CALONIS: An artificial Companion for the care of cognitively impaired patients

Yorick Wilks[1], **Jan Jasiewicz**[2]

Abstract. The paper concerns a prototype MCA or Companion (called CALONIS, a Roman soldier's servant) for a brain-damaged Veteran with no long-term memory intended to provide engagement, diversion and assistance, but with some possibilities of therapy and aids in cognitive testing. Ethical issues arise from how a Companion presents itself in such a situation and how it should manage its relationship with the patient and his or her carer, in particular in regard to information passing between them. The project began with a Wizard-of-Oz (WOZ) version of CALONIS and already at that stage interesting and potentially ethical issues arose in relation to the carer. We discuss the ethical issues in part by answers to fundamental criticisms of this strain of work raised by [1].

The full Companion prototype, beyond the WOZ, is based on a Senior Companion developed as part of a large-scale EU project, designed as a dialogue system to converse with an older person, eliciting knowledge from them about their past through the medium of photographs and so building a knowledge-base about the user's life. CALONIS, however, is designed for a quite specific type of user: a patient with Traumatic Brain injury with impaired executive functioning, short-term and long term memory problems, and difficulty initiating and maintaining conversations and carrying out simple tasks. The purpose of CALONIS is to achieve some level of relationship with the patient, a "stickiness", by any means and to use this so as to, for example, monitor the patient's movements so that he can indicate where he is going when he leaves the building etc. Initial results show a much improved level of engagement on the patient's part.

1 INTRODUCTION

The paper concerns a prototype MCA or Companion (called CALONIS, a Roman soldier's servant) for a brain-damaged Veteran with no long-term memory so as to provide diversion and assistance. Ethical issues arise from how a Companion presents itself in such a situation and how it should manage its relationship with the patient and his or her carer, in particular in regard to information passing between them. The project began with a Wizard-of-Oz version of CALONIS and already at that stage interesting and potentially ethical issues arose in relation to the carer.

[1] Florida Institute of Human and Machine Cognition, Ocala, FL, US. Email: **ywilks@ihmc.us**.

[2] James A. Haley Veterans' Hospital, Tampa, FL US. Email: **Jan.Jasiewicz@va.gov**.

The Companion prototype is based on a Senior Companion developed as part of a large-scale EU project [2] designed as a dialogue system to converse with an older person, eliciting knowledge from them about their past through the medium of photographs and so building a knowledge-base about the user's life, while at the same time providing diversion, news, jokes etc. so as to increase its attractiveness as a conversation partner. The Companion was intended as a partner for a specific individual and not a generic conversation device, its value was intended to be in its knowledge of its user's habits, choices and preferences. The technical basis of the Companion platform, also used in this prototype, include off-the-shelf ASR, TTS and face recognition from photos along with an original content extraction method, based in Information Extraction. In addition it uses a dialogue management model, based on semantic networks in a single stack, along with real time access to web information and a semantic web style inference system, as well as a model of emotional representation that attempts to bring the emotional state of the user back to a point in a two dimensional space. The emotional representation model has the ability to track the emotions of the user as they are being displayed by the system or vice versa. These latter aspects are highly experimental and were not fully tested in the original Companion prototype.

CALONIS is a Companion adapted for a quite specific type of user: a patient with Traumatic Brain injury with impaired executive functioning, short-term and long term memory problems, and difficulty initiating and maintaining conversations. The purpose of CALONIS is to achieve some level of relationship with the patient, a "stickiness", by any means and to use this so as to, for example, monitor the patient's movements so that he can indicate where he is going when he leaves the building etc. CALONIS is not designed essentially for therapeutic purposes, nor is it clear that therapy is of any value in such cases, but were any to arise that would be wholly positive. Engagement, then, is the first essential property of such a system (and we report some encouraging first evidence below) since everything else depends on that: useful assistance in everyday life, possible therapy and even cognitive testing, since such tests tend to be long and arduous for a patient with these impairments. A higher level of engagement than a human psychologist achieves would then make such tests more tolerable and more valuable. The patient has a full time carer, with some respite care, and it is a design requirement that the carer finds CALONIS acceptable and not any kind of interference in the relationship between carer and patient. In the case of patient X, the first to use CALONIS, the carer is his wife.

