Human-computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (Mahwah, NJ: Lawrence Erlbaum Associates).

Jervis, C. (2006) Robot Revolution, E-Health Insider, 12 January. Available at: http://www.e-health-insider.com/comment_and_analysis/robot_revolution

Kitano, M. (2007) Japan Eyes Robots to Support Ageing Population, Reuters UK, 12 September.

Latour, B. (1993) We Have Never Been Modern (London: Harvester Wheatsheaf).

Law, J. and Hassard, J. (1999) (eds) Actor Network Theory and After (Oxford: Blackwell Publishers/The Sociological Review).

Levy, D. (2007a) Falling in Love with a Companion. In OII (2007).

Levy, D. (2007b) Love and Sex with Robots: The Evolution of Human–Robot Relationships (London: Duckworth).

Lighthill, J. (1973) Artificial Intelligence: A General Survey. In Artificial Intelligence: A Paper Symposium (London: Science Research Council).

Lowe, W. (2007) Identifying Your Accompanist. In OII (2007).

Mori, M. (1970) The Uncanny Valley, Energy, 7(4), 33–35 [Originally in Japanese as 'Bukimi no tani', trans. K. F. MacDorman and T. Minato].

Newell, A. (2007) Consulting the Users. In OII (2007).

Newell, A. and Simon, H. A. (1963) GPS: A Program that Simulates Human Thought, in Feigenbaum, E. A. and Feldman, J., Computers and Thought (New York: McGraw-Hill).

O'Hara, K. (2007) Arius in Cyberspace: The Limits of the Person. In OII (2007).

OII (2007) Artificial Companions in Society (Participant Position Papers) available at: http://www.companions-project.org/downloads/Companions_Position_Papers_20071026.pdf

Papert, S. (1993 [1980]) Mindstorms: Children, Computers and Powerful Ideas (New York, London: Harvester Wheatsheaf).

Pask, G. (1976) Styles and Strategies of Learning, British Journal of Educational Psychology, 46:II, pp 128–148.

Pelachaud C. (2005) Multimodal Expressive Embodied Conversational Agent. In MULTIMEDIA '05: Proceedings of the 13th Annual ACM International Conference on Multimedia (New York: ACM Press), 683–689. Available at: http://www.informatik.uni-trier.de/~ley/db/conf/mm/mm2005.html

Pelachaud, C. (2007) Socially-aware Expressive Embodied Conversational Agents. In OII (2007).

Picard, R. W. (1997) Affective Computing (Cambridge, MA: MIT Press)

Pulman, S. (2007) Towards Necessary and Sufficient Conditions for being a Companion. In OII (2007).

Reeves, B. and Nass, C. (1996) The Media Equation: How People Treat Computers. Television and New Media Like Real People and Places (Stanford, CA: CSLI Publications).

Rehm, M. and André, E. (2005) Where do they Look? GAze Behaviors of Multiple Users Interacting with an ECA. In T. Panayiotopoulos et al. (ed.), Intelligent Virtual Agents: 5th International Working Conference, IVA 2005 (Berlin: Springer), 241–252.

Rice, M., Newell, A. F. and Morgan, M. (2007) Forum Theatre as a Requirements Gathering Methodology in the Design of a Home Telecommunication System for Older Adults, Behaviour and Information Technology, 26(4), 232–331.

Romano D. M. (2007) The Look, the Emotion, the Language and the Behaviour of a Companion at Real-Time. In OII (2007).

Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) The Semantic Web Revisited, IEEE Intelligent Systems, 21(3), 96–101. Available at http://eprints.ecs.soton.ac.uk/12614

Sloman A. (1978) The Computer Revolution in Philosophy: Philosophy, Science and Models of mind (Brighton, UK: Harvester Press). Available at: http://www.cs.bham.ac.uk/research/projects/cogaff/crp/

Sloman A. (2006) Do Machines, Natural or Artificial, Really Need Emotions? (Birmingham: School of Computer Science, University of Birmingham). Available at: http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cafe04

Sloman, A. (2007) Requirements and Their Implications. In OII (2007).

Suchman, L. A. (1987) Plans and Situated Actions: The Problem of Human–Machine Communication (Cambridge: Cambridge University Press).

Suchman, L. A. (2006) Human-Machine Reconfigurations: Plans and Situated Actions Communication (Cambridge: Cambridge University Press).

Tanguy, E., Willis, P. and Bryson, J. J. (2007) Emotions as Durative Dynamic State for Action Selection. In Nijholt, A., Pantic. M., Pentland, A. (eds) Artificial Intelligence for Human Computing: Icmi 2006 and IJCAI 2007 International Workshops Banff, Canada, November 3, 2006 Hyderabad, India, January 2007 (San Francisco: Morgan Kaufmann), pp. 1537–1542. Available at: http://www.cs.bath.ac.uk/%7Ejjb/ftp/TanguyIJCAI07.pdf

Taylor, A. S. and Swan, L. (2007) Living with Intelligence? In OII (2007).

