GRESHAM COLLEGE

6 NOVEMBER 2018

# MAKING INFORMATION PERSONAL: ARTIFICIAL COMPANIONS

PROFESSOR YORICK WILKS

One thing we can be quite sure of is that artificial Companions are coming; in a small way they already arrived some decades ago and millions of people have already met them. The Japanese toys Tamagochi (literally "little eggs") were a brief craze in the West that saw sensible people rushing home to play with their Tamagochi so it would not "pine", where pining meant sad eyes and icons on a tiny screen, and playing with it meant pushing a feed button! The extraordinary thing about the Tamagochi (and later the US Furby) phenomenon was that sensible people felt guilt about their behaviour towards a small cheap toy that could not even speak.

The brief history of those toys said a great deal about people and their ability to create and transfer their affections to inanimate objects. This phenomenon is almost certainly a sign of what is to come and of how easily people will find it to identify with and care for automata that talk and appear to remember who they are talking to. This lecture is about what it will be like when much more sophisticated objects are available. Since most of the basic technologies such as speech recognition and simple machine reasoning and memory are already in place, this will not be long. Simple robot home helps are already available in Japan but we only have to look at them and their hard plastic exteriors, and their few tinny phrases about starting the dishwasher, to realise that this is almost certainly not how an ideal Companion should be.

**Companion Scenarios: Senior and Junior Companions, Dragomen**

Companions are not about fooling us, because they will not pretend to be human at all. Imagine the following scenario: An old person sits on a sofa, and beside them is a soft toy or a large furry handbag, which we shall call a Senior Companion; it is easy to carry about, but much of the day it just sits there and chats. Given the experience of Tamagochi, and the fact that old people with pets survive far better than those without, we will expect this to be an essential lifespan and health improving object to own. The elderly population of the EU and the US is the most rapidly growing segment of the population, and one relatively well provided with funds to buy technology.

Other Companions are just as plausible as this one: perhaps a Junior Companion for children, that would most likely take the form of a backpack, a small and hard to remove backpack that always knew where the child was, and which saw it safely across roads and talked to it in elementary French, perhaps, on the way to school. Or consider an artificial *Dragoman* -- the old Ottoman interpreter and guide for travellers -- who, on your vacation, would not only translate and guide you to the best sites, but could sit and converse for you with foreigners you encountered, showing them just the right children's' pictures and displaying theirs to you.

A large proportion of today's old people are effectively excluded from information technology, the web, the internet and some mobile phones because "they cannot learn to cope with the buttons". This can be because of their generation or because of losses of skill with age: there are talking books in abundance now but many, otherwise intelligent, old people still cannot manipulate a TV control or a mobile phone, which has too many small controls for

them with unwanted functionalities. All this is well known and yet there is little thought as to how our growing body of old people can have access to at least some of the benefits of information technology without the ability to operate a PC.

After all, their needs are real and pressing, not just to have someone to talk to as they increasingly live alone, but to deal with correspondence from public bodies, such as councils and utility companies demanding payment, or with the need to set up times to be visited by nurses or relatives by phone, or how to be sure they have taken their pills when keeping any kind of diary may have become difficult, as well as deciding what foods to order, even when a delivery service is available via the net but which is sometimes difficult in practice for them to use.

One can see how a Companion that could talk and understand and also gain access to the web, to email and a mobile phone could become an essential mental prosthesis for an old person, one that any responsible society would have to support. But there are also aspects of this which are beyond just getting information, such as having the newspapers blown up on the TV screen till the print was big enough to be read, and dealing with affairs, like paying bills from a bank account, and remembering the plots of TV serials and reminding their owners of the plot.

It is reliably reported that many old people spend much of their day sorting and looking over photographs of themselves and their families, along with places they have lived and visited. This will obviously increase as time goes on and everyone begins to have access to digitised photos and videos throughout their lives. One can see this as an attempt to establish the narrative of one's life, and to make sense of it: it is what drives the most literate segment of the population to write autobiographies (if only for their children) even when, objectively speaking, they may have lived lives with little to report. But think of what will be needed if a huge volume of digital material is to be sorted, the kind that everyone now acquires?