2. SHARKEY AND SHARKEY'S CRITICISMS

One way of isolating the ethical issues that arise with the kind of companion-carer we describe may be to address some of the fundamental criticisms of the whole idea in the classic paper by Sharkey and Sharkey [1] They ask at the outset: "Who would leave their granny or aging parents in the exclusive care of robots?" And we must disclaim immediately any plan for "exclusive care"; no one currently proposes this. On the other hand, the important ethical issue then is, is a companion carer better than nothing at all? Here one recalls the evidence of very large numbers of old people now alone in their homes with little or no human contact and only the television for company. The Japanese have a word Kodokushi [3] for the old who are left alone, die alone and are discovered dead long afterwards. Which could prompt the response question "Who would *not* leave their granny or aging parents in the care of robots, *rather than nothing*?"

Some of their argument consists in the use of phrases with powerful resonance but no clear meaning such as "a slippery slope towards authoritarian robotics". Again, there is a quite explicit assumption at work in their paper that human interaction is always to be preferred to interaction with robots: "It is the natural right of all individuals to have contact with other humans and to be allowed to socialise." I suggest there is no such right anywhere and no corresponding duty to provide company. It is all assumed self-evident, yet there is a long tradition (at least back to Donald Michie in the 1960s) of arguing that there are situations where robot interaction is to be preferred. Michie's example was arguing that drivers would prefer traffic lights (called robots in some dialects of English) to a policeman on point duty, on grounds of fairness and dispassion. I argued a more complex case [4] on dealing with state bureaucrats for welfare payments etc., where the quirks and prejudices of clerks are notorious. Moving closer to our interests here, might a Muslim not prefer a robot doctor to one of the opposite sex given that a female Saudi student died recently for lack of a female doctor, though males were plentiful at the scene.

Another set of their objections are in fact no different from problems that arise with human carers and humans with information about us: "authoritarian" diet control of the elderly, their complete knowledge of us from the web, just like NSA has and so on. There is a serious issue they raise as to privacy, and in particular what privacy means to someone seriously demented, and I will return to that in the body of the paper. They concede the reported benefit to the old of both real and artificial pets [5] but balance this with a long quote from [6] to the effect that robot pets are a delusion and "It requires sentimentality of a morally deplorable sort. Indulging in such sentimentality violates a (weak) duty that we have to ourselves to apprehend the world accurately. The design and manufacture of these robots is unethical in so far as it presupposes or encourages this".

It is hard to know how to respond to the smugness of this; it could come straight from Plato and his denial of any place in the Republic (chiefly in Book 10 of that work) to artists on the grounds that representation of nature was always delusory and unreal. It is, as has been always known, a disturbing and terrifying book which has been rightly rejected by human history, at least in the West. I know of no such duty to apprehend the world accurately, nor do artists, nor do makers of computer games or toys. This silly argument has nothing really to do with the old specifically nor the provision of Companions.

The "deceit" argument is clearly important to the Sharkeys, even though deceit is no essential part of the use of such Companion agents. Increasingly in the deployment of artificial agents we will not know whether it is human or not and we will not care: I receive many phone calls already from companies where I have no idea if the speaker is real or a computer. Like most people, I do not care: there is no deceit and deciding that issue in each case serves no purpose. There is no reason to think robot carers will be different and, if someone then responds that it will matter if we have to decide ethically how to treat one, then I respond to that that we will at that point need an ethics of how to treat automata, not an accusation of deceit. To assume the old or demented are different from the rest of us is to infantilise them, a danger that I do not feel the Sharkeys are always aware of.

As they sum this argument up: "One the one hand, if you believe that all deceit is morally wrong, then you will categorically oppose the use of care robots. On the other hand, you may believe that the beneficial consequences of colluding in the deceit of an elder is worth a little untruth – 'they are never going to know anyway and so what is the harm.' But even then we still need to consider whether the benefit to the elders outweighs a potential erosion of society's moral standards."
To which the summary answer is: there is no reason to believe deceit is involved in such deployment so the utilitarian argument for it is irrelevant. And again, and more generally, if one's concern was the "erosion of standards" this might well not be the best place to look, lest all science fiction is dismissed along with much else in our culture.