Turing, A. M. (1950) Computing Machinery and Intelligence, Mind, LIX, 236, 433–460.

Turkle, S. (1985) The Second Self: Computers and the Human Spirit (Cambridge, MA: MIT Press). The 20th anniversary edition was published in 2005.

Turkle, S. (2006), Diary, *London Review of Books*, 28 (8), 20 April, http://www.lrb.co.uk/v28/n08/turk01_.html.

Velásquez, J. D. and Maes, P. (1997) Cathexis: A Computational Model of Emotions, Proceedings of the First International Conference on Autonomous Agents, 518–

519. Available at:
http://portal.acm.org/toc.cfm?id=267658&coll=GUIDE&dl=GUIDE&type=proceeding&idx=SERIES134&part=series&WantType=Proceedings&title=AGENTS&CFID=6042890&CFTOKEN=37617687

Warnock, M. (1987) Memory (London: Faber and Faber).

Weizenbaum, J. (1984 [1976)]) Computer Power and Human Reason: From Judgment to Calculation (London: Penguin).

Whitby, B. (1996) Reflections on Artificial Intelligence: The Social, Legal, and Moral Dimensions (Oxford: Intellect Books).

Wilks, Y. (2006) Artificial Companions as a New Kind of Interface to the Future Internet. Oxford Internet Institute Research report No. 13 (Oxford internet Institute). Available at: http://www.oii.ox.ac.uk/research/publications.cfm

Wilks, Y. (2007) Introduction: On Being a Victorian Companion. In OII (2007).

Wilks, Y. (forthcoming)The Semantic Web as the Apotheosis of Annotation, but What are its Semantics? International Journal of Web Semantics. Available at: http://www.dcs.shef.ac.uk/~lucy/yw_pubs/yorick-wilks-semantic-annotation.pdf

Wilks, Y. and Ballim, A. (1990) Liability and Consent. In A. Narayanan and N. Bennun (eds.) Law, Computers and Artificial Intelligence (Norwood, NJ: Ablex).

Winfield, A. F. T. (2007) You Really Need to Know What Your Bot(s) are Thinking. In OII (2007).

Žižek, S. (1997) The Plague of Fantasies (London: Verso).

Zyga, L. (2007) Machines Might Talk with Humans by Putting Themselves in our Shoes, Physorg.com, 10 September. Available at: http://www.physorg.com/news108639466.html

Appendix 1. Resources

The following resources can help to explore in more depth the issue discussed in this paper. They are related to material provided and references at the e-Horizons forum on artificial companions.

Examples of artificial companions

Aibo. Sony's 'intelligent' robot dog companion/pet (http://support.sony-europe.com/aibo/index.asp).

BASIC (Believable Adaptable Socially Intelligent Character). Research project at Sheffield University Department of Computer Science developing synthetically generated characters to produce a believable graphical emotional response to a direct interaction of the user, another character in the BASIC world, or the emotionally charged atmosphere of the environment (http://www.dcs.shef.ac.uk/~daniela/gallery.html).

Beating the Blues. An interactive, computer-based therapy system developed at King's College London to assist people suffering from depression and/or anxiety. Based on Cognitive Behavioural Therapy (http://wwww.mentalhealthcare.org.uk/content/?id=98).

Ecobot. An energetically autonomous robot built at the Bristol Robotics Laboratory using Microbial Fuel Cell technology to extract electrical energy from refined foods such as sugar and unrefined foods such as insects and fruit (http://www.brl.ac.uk/projects/index.html).

ESP (Emotional-Social Intelligence Prosthesis Technology). Developed by MIT Media Lab's Affective Computing Group as a wearable system that can augment and enhance the user's ability to sense non-verbal cues (e.g. facial expressions and tone of voice). It uses common-sense knowledge about people that may not be natural for certain groups, such as those diagnosed with autism. (http://affect.media.mit.edu/projects.php?id=1935).

Furby. An 'intelligent' toy that looks like a furry owl. At first it speaks only its unique Furbish language, but it is programmed to speak less Furbish when it learns more English as it is used more. Voice recognition and increasingly complex facial movements are among the enhancements being introduced (http://www.hasbro.com/monkeybartv/default.cfm?page=SearchResults&criteria=furby).

Greta. 3D virtual agent, part of the HUMAINE project (Pelachaud 2007 and http://emotion-research.net/deliverables/D6f_v1.5%20Feb%2007.pdf).

KASPAR. A child-sized humanoid robot developed by the Adaptive Systems Research Group at the University of Hertfordshire to study human–robot interaction as part of the RobotCub Project. It can be used for developmental studies, including as therapeutic or educational tools to encourage social interaction skills in children with autism (http://kaspar.feis.herts.ac.uk).

