

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

Yorick Wilks (editor)  
(University of Oxford)

Amsterdam: John Benjamins Publishing Company (Natural Language Processing series, edited by Ruslan Mitkov, volume 8), 2010, xxii+315 pp; hardbound, ISBN 978-90-272-4994-4, \$149.00, €99.00; e-book, ISBN 978-90-272-8840-0, \$149.00, €99.00

*Reviewed by*  
*Judith Masthoff*  
*University of Aberdeen*

This book is an edited collection of chapters on artificial companions (ACs) resulting from a workshop. It purports to discuss the philosophical and ethical issues associated with ACs, what ACs should be like and how to construct them, and to provide examples of special-purpose ACs. Table 1 shows the chapters of the book and their respective authors.

When I bought a vacuum-cleaning robot called Roomba, it never occurred to me that this was actually a companion, as claimed by Peltu and Wilks in this book's afterword. Yes, it independently helps me, autonomously recharges its battery, and occasionally talks: "Clean brushes." I had higher expectations of a companion, however. Perhaps it would feel more like a companion if it expressed disgust with my mess (cf. Bee et al.), inquired knowledgeably about my holidays and family (cf. Wilks), empathized with my feelings (cf. Bevacqua et al.), and advised scraping off spilled porridge with a spoon (cf. Sloman).

There is no consensus in this book on what an AC should be like; each chapter tends to express an alternative (often quite radically different) opinion. In the foreword, Wilks defines ACs as "conversationalists or confidants" that get to know their owners, assist with Internet interactions, provide company and companionship, and build their owner's biography. However, when analyzing the necessary conditions for being an AC, Pulman concludes that conversation is not needed, and Boden strongly objects to ACs as confidants because of privacy concerns. Rather than providing company, according to O'Hara, ACs will represent (technophobe) owners in complex dealings with technology. Lowe takes this one step further and questions whether ACs are really others or whether they are cooler-headed versions of ourselves. Whereas Romano defines an AC as a good friend who, among other things, makes you laugh, shares your emotions, and listens, Sloman focuses on the (in his opinion) more difficult problem of creating ACs that provide help and assistance. Turkle sees ACs as offering care as well as being good company, teachers, and lovers (!). Taylor et al. deviate from the anthropomorphic view of ACs, and discuss a robot that has no human-like properties other than that it is fed on organic material.

There is also disagreement on whether ACs have bodies. In the foreword, Wilks states that ACs are not robots but software agents (though later he mentions that they could be furry handbags). In contrast, Turkle, Bryson, and many others clearly talk about robots. Winfield even talks about a family of robots, rather than just one.

Finally, there is disagreement on the role of emotions. Romano believes that the ability to interpret and express emotion is crucial, and some chapters are completely

**Table 1**Contents of *Close Engagements with Artificial Companions*.**Section I. Setting the Scene**

"In good company? On the threshold of robotic Companions" by Sherry Turkle

"Introducing artificial Companions" by Yorick Wilks

**Section II. Ethical and Philosophical Issues**

"Artificial Companions and their philosophical challenges" by Luciano Floridi

"Conditions for Companionhood" by Stephen G. Pulman

"Arius in Cyberspace: Digital Companions and the limits of the person" by Kieron O'Hara

**Section III. Social and Psychological Issues: What Should a Companion Be Like?**

"Conversationalists and confidants" by Margaret A. Boden

"Robots should be slaves" by Joanna J. Bryson

"Wanting the impossible: The dilemma at the heart of intimate human-robot relationships"  
by Dylan Evans

"Falling in love with a Companion" by David Levy

"Identifying your accompanist" by Will Lowe

"Look, emotion, language, and behavior in a believable virtual Companion"  
by Daniela M. Romano

"New Companions" by Alex Taylor, Anab Jain, and Laurel Swan

"On being a Victorian Companion" by Yorick Wilks

**Section IV. Design Issues: Building a Companion**"The use of affective and attentive cues in an empathic computer-based Companion"  
by Nikolaus Bee, Elisabeth André, Thurid Vogt, and Patrick Gebhard"GRETA: Towards an interactive conversational virtual Companion" by Elisabetta Bevacqua,  
Ken Prepin, Radoslaw Niewiadomski, Etienne de Sevin, and Catherine Pelachaud"A world-hybrid approach to a conversational Companion for reminiscing about images"  
by Roberta Catizone, Simon F. Worgan, Yorick Wilks, Alexiei Dingli, and Weiwei Cheng

"Companionship is an emotional business" by Roddy Cowie

"Artificial Companions in society: Consulting the users" by Alan Newell

"Requirements for Artificial Companions: It's harder than you think" by Aaron Sloman

"You really need to know what your bot(s) are thinking about you" by Alan F. T. Winfield

**Section V. Special Purpose Companions**

"A Companion for learning in everyday life" by Rebecca Eynon and Chris Davies

"The Maryland virtual patient as a task-oriented conversational Companion"  
by Sergei Nirenburg

"Living with robots: Ethical tradeoffs in eldercare" by Noel Sharkey and Amanda Sharkey

**Section VI. Afterword**

"Summary and discussion of the issues" by Malcom Peltu and Yorick Wilks

dedicated to this topic (Bee et al., Bevacqua et al.). In contrast, Boden believes that ACs will not be able to show real empathy, a view also expressed by somebody interviewed by Turkle: "How can I talk about sibling rivalry to something that never had a mother?" Wilks prefers ACs as Victorian companions, restrained in showing emotions. Romano also mentions the importance of politeness, although Wilks claims that politeness may not be needed, that the user may have personality preferences. In fact, Evans talks about ACs matching their owners' personality. Lowe goes even further and discusses that users may select their own AC, matching, for example, their religion and world views.