Other arguments in the paper are more easily dealt with: a Reuters robot [7] is said not to engage the old, which is poor performance not ethics. They cite figures for the lonely old at home and say that "Safe robot care could exacerbate this situation" which is hard to follow, especially give the spectre of Kodokushi. Finally, they cite [8] work which does show the increased engagement with others by elders after exposure to the AIBO robot dog, rather along the lines we report here with the use of the WOZ at Tampa, but they balance this with claims from care workers and others that "emotionally engaging with the artefacts was demeaning, patronising and inappropriate." Given the flood of cases in the UK recently of careworkers being jailed for abuse of the elderly (e.g. Carers) one might begin to see beyond fear for their own jobs that many see in automation to a worse world where a robot carer may well be safer bet than a careworker.

3. ETHICAL ISSUES MORE GENERALLY

In an earlier paper [9] we distinguished three clusters of notions associated with Machine Ethics in connection with this kind of artificial companion: first, issues concerning how an artificial companion should itself be treated; secondly, design issues in the construction of a Companion and what values if any should be maximised in its interactions and, thirdly, issues to do with aspects of the decisions the Companion itself takes in its interactions with a user. In this paper we focus on the third type, which is closer to the standard usage of "Machine Ethics".

However, the three are not totally independent in the context we describe: for example, how the patient treats the Companion may depend on whether he views it as a person or as a machine (and in a WOZ environment, he believes it to be a machine when it is not). However, if it turns out that the patient shows much more willingness to interact with the WOZ-Companion while believing it to be a machine, it is not clear that he should be told by the Companion that it is, in fact, a person. Again, if the high levels of engagement we have observed with the WOZ-Companion (see below, for these are higher than he achieved with a real person in some test conversations) are continued with the automated Companion, it would seem that the Companion should not attempt to pose as a person. If so, this strategy for introducing the automated Companion would bypass many of the Sharkey and Sharkey criticisms that rest on the presence of deceit. All this rests on the assumption that a patient with this degree of cognitive impairment can fully grasp the machine-person distinction in the first place. This condition may also make many of the ethical arguments to do with privacy somewhat moot: it is unclear that privacy has much meaning for someone with little or no short term memory, since they may not be able to have access to any information that they would wish to keep from others---an essential condition of privacy. Their ability to function at all may in fact rest on their carers knowing everything about them there is to know, which makes the notion of privacy (and any associated ethical issues) meaningless.

On Machine Ethics issues in general we are arguing from a position---one there is no space to establish here---that much writing on these issues from AI-researchers (e.g. [10][11]) has overemphasised the role of rational decision-making in ethics in a way consistent with the centrality of machine reasoning in the traditional AI cognitive paradigm. We have suggested [9] that, on the contrary, and somewhat in line with the moral sentiment tradition of British philosophy (e.g. [12][13][14]) the source of ethics is in the emotions rather than the reason and, given how AI research on emotion and dialogue (e.g. [15] has developed considerably over the last decade, this is a position that it is now reasonable to seek to embody in a Companion as a way of expressing and analysing ethical concepts within dialogue interchanges with an automaton.

4. WORKING WITH CALONIS

Initially, we deployed CALONIS with patient X in a WOZ mode, so as to gather data on his responses in order to develop the full prototype from that corpus of interactions. However, and to our surprise and that of his therapists, patient X responded far better to the WOZ system than when responding to the same utterances from a person. This was confirmed by counts of non-yes/no answers being greater for the WoZ than for the same human conversation. The patient's wife/carer said she would like to have had those conversations with him, which she never did, even though he had said in the dialogue, and in the wife/carer's hearing, that he was not married. This interchange clearly illustrates the ethical problems that will arise when the patient and carer are having separate, not overheard, conversations with CALONIS, and the issue of what should not be reported back from the patient to the carer.

In the CALONIS prototype we can distinguish immediate goals and possibilities (to be installed in the current prototype) from longer-term ones. The nature of patient X, and those like him, makes certain aspects of computer dialogue management easier than in the standard case: for example, mixed initiative is hardly necessary since the patient is happy for the system to monopolise the conversational initiative; he never initiates any topic at all. Again, utterance repetition that might lead to boredom is less of an issue with a patient with impaired short-term memory. At the moment we are testing substitution of the prototype within the WOZ environment –the classic Turing test scenario!-----to see if the measures of engagement listed about shift at the point of substitution.