Kismet. Developed by the Sociable Machines Project in MIT's Humanoid Robotics Group. Able to engage people in natural and expressive face-to-face interaction (http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html).

Laura. A virtual fitness health adviser that can ask the user questions and respond empathetically and encouragingly to the answers (see Bickmore 2003 and http://www.ccs.neu.edu/home/bickmore/agents/).

Nabaztag. A programmable Wi-Fi wireless-enabled talking device that looks like a rabbit. It has with movable ears and can speak information it downloads from the Internet, such as weather forecasts and email alerts (http://www.nabaztag.com).

Nursebot Pearl (Personal Robotic Assistants for the Elderly). An interdisciplinary university research initiative at the University of Pittsburgh and Carnegie Mellon University focused on robotic technology for the elderly, particularly mobile personal service robots (http://www.cs.cmu.edu/~nursebot/).

Paro. Developed by the Japanese National Institute of Advanced Industrial Science and Technology as a 'mental commitment robot'. It is designed to interact with human beings to make them feel emotional attachment to its embodiment as a cuddly baby seal. (http://paro.jp/english/index.html).

Primo Puel. A talking doll originally designed to be a substitute boyfriend for young single girls in the Japanese workforce, but has become unexpectedly popular among elderly people across Japan; has a vocabulary of a few hundred words and can talk, laugh and even ask for a kiss (BBC News 24 2007).

Radar (Reflective Agents with Distributed Adaptive Reasoning). Developments at Carnegie Mellon University's Robotics Institute focused on creating a cognitive assistant that embodies machine learning technology that is able to function 'in the wild'—by using technology that need not be tuned by experts, and for which the person using it need not be trained in any special way (http://www.radar.cs.cmu.edu/).

Roomba. Robot vacuum cleaner from iRobot Corp, whose seemingly autonomous movements can make it seem 'alive', at least to real pets (http://www.irobot.com).

RoboCup. An international challenge aiming to develop a team of fully autonomous humanoid robots that can win against the human world champion team in soccer— by the year 2050 (http://www.robocup.org).

Tamagotchi. A virtual creature that 'lives' on small screens housed in a small plastic egg. It can communicate its needs and 'grow' from a 'child' to 'healthy adult' if looked after, but dies if its needs are not met (http://www.tamagotchi.com).

Virtual Woman. Software that presents a 3D animated image of a woman who can engage the user in conversation. Users can choose characteristics of the presentation—such a ethnic type, personality and clothing—to personalize the image of their own Virtual Woman (http://virtualwoman.net).

Related research projects

Adaptive Systems Research Group, University of Hertfordshire. A multidisciplinary group studying artificial life, socially intelligent agents and AI (http://adapsys.feis.herts.ac.uk).

Affective Computing Research Group, MIT Media Laboratory. **F**ounded and directed by Rosalind Picard, the group focuses on computing developments that relate to, arise from or deliberately influence emotion or other affective phenomena (http://affect.media.mit.edu/index.php).

AMIDA (Augmented Multi-party Interaction with Distance Access). European Commission project addressing many scientific challenges in multimodal processing, focusing on live meetings with remote participants (http://www.onderzoekinformatie.nl/en/oi/nod/onderzoek/OND1320459/).

University of Bath. The AmonI (Artificial Models of Natural Intelligence) group is seeking to understand human and animal intelligence by engaging in research in the natural sciences, while using their knowledge as engineers to develop software tools and AI techniques to make their models more accurate and easier to develop (http://www.cs.bath.ac.uk/ai/AmonI.html). The University's Media Technology Research Centre studies computer technology for animation, graphics, image processing, music, rendering and virtual reality, including the Animating Virtual Humans project that has developed a Dynamic Emotion Representation (DER) for use in controlling the expressiveness of virtual characters (http://www.cs.bath.ac.uk/~pjw/media/avatars.htm).

Bristol Robotics Laboratory. A multidisciplinary venture from the University of Bristol and the University of the West of England. It includes work on the Ecobot and a range of developments in the Humanoid Robotics and Social Interaction group. The Laboratory leads the EPSRC UK public engagement network 'Walking with Robots', one aim of which is to create public debate about the ethical questions raised by intelligent robots in society (http://www.brl.ac.uk/projects/index.html).

CALLAS (Conveying Affectiveness in Leading-Edge Living Adaptive Systems). European Commission project designing and developing a multimodal architecture that includes emotional aspects (http://www.callas-newmedia.eu).

CALO (Cognitive Assistant that Learns and Organizes). US project led by SRI International's Artificial Intelligence Centre. Supports decision making by creating cognitive software systems that can reason, learn from experience, be told what to do, explain what they are doing, reflect on their experience and respond to surprise (http://www.ai.sri.com/project/CALO).

