



**HAL**  
open science

# The Robot Who Loved Me: Building Consciousness Models for Use in Human Robot Interaction Following a Collaborative Systems Approach

Adamantios Koumpis, Maria Christoforaki, Siegfried Handschuh

► **To cite this version:**

Adamantios Koumpis, Maria Christoforaki, Siegfried Handschuh. The Robot Who Loved Me: Building Consciousness Models for Use in Human Robot Interaction Following a Collaborative Systems Approach. 19th Working Conference on Virtual Enterprises (PRO-VE), Sep 2018, Cardiff, United Kingdom. pp.409-416, 10.1007/978-3-319-99127-6\_35 . hal-02191182

**HAL Id: hal-02191182**

**<https://inria.hal.science/hal-02191182>**

Submitted on 24 Jul 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# The Robot Who Loved Me: Building Consciousness Models for Use in Human Robot Interaction Following a Collaborative Systems Approach

Adamantios Koumpis, Maria Christoforaki, Siegfried Handschuh

Universität Passau, Fakultät für Informatik und Mathematik, Germany

<sup>1</sup>[name.lastname@uni-passau.de](mailto:name.lastname@uni-passau.de)

**Abstract.** We build on the results of our MARIO project to support the benefit of informing research taking place in the area of development of human consciousness models for use by robots to facilitate human robot interaction with research results from the area of collaborative systems. The main outcome of such a research would be a software model of a robot assistant that might offer functionality and features of a ‘virtual coach’ with an in-built evolution capability.

**Keywords:** human-robot interaction, consciousness modelling, distributional semantics, surrogate (virtual) objects, data curation of semi-structured and unstructured data

## 1. Introduction

We shouldn’t see the future as black. On the other hand, our daughters may be thousands of miles away from us – and our partners either dead or on a cruise ship in the Caribbean. So we are ‘home alone’ – but in contrast to the (now rather old) movie hero that was 8 years old, we are 80+. Luckily, we are not totally alone as we have a robot that lives with us. The robot follows our routine and for some few days now is spotting that:

- We are missing meals
- We are not taking any calls
- We are not leaving home
- We skip exercises and avoid taking our medication

Is it depression? Is it loneliness? Is there something that the robot should do to help? In the European Horizons2020 MARIO project ([www.mario-project.eu](http://www.mario-project.eu)) we have been building Apps that would help the robot entertain persons with dementia. But we haven’t thought about reaching a higher level, namely of equipping the robot with the necessary consciousness model that will help it become a self-starter in many everyday situations that a human might operate well. There is surprisingly much research work and bibliography in the area of collaborative systems, some of which also relevant to elderly care, that can be of direct value and use in the area of human robot interaction. In this paper we aim to make clear the need for synergies between the two areas.

By January 2018 our European research project MARIO ([www.mario-project.eu](http://www.mario-project.eu)) came to its end. In MARIO we introduced a humanoid care robot to help and facilitate interactions with persons with dementia – an all in all difficult ‘target’ group [1]. However, and going one level deeper in the project and its research agenda, we were addressing the difficult challenges of loneliness and isolation in older persons with dementia through innovative and multi-faceted inventions delivered by service robots. The effects of these conditions are severe and life-limiting. They burden individuals and societal support systems. Human intervention is costly but the severity can be prevented or mitigated by simple changes in self-perception and brain stimulation mediated by robots. In [2] is stated that until the present, ‘loneliness is being treated as a symptom of mental health problems; however, for elderly (aged 60 years and above), loneliness has become a disease in itself. There are epidemiological, phenomenological, and etiological reasons to say that’.

What we have identified during the project lifetime and both as result of our research activities, but also as result of our interactions with patients, their relatives and the carers, is that there is well identifiable space which may have been treated as a blindspot as of today and which concerns the area of modelling of human consciousness mechanisms to govern the robot. Human consciousness in our context refers both to the robot that may need to build some primitive ‘consciousness’ to support its interactions with the humans in the form of empathy or affection, but also from the side of the human user so that in case of loss of abilities due to dementia or some other reason, they will be assisted for a variety of functions and tasks.