We shall also start to deploy the standard Companion scenario of discussion of images of the past to see if any recall is possible from such conversations with the prototype. An important issue here will be the consistency, and possibly truth (as defined by the carer) of the "memories" so elicited, since patient X will always reply something, even if quite random, as well as the consequent patient well-being, as defined by his willingness to continue interacting.

In the longer term, we shall incorporate dialogues between Companion prototype and the carer, although separately and not as three person-dialogues, since the theory of multi-party speech acts almost certainly makes that impractical. However, the ability to conduct some form of dialogue with both patient and carer is an ideal test bed for ethical issues of the limits on "cross information" transfer between them. A more adventurous goal will be the incorporation of a more systematic person-model within CALONIS, modelling the patient, the carer and their beliefs about each other of the kind set out in [16] and which we have implemented concurrently within other projects. A key assumption of such a system is that machine belief is defined by the possibility of holding alternate belief structures; in which sense an ATM does not *have beliefs* as it has only a single possible view of the state of my bank account. This fits closely with McDermott's claim that ethical belief is defined (as a necessary feature of machine agents to be ethical) as the possibility of contemplating alternative courses of action between which an agent must choose.

ACKNOWLEDGEMENT

This work is supported by grants from the Veterans Administration and from Ubisense plc.

REFERENCES

[1] Sharkey, N. and Sharkey, A. Living with robots: ethical tradeoffs in eldercare, In Wilks, Y. *Artificial Companions in Society: scientific, economic, psychological and philosophical perspectives.* Amsterdam: John Benjamins. (2010)

[2] Wilks, Y. Artificial Companions, In *Artificial Companions in Society: scientific, economic, psychological and philosophical perspectives*. John Benjamins: Amsterdam. (2010)

[3] https://en.wikipedia.org/wiki/Kodokushi

[4] Wilks, Y. The elimination of the bureaucracy, In *Theoria to Theory*. Vol 1. (1967)

[5] Libin, A., and Cohen-Mansfield, J., Therapeutic robocat for nursing home residents with dementia: Preliminary inquiry, *American Journal of Alzheimer's Disease and Other Dementias,* 19, 111-116, (2004)

[6] Sparrow, R., The march of the robot dogs. *Ethics and Information Technology*, 4, 305–318. (2002).

[7] http://www.reuters.com/article/2007/09/20/us-japan-ageing-gadgets-idUST29547120070920?sp=true
[8] Kanamori, M., Suzuki, M., & Tanaka, M., Maintenance and improvement of quality of life among elderly patients using a pet-type robot. *Japanese Journal of Geriatrics*, v. 39, p. 214-8. (2002)
[9] Wilks, Y. Cognitive issues of sentiment in Machine and Human Ethics, In *Proc. Workshop on Moral Cognition and the Theory of Mind*, AISB-IACAP-workshop 2012 Birmingham, (2012)
[10] McDermott, D. Why ethics is a high hurdle for AI. In *Proc. North American Conference on Computers and Philosophy (NA-CAP)* Bloomington, Indiana, and see his contribution to Anderson, M. and Anderson, S. (2011) *Machine Ethics:* Cambridge University Press: Cambridge. (2008)
[11] Thompson, H. S. Computational systems, responsibility and moral sensibility. *Technology in Society*. (1999)
[12] Hume, D. *An Enquiry Concerning the Principles of Morals*. David Hume, *Essays Moral, Political, and Literary* edited with preliminary dissertations and notes by T.H. Green and T.H. Grose, Longmans, Green : London. (1751/1907).
[13] Smith, A. *The Theory of Moral Sentiments*. Millar: London. (1759)
[14] MacIntyre, A. *After Virtue*, second edition, Duckworth, London. (1985).
[15] Gratch, J. and Marsella, M. Modeling the cognitive antecedents and consequences of emotion, *Journal of Cognitive Systems Research*, vol. 10, no. 1, 1-5, 2009
[16] Wilks, Y., and Ballim, A. Liability and Consent. In N. Narayanan & T. Bennun (eds.) *Law, Computers and Artificial Intelligence*. Norwood, NJ: Ablex. (1990).