One can see all this as democratising the art of autobiography: it will mean far more than simply providing computational ways in which people can massage photos and videos into some kind of order on a big glossy screen: it will require a guiding intelligence to provide and amplify a narrative that imposes a time order. Lives have a natural time order, but this is sometimes very difficult to impose and to recover for the liver; even those with no noticeable problems from aging find it very hard to be sure in what order two major life events actually happened: "I know I married Lily before Susan but in which marriage did my father die?"

This family example is to illustrate how an artificial agent might assist in bringing the events of a whole human life, whether in text or pictures, into some single coherent order. This is the kind of thing today's computers can be surprisingly good at, although it is a very complex and abstract notion, that of the time ordering of events, which can be simple (I know James was born before Ronnie) or only what is called a partial order in some situations ( I know my brother's children were born after my marriage and before my wife died, but I'm not sure in what order they arrived). These may seem odd or contrived but they do represent real problems at the border of memory and reasoning for many people, especially the old.

**Companions as a Core AI Project**

The much-advertised home question answerers *Siri* and *Alexa* are the first commercial attempts to get talking Companions into our lives, but the notion has been around for decades, as with everything in AI. Companions are a crucial form of AI because they embody so much of what is central to the AI project: language, understanding, reasoning, empathy, planning and so on. Some Companions might be robots in the future, but there is no need for them to have human form, as opposed to being a toy or a phone. I will be concerned here with aspects of Companions such that embodiment is a secondary matter, provided they can converse with an owner and can reach out to the world via the internet for information and to establish action and control.

We can distinguish Companions from conversational internet agents that carry out specific tasks, such as train and

plane scheduling and ticket ordering applications of speech dialogue. Those go back to early MIT systems in the 1990s, and have now evolved into such things as the intelligent microwaves (recently reported in *Wired*) which will actually fill an important niche, like a more intelligent TV controller, since the complex systems of buttons on microwaves and TV monitors are often very hard for people to manage, and would be much better replaced by a conversational system. But we shall mean something more general by Companion: an entity that in principle knows all about you.

I also want to separate a Companion from chatbots, of the kind you now meet giving help on airline web sites: Alexa and Siri are simply more sophisticated versions of chatbots, with no memory or real knowledge of their owners; for them each new input starts them all over again, even though they embody far more sophisticated techniques like answering questions from the web.

I take the distinguishing features of a Companion agent to be:

1) that it has no central or over-riding task and there is no point at which its conversation is complete or has to stop, although it may have some tasks it carries out and completes in the course of a conversation;
2) That it should be capable of a sustained discourse over a long-period, possibly ideally the whole life-time of its principal user;
3) It is essentially the Companion of a particular individual, its principal user, about whom it has a great deal of personal knowledge, and whose interests it serves—it could, in principle, contain all the information associated with a life;
4) It establishes some form of relationship with that user, if that is appropriate, which would have aspects we associate with the term "emotion";
5) It is not essentially an internet interface, but since it will have to have access to the internet for information (including everything about its user and their media use and searches) we may as well take it to be a kind of internet agent.

We can now ask questions about a Companion so defined, such as:

i) What aspects of a relationship should one aim at with a Companion, in terms of such conventional psychological categories as emotion, politeness, affection etc.?
ii) Even if it is not a robot, in the sense of a free-moving entity, should it have a screen, and should it have a visible avatar for communication, whether human, animal or abstract?
iii) Need it have one identifiable personality, or perhaps several, and should the user be able to choose the Companion's personality or shift between them if there are several—making a Companion lively and chippy one day, and subservient the next?
iv) Does the Companion have any goals of its own, beyond carrying out a user's commands, if that is possible: should there be other overriding ethical constraints on what can be commanded, such as avoiding harm to the user, even if requested? Should there be ethical constrains *on the user* as to how the Companion can be treated?
v) What safeguards are there for the information content of such a Companion, in the sense of controlling access to its contents by the state or a company, and how should a user best provide for its disposal in case of his/her own death or incapacity?
vi) What if anything does a Companion have to *know* to be plausible, and does it need a certain level of inference and memory capacity over the material of past conversations with the user?

Let us look at these issues in turn:

### i) Emotion, politeness and affection

Cheepen and Monaghan presented results some years ago that customers of automata such as ATMs, are repelled by excessive politeness and endless repetitions of "thank you for using our service", because they know they are dealing with a machine and such feigned sincerity feels wrong. This suggests that politeness is very much a matter of judgment in certain situations, just as it is with humans, when inappropriate politeness is encountered.