Of course, there is nothing wrong with scientists having different views on what an AC should be. However, for clarity, it would help to have more agreement on what should be covered by the term AC. If every autonomous software programme or robot (such as my Roomba) is called an AC, then AC overlaps with well-established terms such as *intelligent agent* and it is not clear what introducing AC as a concept contributes, nor what research is required to construct them. For this reason, I would have excluded, for example, technology assistants and problem solvers (close to existing intelligent

agents), patients (I do not think doctors regard their patients as companions), and teachers (there is a lot of work on teachers in the intelligent tutoring systems community). At the very least, a taxonomy of AC types should be established. Sloman provides a limited starting point, distinguishing between ACs meant to entertain and ACs meant to help, with some subcategories in the latter. Pulman's necessary conditions (having intentions towards their owner, recognizing their owner as an individual, having predictable behavior, being able to predict their owner's behavior, being independent) seem a sensible start towards a consensus definition. These necessary conditions are still rather vague, however. For example, what aspects and proportion of behavior should an AC be able to predict, given that even humans have difficulty predicting exactly what their human partners will do and say? Also, should ACs not be able to communicate in some form (even if not necessarily using natural language) to distinguish them from other types of adaptive systems? Peltu and Wilks provide another list of features in the afterword. While being more comprehensive, this list will also undoubtedly lead to more disagreement. Is it really essential that ACs are "cheerful even when faced with difficult problems"? I am not sure I could cope with an eternally cheerful companion.

The book raises many ethical issues. Should ACs be legally and morally responsible for their actions? Should people be allowed to befriend, get emotionally attached to, or love and marry ACs? Is it morally allowable that people are deceived into believing that ACs care about them if, in fact, they don't? Will ACs make us live beyond death (called e-mortal by Floridi)? Will ACs be preferred to human companions, regarded as better companions and better lovers? Turkle even muses "what are the purposes of living things?" This may all sound far-fetched, but the book provides many examples of how people are already reacting irrationally to existing very simple ACs, such as refusing to board a plane as their Tamagotchi would die, or giving their bed to a doll so it can have a good night's sleep. In the foreword, Wilks writes about human-Companion interaction, the capitalization almost implying that the AC will be more important than the human. Some ACs described in this book are clearly scary, often intentionally to highlight ethical issues. Wilks mentions Frankenstein's monster as the first AC. Sharkey and Sharkey discuss robot carers that may restrain an elderly person when they are doing something potentially dangerous (such as taking a knife out of a drawer), and that may take them back inside if they were to escape. Lowe discusses some potential ACs as "the punitive hand of the state," making you behave properly. Several ACs (such as the Junior Companion mentioned by Wilks) sound like Big Brother watching you (how would it make children feel if their parents knew what they were doing all the time?). A debate on what limits there should be to AC behavior is clearly required. This book provides many of the arguments for such a debate. Again, there is no clear consensus. For example, on the one hand, Levy argues that many people will no doubt fall in love with ACs and that this is completely normal; on the other hand, Bryson argues that ACs should just be "servants you own," machines that you can switch off whenever you like.

Overall, this book is an interesting read that raises many questions on what ACs should be and related ethical issues. It will generate discussion and provoke reflection. It provides insight into the kinds of ACs people are working on, and gives a hint of the technologies required. It does not, however, provide an overview of the current state-of-the-art and what current ACs are capable or incapable of (there is some discussion in Sloman's chapter on research that is still needed, but more of this would have been nice). There is also not much discussion on how to construct ACs. There is some advocacy of situated cognition (Taylor et al.), and some of a more traditional AI approach (Sloman). The most comprehensive chapters are those on emotions (Bee et al., Bevacqua et al.). The book further lacks discussion on user studies with

existing ACs: There are anecdotal user quotes and a much needed chapter by Newell highlighting the need for user involvement, but most chapters only discuss what ACs should be like or what ACs are being built without any mention of user feedback or user input into the design. Also, despite being part of a Natural Language Processing series, this book barely touches upon natural language processing. It is more or less restricted to Sloman indicating that work is needed on referring expressions such as “the mug I used yesterday” and Nirenburg showing how natural language processing is done in a Virtual Patient. Although in most chapters ACs are depicted as having conversations, there is hardly any mention of the natural language generation and interpretation required to obtain this ability. Despite these shortcomings, however, this remains an interesting book to read, discuss at lunch, and, occasionally, have a laugh about.

*Judith Masthoff* is a Senior Lecturer in Computing Science at the University of Aberdeen (UK). Her research interests include intelligent user interfaces, persuasive technology, affective computing, and the evaluation of adaptive systems. Masthoff’s address is University of Aberdeen, Department of Computing Science, Aberdeen AB24 3UE, UK; e-mail: [j.masthoff@abdn.ac.uk](mailto:j.masthoff@abdn.ac.uk).