This means that while in MARIO project we mainly cared about offering interventions, there is need for some new research that will analyse human consciousness mechanisms, identify their main features, and identify the border between human consciousness and unconsciousness to support reasoning, learning and adaptation to varying emotional and behavioural patterns, conditions and preferences.

The main outcome of such a research would be a software model of a robot assistant that might offer functionality and features of a ‘virtual coach’ with an in-built evolution capability. At a first level, such a capability should be based on routine tasks and some initial relevant physiological and behavioural data; over time it should become more specialised to reflect the particular needs and preferences of each individual older person.

In [3], the authors pose a rhetoric question, asking whether it is ‘possible to devise a general architectural principle that might lead an artificial agent to exploit what is called consciousness in human beings and various animals?’, stating also that ‘in a scientific and technological context, whatever formulation of consciousness one opts to consider, consciousness should not come at the end of the day as an addition to an otherwise working description of a physical system’. For such a ‘working description of a physical system’ it is that the authors of [4] recognize that one has to look into the construction of agent-oriented design processes, and more specifically what the authors call ‘the definition of the best way to create ad-hoc agent oriented design processes’. In this respect, the development of consciousness as a multi agent system would, following the authors’ argumentation line, require ‘great efforts in learning and using an existing design process’ and the main idea of the approach taken is one has to include in the design process also the system metamodel.

In [5] the author reports about the advances made in the IT industry in the area of machine reading (to avoid the term distant reading introduced by Moretti in [6] and which has been in the last years mainly focused on the ‘reading’ of literature, and what was considered as one of artificial intelligence’s holiest grails for many years. What is interesting and relevant to our context here is a remark that one of the interviewed researchers (Ernest Davis, a New York University professor of computer science and AI researcher) provided, namely that this (: machine reading) is ‘technically an accomplishment, but it’s not like we have to begin worshipping our robot overlords’, adding what epitomizes most successfully the point we take in our research namely that ‘When you read a passage, it doesn’t come out of the clear blue sky: It draws on a lot of what you know about the world’.

## 2. The Main Idea

The idea of approaching real world objects through knowledge graphs automatically created from multimodal information sources is, at least to our knowledge from the literature, uncommon. However, what might be a more uncommon idea is the case of automatically building and populating human consciousness models with the use of knowledge graphs.

Knowledge discovery from digital big multimodal data can be done effectively by employing techniques like machine learning (such as Deep learning, SIFT), distributional semantics in NLP e.g. for text mining purposes and semantic augmentation (use of domain specific formal ontologies) [7].

However, and in order to facilitate the modeling of human consciousness, such an approach can be seen as a process where studying the actual artefact (i.e. the human consciousness) is replaced by studying a latent (virtual) object consisting of a dense network of related information. This latent object is produced from secondary sources, namely digital or digitised images and texts or other modalities, where the image or the text is used as an interface to the latent object, to which the human user has no access to.

We see a similarity with the case of an art historian studying artefacts from images and catalogues or essays, or a medical doctor trying to identify pathogenies from medical images and test data. In both cases the form of information can be:

- structured data (e.g. data from a collections management system records or medical test results)
- unstructured text (e.g. free text description of an artefact)
- images (e.g. digital or digitized images of museum objects, or CAT scans)
- formalized domain knowledge existing in bibliography and in the experience of the person examining the material

These sources constitute already a limiting approach to the true nature of the object as such, since they constitute already ‘processed’ knowledge, thus eliminating facets of it (e.g. same as 2D images provide us with information about the form of an object from a certain perspective only, textual descriptions offer information only deemed important by the author of the document, etc.).

Moreover, the application of algorithms in these already secondary sources are perceived as black boxes, thus giving the domain experts no transparent results to construct an argumentation supporting their research.

### 3. The Need for a Collaborative Systems Perspective

There is a qualitative difference between the term *need* and this of a *necessity*; though both words are pretty close, denoting some similar condition, the difference between them lies on the prioritization, as while for a need there is an absolute demand, the necessity also relates to a craving without which a person manages to survive.