We know, since the original work of Nass and Reeves in the 1990s that people display some level of feeling for the simplest machines, even their own PCs in the original experiments, where people avoided criticizing the performance of their own PCs if they could! And David Levy has argued persuasively that the trend is towards high levels of "affectionate" relationships with machines in the next decades, as realistic hardware and sophisticated speech generation make machine interlocutors increasingly lifelike and increasingly available for physical contact, including sex, but also simply affection.

The AI area "emotion and machines" is somewhat confused and contradictory: it has established itself as more than an eccentric minority taste, but as yet has little concrete to show beyond procedures for detecting "sentiment" in text and, even though programs to do that have been in great commercial demand, they rest on little more than detecting the use of certain emotionally "loaded" words. This strand of work began as "content analysis" at the Harvard psychology department many decades ago and, while prose texts may offer enough length to let a measure of sentiment to be assessed, this is not plausible with short dialogue interactions. That technology rested almost entirely on the supposed sentiment value of individual words, which ignores the fact that their value is so content dependent. "Cancer" may be marked as negative word but the utterance "I have found a cure for cancer" is positive and detecting that fact and the appropriate response to it rests on the ability to extract information way beyond single terms. Not being able to do that has led to many of the classic foolishnesses of chatbots such as congratulating people on the death of their relatives, and so on.

At deeper levels, there are conflicting theories of emotion for automata, not all of which are consistent and which apply only in certain kinds of discourse. So, for example, the classic theory of Marsella and Gratch that emotion is a response to the failure or success of a machine's plans covers only those situations that are clearly driven by plans. But Companionship dialogue is not always closely related to plans and tasks; people just like to chat much of the time and without wanting anything done.

All this makes many emotion theories look primitive in terms of developments elsewhere in AI. John Wisdom once said of philosophical discoveries that they are often the "running of a platitude up a flagpole", and theories of emotion have something of that in them, but there is no doubt that there has been progress in incorporating emotion into artificial devices and that that will be very important to human users of the technology.

Not all successful emotionally aware Companions have language. One of the most attractive of all artificial pets is the Japanese PARO seal: it is furry and has no language, but wiggles like a real animal or baby. Its secret is its many servo motors under the skin that give it an animal-like feel. and let it give pushback, or feedback, when held. The notion of *feedback* is an old one, going back to cybernetic ideas and in particular to Wiener's notion that activities like walking are only possible because of constant information feedback from our "servo" muscles in contact with the ground and sent to the brain. Wiener was emphasizing information feedback, as opposed to the "haptic" transfer from muscles, and in a computational setting everything must at some stage reduce to information expressed digitally in a machine.

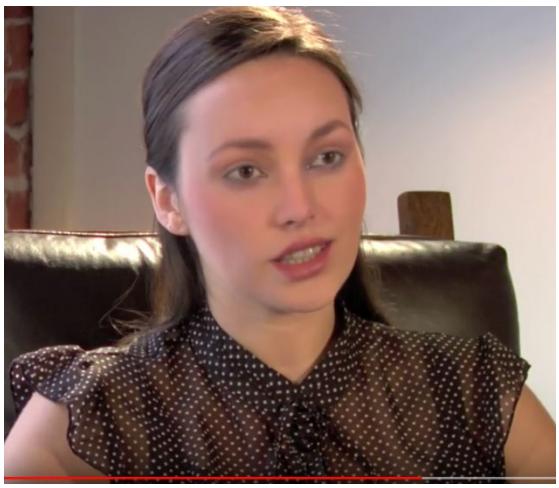
The French *Nabaztag* rabbit toy, in its original design, glowed in a number of colours to indicate the feelings of the sender (such as blue for "sad") and two Nabaztags and their respective owners, usually far apart, would be a classic feedback loop, but with no language involved, just the mutual sending of feeling by colour. The importance of these early Companion toys is that they make emotion central, not peripheral, to communication and relationships rather

than language. Everyone knows that in relationships with pets, a central relationship for many, this is the case: strong emotions are aroused by actions like stroking, but there is no verbal content.

An important issue to be settled in a Companion's design is whether a Companion should invariably try to cheer a user up if miserable i.e. be sad with a sad user and happy with a happy one? There is no general answer to this question and, indeed, in an ideal Companion, which method should be used would itself be a conversation topic e.g. "Do you want me to cheer you up today or would you rather stay miserable?".