CARTE (Center for Advanced Research in Technology for Education). US research project based at the University of Southern California's Information Sciences Institute (ISI). Developing tools to support individualized language learning (http://www.isi.edu/isd/carte/proj_tactlang/).

Cathexis. A computational model for the generation of emotions and their influence in the behavior of autonomous agents (see Velásquez and Maes 1997).

COMPANIONS: European Commission project developing conversational artificial companions (including two demonstrators: a Senior Companion for elderly people and a Health and Fitness Companion). The COMPANIONS demonstrators will be agents or 'presences' that stay with the user for long periods of time, developing a relationship and 'knowing' its owners preferences (http://www.companions-project.org).

CoSY (Cognitive Systems for Cognitive Assistants). European Commission project involving seven centres in constructing physically instantiated systems that can perceive, understand and interact with their environment. They will also be able to evolve to achieve human-like performance in activities requiring context- and task-specific knowledge (http://www.cs.bham.ac.uk/research/projects/cosy/index.php).

Emotionally Intelligent Interface. Research led by Peter Robinson at Cambridge University. Provides a taxonomy of facial expressions and the emotions they represent, which has been used as the basis of a 'Mind Reading' DVD: an interactive computer-based guide to reading emotions from the face and voice (http://www.cl.cam.ac.uk/research/rainbow/emotions/).

HUMAINE (Human-Machine Interaction Network on Emotion). European Commission project developing 'emotion-oriented systems' that can register, model and/or influence human emotional and emotion-related states and processes (http://emotion-research.net).

INDIGO (Interaction with personality and dialogue-enabled robots). A European Commission project developing technology that could enable robots to perceive natural human behaviour and to act in ways that are familiar to humans (http://www.ics.forth.gr/indigo/)

Oz. Completed project at Carnegie Melon University that developed technology and art approaches to help create high quality interactive drama, based in part on AI technologies (http://www.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/oz.html).

RobotCub. A European Commission project which aims to build an open-source humanoid robot platform, such as KASPAR, for cognitive development research (http://www.robotcub.org).

Robot ethics. Roboethics.org is a group coordinating work on the application of ethics to the development and use of robots (http://www.roboethics.org). Aaron Sloman has led research in this field at the School of Computer Science, Birmingham University (e.g. see Sloman 1978 and http://www.cs.bham.ac.uk/research/projects/cogaff/crp/epilogue.html).

Sheffield University Department of Computer Science. Includes a range of work related to artificial companions, including cognitive systems, natural language processing and speech technology (http://www.shef.ac.uk/dcs/research)

UTOPIA. A consortium of Dundee, Abertay, Glasgow and Napier Universities funded by the Scottish Higher Education Funding Council to research the relationship between older people and technology (http://www.computing.dundee.ac.uk/projects/UTOPIA/).

Appendix 2. Forum participants

Elisabeth André, Professor, Multimedia Concepts and their Applications, University of Augsburg

Margaret Boden, Research Professor of Cognitive Science, Centre for Cognitive Science, University of Sussex

Joanna Bryson, University of Bath and Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria

Roberta Calzione, Department of Computer Science, The University of Sheffield

Roddy Cowie, Professor of Psychology, Queens University, Belfast and Co-Ordinator HUMAINE project

Chris Davies, University of Oxford Department of Educational Studies

Bill Dutton, Director Oxford Internet Institute and Co-director e-Horizons Institute

Rebecca Eynon, OII and University of Oxford Department of Educational Studies

Luciano Floridi, University of Hertfordshire and St Cross College, Oxford University

Joanie Gillespie, author *Cyberrules*

David Levy, CEO, Intelligent Toys, London

Will Lowe, Methods and Data Institute, University of Nottingham

Alan Newell, Queen Mother Research Centre for IT to Support Older People, School of Computing, University of Dundee

*Kieron O'Hara, Intelligence, Agents, Multimedia Group, School of Electronics and Computer Science, University of Southampton

Malcolm Peltu, Editorial Consultant, OII

Catherine Pelachaud, Professor, IUT de Montreuil, Universite de Paris 8, INRIA

Stephen Pulman, Professor of Computational Linguistics, Oxford University Computing Laboratory

Daniela Romano, Department of Computer Science, The University of Sheffield

Aaron Sloman, Honorary Professor of AI and Cognitive Science, School of Computer Science, University of Birmingham

Laurel Swan, Information Systems, Computing and Mathematics, Brunel University

Alex Taylor, Microsoft Research

David Traum, Institute for Creative Technologies, University of Southern California

Sherry Turkle, Abby Rockefeller Mauze Professor of the Social Studies of Science at MIT and Technology Director, MIT Initiative on Technology and Self

Yorick Wilks, Oxford Internet Institute, University of Sheffield and Director of the COMPANIONS project

*Alan Winfield, Bristol Robotics Laboratory, University of West of England

Provided papers for the workshop but were not able to attend.