Our main position in this position paper is that there is a well identifiable space for research to take place in the following years that will aim to support the field of human-robot interaction and in particular the area of content-creation with research that will be, at a first level informed, and at a further level building on the work of collaborative systems. Work conducted and reported in [8] and [9] is of direct relevance in a variety of ways as we shall be able to indicate below, where we see a great potential in the collaboration and the synergies for future research with the help of the PRO-VE community.

In [10], the authors state that ‘in order to have a robotic system able to effectively learn by imitation, and not merely reproduce the movements of a human teacher, the system should have the capabilities of deeply understanding the perceived actions to be imitated’. In our context, we need to combine all existing knowledge patterns and information sources that may be considered as relevant to a situation that our MARIO robot may encounter, while also relate them or, if more appropriately, match them with a situation that the robot may have just or recently encountered with a human user, like this of holding a conversation on a family matter (planning for the visit of a relative or need to have some medical tests). To this aim we take a holistic approach that combines bottom-up and top-down approach. In particular:

- bottom-up approach that is based on the employment of Big Data analytics techniques to explore what may in the future constitute some potentially relevant patterns by having access to a vast pool of online media resources like social media, internet discussion forums, blogs, etc. that are relevant to the discussion topic between the human and the robot, as well as a
- top-down approach that takes as starting points existing patterns or instances of past interactions between the robot and the human in the recent past, and check for their relevance.

As one may expect, the superiority and the longer term sustainability and value of our approach relates to the combination of both bottom-up and top-down approaches. Only the top-down would result into yet another ‘spell-checker’; only the bottom-up would result into an endless crawling of the Web where patterns might be ignored even if they would exist because they couldn’t be identified using the current information. Same as in crime forensics where DNA profiling may have been used after several years to identify individuals by characteristics of their DNA, our tools shall offer the capability to store patterns that will be identified after some time, when some increased information will have been collected. This is an important feature of our approach.

Further to this, complementary and potentially relevant patterns will be identified by means of algorithms in statistical analysis and machine learning for sentiment extraction and social network analysis (Bottom-Up-Approach). In particular, this approach has the potential to uncover hitherto unknown patterns in the source data and therefore enable the construction of new hate speech categories related to a specific field and knowledge base. The outcomes of both approaches will be united in a consistent, unified formal knowledge representation, to allow both for an extended and improved understanding of the subjects of investigation and for a long-term, versatile reusability of the collected data. At our group we have developed data-driven methods for data-driven decision making, recommendation and contextual suggestion engines, using and/or developing algorithms for classification, clustering, rule extraction and sentiment analysis.

To the above one may find an instance of the problem of origins and first cause or what we usually call as a chicken and egg situation. We have been all educated to use in relational data and database modeling as this has been typically the prevailing paradigm that was also driven by the structure of available data for many years now. The key question that has been driving the development, the research and the business there and is also used as the main design theme is: “*What answers do I have?*”

However, today’s problems and challenges as well as modern data and database modeling (what is also usually called as ‘NoSQL’) is typically driven by application-specific access patterns, i.e. the types of queries to be supported. So the main question and also the main design theme here is: “*What questions do I have?*”

It is for this reason that collaborative systems are relevant and highly necessary to consider as an enabling technology for all future developments in the areas of intelligent systems and in particular for the area of human robot interaction.

One should be able to intuitively see the difference and also the parallels we see with the modelling of human consciousness. As an example we can imagine a library with all sorts of books – the books that a person may have read in their lives. We are all aware of the Dewey system – or as it is properly called: the Dewey Decimal Classification. This is an ideal role model example on why people came up with the concept and the idea of relational databases. And, by the way, this is a system that started being shaped back in the 1870s.

However, if one imagines that our driver is ***not*** to find:

- if a book exists in our library or
- who has borrowed it,
- who is the publisher,
- when it was published or
- other similar ‘relational’ questions,

but:

- anything about the contents of the books,
- anything that might be unexpected, unplannable, impossible or even at some point to consider as unworthy to even model when building our ‘neat and cute’ relational database reality,

it is in this case, ***where we don’t know which possible questions we have***, where the need for adopting a collaborative systems approach comes on stage.