### ii) What should a Companion look like?

Personally, I have an affection for a faceless Companion—the proverbial furry handbag, warm and light to carry, chatty but with full internet access and probably no screen. However, this may be a minority taste; and such a Companion could always take control of a nearby screen or a phone if it needed to show itself. If there is to be a face, the question of the *uncanny valley effect* always comes up. The phrase was due to Mori who argued that people get more uneasy the closer something artificial is to ourselves. I personally do not feel this, and cannot feel it with an avatar so good that one cannot be sure it is artificial. This is what I feel about the *Emily* avatar from a Manchester company, who at the end of her YouTube video takes her own face off!



It may be worth making here a small clarification about the word "avatar" that sometimes distorts discussion: researchers often use the word to mean any screen form, usually two-dimensional, that simulates a human being, but not any particular human being. On the other hand, in the virtual reality and game worlds, such as Second Life, an avatar is a manifestation of a particular human being, an alternative identity that may or may not be similar to the owner in age, sex, appearance etc. These are importantly different notions and confusion can arise when they are confused. However, in the case of a long-term computer Companion, it should perhaps also train its own voice in imitation of its owner, since research shows that more successful computer conversationalists are as like their owners as possible.

### iii) One Companion personality or several?

Some have argued that having a consistent personality is part of being a Companion, but one could disagree and argue that, although that is true of people—multiple personalities being a classic psychosis—there is no reason why we should expect this of an artificial Companion. Perhaps a Companion should have a personality adapted to its particular relationship to a user at a given moment: one might want a Companion to function at times as a gym trainer, in which case its having a rather harsh attitude might be the best one. It might be better to give a user access to, and some control over, the display of a multiple-personality Companion, something one could think of as a plural "agency" of Companions, a sort of caring subcommittee, rather than a single "agent", all of which shared access to the same knowledge of the world and of the state and history of the user.

### iv) What must a Companion know?

There is no clear answer to this question: dogs make excellent Companions and know nothing. On the other hand, it is hard to relate over a long term to an interlocutor who knows little or nothing and has no memory of what it or you have said in the past. It is hard to attribute much personality to an entity with no memory and little or no knowledge.

Much of what a Companion knows about its owner it should elicit via conversation; yet much could also be gained from publicly available sources, such as going off to Facebook to find out who its user's friends are. Current language technology allows a reasonable job to be made of going to Wikipedia for general information when, say, a world city is mentioned; the Companion can then glean something about that city there and ask a relevant question such as "Did you see the Eiffel Tower when you were in Paris?" which again gives a plausible illusion of general knowledge.

John McCarthy, an AI founder at Stanford, maintained fifty years ago that the real challenge for AI was not having exotic or detailed knowledge but common-sense knowledge, what exists below our levels of consciousness, such as that dropped thing fall, and fingers go into water when pushed but not into tables. Some of this can be coded in the logical inference rules a Companion will need, such as that sisters share parents, but much of it is below the level of straightforward rules, which is exactly what led the philosopher Hubert Dreyfus to argue in the 1960s that plausible AI would need the ability to learn as we do by growing up, rather than by using existing forms of machine learning or hand-coding. However, the great improvements in such learning in recent years, from speech recognition to machine translation, suggests that the jury is still out on this, even if the methods that have proved successful in computers are clearly not those humans themselves use.

## Another Companion Paradigm: The Victorian Companion

A colleague once said to me that James Boswell was a clear case of an inaccurate Companion: his account of Johnson's life is engaging but probably exaggerated, but none of that now matters. Johnson has become *Boswell's* Johnson, by and large, and his Companionship made Johnson a social property in a way he would never have been without his biographer. This observation brings out some of the complexity of Companionship, as opposed to a mere amanuensis or recording device, and its role between the merely personal and the social.

The first Artificial Companion is, of course, Frankenstein's monster in the 19C. That creature was dripping with emotions, and much concerned with its own social life:

> *"Shall each man," cried he, "find a wife for his bosom, and each beast have his mate, and I be alone? I had feelings of affection, and they were requited by detestation and scorn. Man! you may hate; but beware! your hours will pass in dread and misery, and soon the bolt will fall which must ravish from you your happiness for ever (Shelley, 1831, Ch. 20).*

This is clearly not quite the product that any modern computer Companion would be aiming at but, before just dismissing it as an "early failed experiment", we should take seriously the possibility that things may turn out differently from what we expect and Companions, however effective, may be less loved and less loveable than we might wish. Some have argued that we must actually find out what kinds of relationship people want with Companion entities, as opposed to being technologists and just deciding that for them, and then building what we believe they want.

It is no longer fashionable to explore a concept by reviewing its various senses, but the main Google-sponsored when searched for "Companions" still announces "14.5 million girls await your call?" For others, a Companion is still primarily a domestic animal, and pet-animals still play a key role in the arguments on what it is to be a Companion: especially the features of memory, recognition, attention and affection, found in dogs but rarely in snakes or newts.

Let us jump to another sense of the word and spend a few moments reminding ourselves of the role of the Victorian Lady's Companion. Forms of this role still exist, as in the web posting not too long ago:

> *Companion Job*
> *posted: October 5, 2007, 01:11 AM*
>
> *I Am a 47-year-old lady looking seeking a position as Companion to the elderly, willing to work as per your requirements. I have been doing this work for the past 11 yrs. very reliable and respectful.*
> *Location: New Jersey*
> *Salary/Wage: Will discuss*
> *Education: college*
> *Status: Full-time*
> *Shift: Days and Nights*

This role has now become more closely identified with the social services than it would have been in Victorian times, where the emphasis was on company, preferably educated company and diversion, rather than care. However, this was not always a particularly desirable or even tolerable role for a woman. Fanny Burney refers to someone's Companion as a "toad-eater" which Grose (1811) glosses as:

> *A poor female relation, and humble Companion, or reduced gentlewoman, in a great family, the standing butt, on whom all kinds of practical jokes are played off, and all ill humors' vented. This appellation is derived from a mountebank's servant, on whom all experiments used to be made in public by the doctor, his master; among which was the eating of toads, formerly supposed poisonous. Swallowing toads is here figuratively meant for swallowing or putting up with insults, as disagreeable to a person of feeling as toads to the stomach.*

But one could nevertheless, and in no scientific manner, risk a listing of features of the ideal Victorian Companion:

1. Politeness
2. Discretion
3. Knowing their place
4. Dependence
5. Emotions firmly under control
6. Modesty
7. Wit
8. Cheerfulness
9. Well-informed
10. Diverting
11. Looks are irrelevant
12. Long-term relationship if possible
13. Trustworthy
14. Limited socialization between Companions permitted off-duty.

The emphasis in the list is on what the self-presentation and self-image of a tolerable Companion should be; its suggestion is that overt emotion may not be what is wanted at all. It is important to say this because many research Companions now being produced press their explicit "emotion" on an audience all the time, and I think this may be a mistake.

On the other hand, the pet-Companion analogy could suggest that overt demonstrations of emotion are sometimes desirable and are sought by pet owners, especially with dogs. Language, however, disguises emotion as much as it

reveals it, and its ability to please, soothe and cause offence are tightly coupled with linguistic expertise -- as opposed to the display of gestures and facial expressions -- as we all know with non-native speakers of our languages who frequently offend, even though they have no desire to do so, and often have no awareness of the offence they cause. What name to call someone by, or whether or not to use "Sir", "Mister", "Miss", "Missus" are complex matters—now made harder with the rise of new gender pronouns -- known intuitively to native speakers but not to outsiders, who are never taught them and have nowhere to go for advice or instruction.

I personally find the Lady's Companion list above an attractive one: it avoids emotion beyond the linguistic, it implies care for the mental and emotional state of the user, and I would personally find it hard to abuse any computer with the characteristics listed above. Many of the situations discussed above are, at the moment, wildly speculative: that of a Companion acting as its owner's agent, on the phone or World Wide Web, perhaps holding power of attorney in case of an owner's incapacity and, with the owner's advance permission, perhaps even being a source of conversational comfort for relatives after the owner's death. Companions may not all be nice or even friendly: Companions to stop us falling asleep while driving may tell us jokes, but will probably also shout at us and make us do stretching exercises. Long-voyage Companions in space will be indispensable cognitive prostheses for running a huge interplanetary vessel and its experiments: Hollywood already knows all that, and created the terrifying nightmare Companion in HAL9000 in the film *2001*. All these situations are at present absurd, but perhaps we should be ready for them.