There is always a risk with the quality of the information acquired from unknown sources – a good example for this relates with the case of fake news. If the consciousness model for our case relies on any type of lesser value content such as the case of fake news, it is obvious that this will affect the quality of the model as such. To this problem, however, we can take a similar stand to the one we take in real life: not all people build their consciousness on the same sources or may be able to exhibit some evidence-based consistency in the way their consciousness models are built. But it is therefore that a CN-based approach would increase the value of the end system, as the aggregate results of such a consciousness model that comes as outcome of the collaboration amongst can will tend to be more accurate than any single entity alone. This collaboration aspect should be not mistakenly related to with crowdsourcing that is rather the process where knowledge is obtained from the public at large through a system of invitations and incentives [12, 13].

### **3.1. A Word on Ethics**

The ethics work during the MARIO project provided an innovative, integrated approach to addressing ethical issues in robot design, research and implementation, addressing the entire development process of the project [11]. As developed within the project lifetime, the ethics framework we developed provided a new comprehensive review of relevant interdisciplinary literature with relevance for understanding ethical concerns regarding the use of care robots, a useful contribution to the field of the ethics of care robots. In addition, its findings informed the approach to data management in the project.

On the basis of the MARIO ethics framework, a set of innovative practical tools in the form of five ethics checklists were developed to guide ethical practice for different stakeholders in robot development, research and implementation. These were tested by different stakeholders, who found them to be helpful practical tools for their respective practice areas. An “ethical encountered issues” form was employed during the pilots, to capture evidence on the incidence of actual ethical issues that stakeholders encountered during trials. These indicated that careful implementation was required, but that users generally tolerated the robot well and distress was rare.

These tools have the potential to be employed in different care robot projects, and the ethics tools and experiences from the project will be published to make them available to the international robot community, contributing to the development of a solid evidence base of ethical concerns in care robotics, which, if those results remain as positive as during the MARIO project, is likely to facilitate wider societal acceptance of the use of care robots in the care setting. It is interesting in this context to take into account that in [14] the authors identify in relation to service robots a problem related, amongst others, with loss of control from the side of the elderly person due to robot monitoring and the interventions that the latter undertake within their daily activities and routine. This is true at a great extent, and same as the (feeling of a) loss of privacy and the (feeling of) infantilisation. However, one has to admit that the benefits of such a technology equal out at a great extent all potential drawbacks.

#### 4. Conclusions

Developments in the addressed area of modeling human consciousness in computable and computer-processable forms are still in a fluid form. This poses a challenge for bringing together a corpus of research that has taken and is still taking place in the area of collaborative systems to identify and, if possible, also seize challenges that will change the way we perceive our idea for the use of robots in everyday tasks. The result of a robot intelligence will not anymore rely on its ‘hardwired’ intelligence but, with use of sentiment analysis and deep learning technologies, their capabilities to extract information from a variety of modalities and media including videos, images, natural language, as well as structured data and then automatically transform them into ‘actionable information’ by means of deploying them in their interactions with the human users.

It is worth to mention here that in the MARIO project, as well as in the context of the present paper, we refer to robots that act as an extension of humans offering to them a symbiotic perspective, so these are not what one would consider as ‘truly autonomous robots’. However, the aim to model aspects of the human consciousness is dual namely on the one hand to replicate it in robots, while on the other hand to help robots extend and complement the consciousness of people with decaying cognitive ability.

From our research expertise it is evident that we promote the idea of exploiting developments in the areas of semantic technologies and statistical based natural language processing to facilitate the task of building data-driven models of body consciousness that will be possible to be easily populated with big data acquired from a variety of sources. Furthermore, we are able to see that the improvements to be made will seriously lag behind the needs that exist if our research shall not be informed and enriched with the perspective taken in the area of collaborative systems.

Human intelligence is in all its aspects and manifestations such a system, consisting of a variety of entities and artifacts, demonstrating also characteristics of autonomy as well as synergy, geography is also of importance with the ability to take advantage of geographical distribution as well as of geographical concentration, while last but not least, heterogeneity is also a key term as it affects all levels like the culture, the socio-economic context, or even the daily (operational) routines.

It may seem like a novelty to think of a future with networks of collaborating robots that are working together to offer a service to an individual. However, and while from a technology stance this may be a novelty, from an organizational point this is not at all: currently, and for the care to an old person, there may be their children that will interact with the physicians and the nurses. In the future, all these interactions may take place in an automated fashion by robots that will live with us same as the MARIO robot does with the persons with dementia.

#### Acknowledgements

This work was supported in part by a grant from the European Union Horizons 2020 – the Framework Programme for Research and Innovation (2014–2020) under grant



agreement 643808 Project MARIO “Managing active and healthy aging with use of caring service robots”.

The authors would like to our thanks and appreciation to the two anonymous reviewers that provided us with good advice and input to our ideas and thoughts.

## References

1. D. Casey, H. Felzmann, G. Pegman, C. Kouroupetroglou, K. Murphy, A. Koumpis, and S. Whelan, What People with Dementia Want: Designing MARIO an Acceptable Robot Companion, 15th International Conference on Computers Helping People with Special Needs, July 13-15, 2016, Linz, Austria.
2. Tiwari, Sarvada Chandra, Loneliness: A Disease?, Indian Journal of Psychiatry, 55.4 (2013): 320–322. PMC. Web. 29 Jan. 2018.
3. R. Manzotti, A. Chella, Physical integration: A causal account for consciousness, Journal of Integrative Neuroscience, 13, 403, 2014.
4. V. Seidita, M. Cossentino, and A. Chella, A Proposal of Process Fragment Definition and Documentation, in Massimo Cossentino, Michael Kaisers, Karl Tuyls Gerhard Weiss (Eds.) Multi-Agent Systems, 9th European Workshop, EUMAS 2011, Maastricht, The Netherlands, November 14-15, 2011, Springer Verlag, Lecture Notes in Artificial Intelligence 7541, 2012.
5. D. Harwell, AI’s ability to read is hailed as milestone, but there’s still long way to go, originally published January 21, 2018, <https://www.seattletimes.com/business/ais-ability-to-read-is-hailed-as-milestone-but-theres-still-long-way-to-go/>, accessed on 3 May 2018.
6. F. Moretti, Distant Reading, London, Verso, 2013.
7. J. Efon Sales, A. Freitas, B. Davis, S. Handschuh, A compositional-distributional semantic model for searching complex entity categories, Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics, pp. 199-208, 2016.
8. Baldissera, T. A., & Camarinha-Matos, L. M. (2016). Towards a collaborative business ecosystem for elderly care. In L. M. Camarinha-Matos, A. J. Falcão, N. Vafaei & S. Najdi (Eds.), Technological Innovation for Cyber-Physical Systems, IFIP AICT series 470/2016, Springer, pp 24-34.
9. Camarinha-Matos, L. M., Rosas, J., Oliveira, A. I., & Ferrada, F. (2015). Care services ecosystem for ambient assisted living. *Enterprise Information Systems*, 9(5-6), 607-633.
10. A. Chella, H. Dindo, and I. Infantino, Anchoring by Imitation Learning in Conceptual Spaces, in S. Bandini and S. Manzoni (Eds.): Advances in Artificial Intelligence, AI\*IA 2005, IX Congress of the Italian Association for Artificial Intelligence (AI\*IA), Milan September 21–23, 2005, LNAI 3673, pp. 495 – 506, Springer Verlag, 2005.
11. Felzmann, H.; Murphy, K.; Casey, D.; Beyan, O. (2015) Robot-assisted care for elderly with dementia: is there a potential for genuine end-user empowerment? The Emerging, Oral presentation to the Policy and Ethics of Human Robot Interaction Conference, DOI <http://dx.doi.org/10.13025/S8SG6Q>
12. Prpić, J., & Shukla, P. (2016). Crowd Science: Measurements, Models, and Methods. In Proceedings of the 49<sup>th</sup> Annual Hawaii International Conference on System Sciences, Kauai, Hawaii: IEEE Computer Society.
13. Gerber, E. M., & Hui, J. (2013). Crowdfunding: Motivations and deterrents for participation. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(6), 34.
14. Sharkey, A., and Sharkey, N. 2012. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology* 14, 1, 27-40.